Article

Benchmarking Quantum Mechanical Levels of Theory for Valence Parametrization in Force Fields

Pavan Kumar Behara,* Hyesu Jang, Joshua T. Horton, Trevor Gokey, David L. Dotson, Simon Boothroyd, Christopher I. Bayly, Daniel J. Cole, Lee-Ping Wang, and David L. Mobley

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A wide range of density functional methods and basis sets are available to derive the electronic structure and properties of molecules. Quantum mechanical calculations are too computationally intensive for routine simulation of molecules in the condensed phase, prompting the development of computationally efficient force fields based on quantum mechanical data. Parametrizing general force fields, which cover a vast chemical space, necessitates the generation of sizable quantum mechanical data sets with optimized geometries and torsion scans. To achieve this efficiently, choosing a quantum mechanical method that balances computational cost and accuracy is crucial. In this study, we seek to assess the accuracy of quantum mechanical theory for specific properties such as conformer energies and torsion energetics. To comprehensively evaluate various methods, we focus on a representative set of 59 diverse small molecules, comparing approximately 25 combinations of functional and basis sets against the reference level coupled cluster calculations at the complete basis set limit.
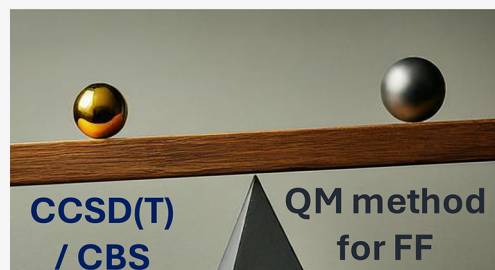
1. INTRODUCTION

General force fields offer a computationally efficient alternative to quantum mechanical (QM) calculations,[1−13] in particular for pharmaceutical and biomolecular applications. Parametrizing general force fields or training machine learning (ML) potentials, for molecular mechanics (MM), requires extensive quantum chemistry data sets, with molecules covering a large chemical space, that inform the bonded parameters, typically bonds, angles, and torsions as well as nonbonded terms.[14−21] The data sets include optimized chemical structures of different conformers of molecules and associated properties such as charges, bond orders, dipole moments, Hessian matrices, and torsion energy profiles for rotatable bonds, dimer interaction energies, among other data. A quantum mechanical method is needed that is globally accurate for different chemistries and properties, is computationally cost-effective, and can converge with a predefined set of hyperparameters (or a template with minimal changes) without too much human intervention, thus enabling automated generation of larger data sets essential for this effort. There are hundreds of density functionals, post Hartree−Fock methods, and basis sets to choose from, and it is cumbersome to pick one without a proper benchmark study.

Recent benchmark studies address the issue of how accurate a quantum mechanical method is with respect to a highly accurate gold standard such as coupled cluster with singles, doubles and perturbative triples in the complete basis set limit (CCSD(T)/CBS). Some studies are from the viewpoint of quantum chemistry method developers, where one would be concerned by a difference of even 0.1 kcal/mol in mean accuracy of absolute energies, with respect to the reference level of theory. However, from the perspective of practitioners engaged in force field development and general biochemical simulations, where relative energies such as conformer energetics and torsional profiles are of primary concern (as emphasized in Sellers et al.[22]), it is crucial that the quantum mechanical (QM) methods used to generate reference data for force field parametrization can reproduce torsion energy profiles with deviations of approximately 0.5 to 1.0 kcal/mol or smaller. This level of accuracy, as demonstrated in high-level QM calculations, is vital for ensuring the accuracy of trained force fields, and boosts the reliability of property predictions like protein−ligand binding affinities.[22,23] By achieving accuracy within this specified range for the QM reference data employed in force field parametrization, and assuming the trained force field mimics the QM training data quite well, we can effectively capture the essential conformational behavior and energetic trends exhibited by the molecules under investigation. Our focus here is precisely in this area, conformational energetics, and not on modeling long-range electrostatic interactions, which may involve different considerations.[24−27]
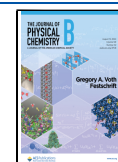
Our primary focus is on accurately modeling small molecule organic compounds within pharmaceutically relevant regions of chemical space using force fields. As a result, we do not concern ourselves with evaluating the performance of quantum mechanical (QM) methods for compounds involving transition elements, lanthanides, noble gases, and other elements not directly implicated in the fundamental molecular interactions governing the binding and recognition of small molecule drugs. At the Open Force Field (OpenFF) Initiative,[3,28] our attention lies on achieving performance benchmarks for organic molecules within the chemical space that is pertinent to small molecule drugs and that our force fields are currently capable of modeling. This chemical space encompasses C, H, O, N, P, S, F, Cl, Br, and I, as well as the monatomic ions Li^+, Na^+, K^+, Rb^+, F^-, Cl^-, Br^-, and I^-.

A variety of different method and basis set combinations in gas-phase calculations have been used in building various general force fields as well as ML potentials, depending on the force field and application. Some other common models were derived as follows:

- GAFF: MP2/6-31G* was used as one of the sources of training bond and angle parameters. For the torsional parameters MP4/6-311G**//MP2/6-31G* was used.[4]
- CHARMm: valence parameters were trained on QM data calculated at the MP2/6-31G* and MP2/6-31+G* levels.[29]
- ANI-1: Training data generated at ωB97X/6-31G* level of theory. ANI-1ccx used CCSD(T)/CBS extrapolated structures.[30]
- OPLS3e: M06-2X/cc-pVTZ(-f)//B3LYP/6-31G* level of theory was used for their training set.[10]
- TorsionNet: a mix of data generated at B3LYP/6-31G**//HF/3-21G and B3LYP/6-31+G**//HF/3-21G.[31]
- OrbNet Denali: ωB97X-D3/def2-TZVP level of theory.[32]
- Open Force Field (OpenFF): B3LYP-D3(BJ)/DZVP was the default level of theory.[2]
- Espaloma: A mix of SPICE[16] data set and OpenFF data sets, all of the data generated at B3LYP-D3(BJ)/DZVP level of theory.[18]
- MACE-OFF23: SPICE[16] data set at ωB97M-D3(BJ)/def2-TZVPPD level of theory.[21]

Traditionally there are five rungs on "Jacob's ladder" of chemical accuracy corresponding to the density functional theory (DFT) method used.[33] Starting with the Local Density Approximation (LDA), we progress one level up the accuracy ladder each with Generalized Gradient Approximation (GGA), meta-GGA, hybrid functionals, and then double hybrid functionals. Although, all hybrid functionals do not necessarily include the ingredient added in "meta-GGA" which uses the second derivative of the density in the energy functional. For example, B3LYP does not have the meta-GGA term, but M06 does. Hybrid and double-hybrid functionals are distinguished by including a percentage of HF exchange, and a percentage of MP2 correlation, respectively. Therefore, double hybrid functionals are at least the cost of a MP2 calculation. Range separated functionals (RSF) are another very important development in DFT methods. In RSFs, such as ωB97 family, the percentage of HF exchange depends on the distance between the electrons. This accounts for the electron self-interaction error, and has become incorporated into several of the most modern functionals (i.e., the ones starting with LC or ω). DFT methods

with dispersion corrections, which can be either *ab initio*,[34−36] or empirical in nature,[37] are notably more accurate than their standard forms as they improve the description of the noncovalent interactions. Grimme's DFT-D3,[38,39] along with Becke-Johnson (BJ) damping,[40] is one of the most prominent dispersion corrections widely used in the field. On the other hand, wave function based methods (WFT) include Hartree−Fock, post-Hartree−Fock methods viz., Møller−Plesset perturbation theory of different orders (MP2, MP3, MP4), configuration interaction (CI), coupled cluster (CC), and multireference methods such as complete active space self-consistent field (CASSCF), in increasing order of complexity in describing electron correlation, and approaching higher accuracy.

Recent studies have utilized several benchmark data sets to assess the general quality of QM methods for specific molecular properties. Notable among these data sets are Database2015B,[41] GMTKN30[42] and its expanded version GMTKN55,[43] MGCDB84,[44] CHAL336,[45] MPCONF196,[46] YMPJ,[47] S22,[48] S66,[49] HB375,[50] TorsionNet500,[31] Dakota Folmsbee's set,[51] XB18,[52] CTB-279.[53] The subsets of these data sets emphasize performance on predicting:

- noncovalent interactions, both inter- and intramolecular
- isomerization energies
- chemical reaction barrier heights
- relative conformer energies for various chemical moieties
- torsional strain, and barriers

In our present study we want to address

- the extent to which benchmark results of quantum mechanical methods obtained from larger basis sets, as documented in existing literature, are informative when deployed with smaller basis sets
- how good OpenFF's current choice of DZVP basis set is in describing conformational energetics for charged entities, and whether the error in relative energies of torsion profiles is within the desired range (0.5−1 kcal/mol) without addition of more diffuse functions to the basis set
- how well different QM methods predict torsional energy profiles and whether the accuracy gain is significant enough that it warrants a change in QM reference for force field training

2. MOTIVATION AND APPROACH

2.1. Prior Studies and Selection of Molecules. Based on earlier benchmark studies,[2,54] the OpenFF initiative initially chose B3LYP-D3(BJ)/DZVP as the method for generating QM training data to train OpenFF force fields, with the expectation that this choice might need to be revisited at a later date. Such a "revisit" is our focus here. The prior benchmarks[2] included assessing accuracy of conformer energetics on the MPCONF196 set,[46] which is a data set of conformers of smaller peptides and medium-sized macrocycles, *that were all neutral*, and of relative energies of torsional profiles on a curated set of 15 one-dimensional torsion scans. A good compromise between accuracy and cost on MPCONF196 and the smaller set of torsion scans led to the choice of the B3LYP-D3(BJ) functional along with Salahub's double-ζ split-valence + polarization (DZVP) basis set[55] for building OpenFF force fields.

Although very insightful, this prior benchmark study did not include charged molecules, and was limited to [C, H, N, O]
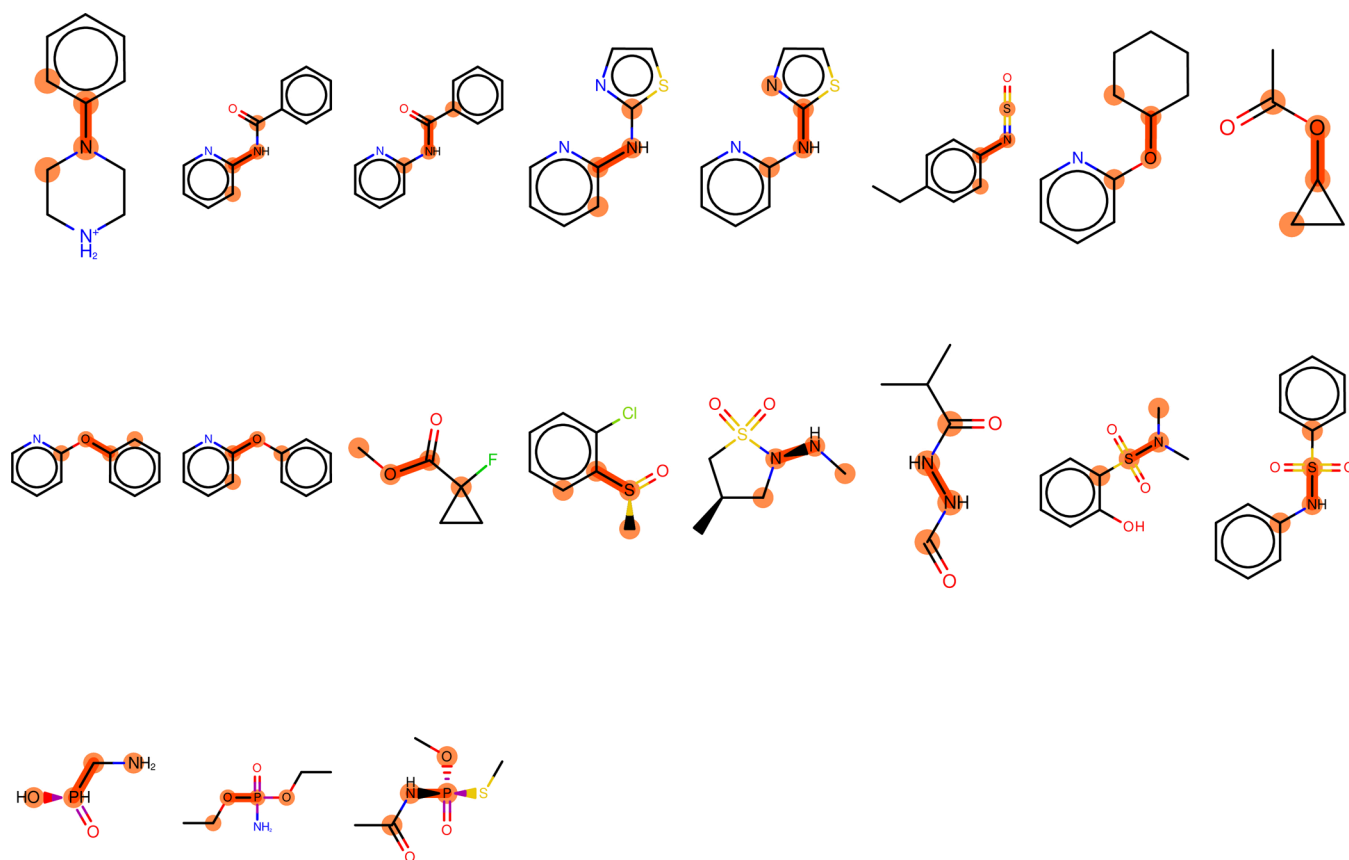
Figure 1. continued
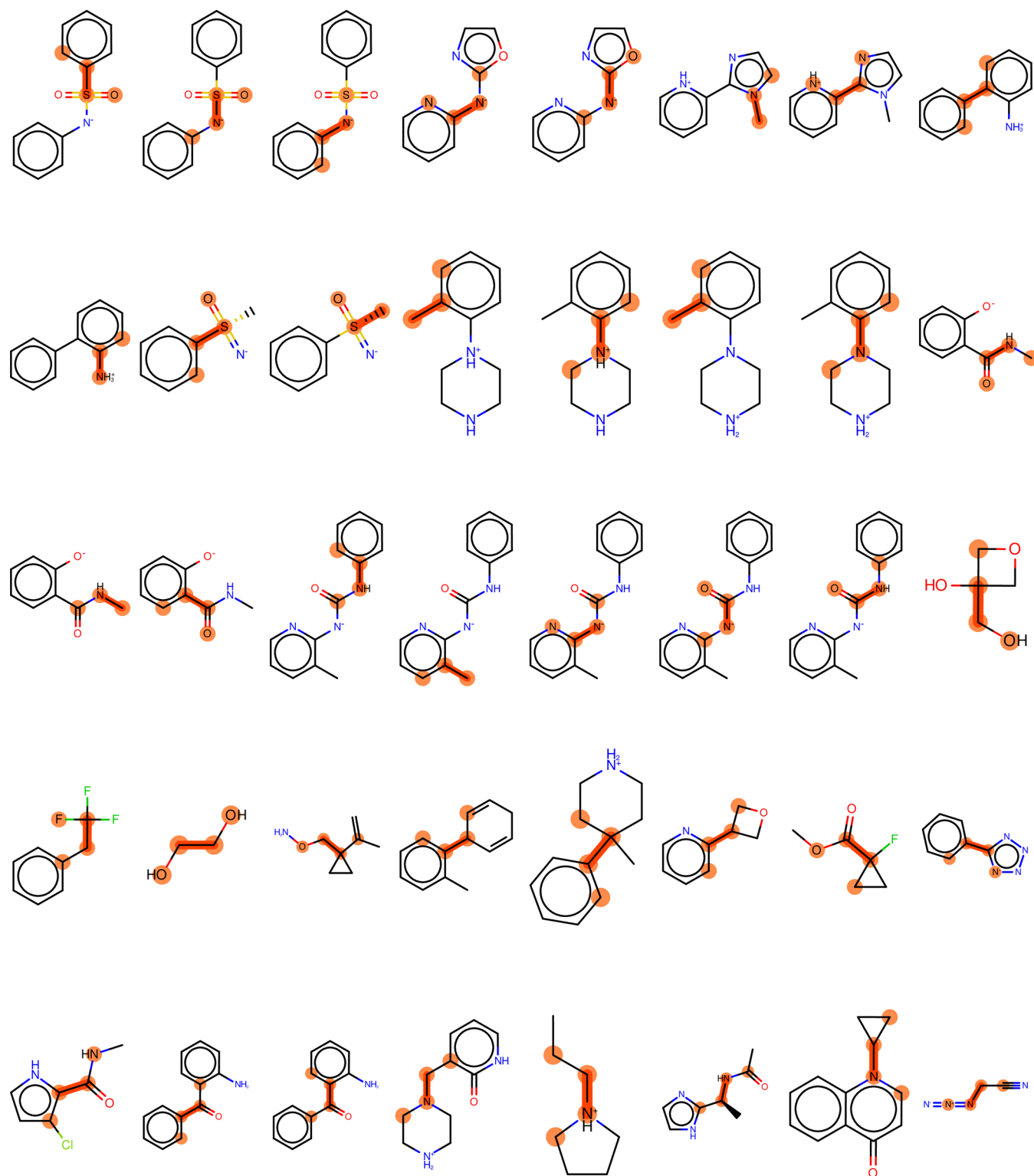
Figure 1. Molecules used in this benchmark study. The atoms and central bonds engaged in the driven torsions are highlighted.

chemical space. Because of these limitations, here we focused on running a new benchmark which would more adequately represent the pharmaceutically relevant chemical space that our force field needs to describe accurately. Thus, to expand the scope of the prior benchmark we selected 59 torsions from molecules in OpenFF's Roche (general, and tautomer+protomer data set) and Coverage molecule sets.

Molecule selection followed these criteria:

- ensuring a comprehensive representation of chemical diversity
- prioritizing size similarity among molecules to mitigate the influence of intrinsic size-related effects on electronic properties
- incorporating variations in central bonds, formal charges, element compositions, and intramolecular interactions
- and, inclusion of molecules that added additional complexity beyond those in the prior set, such as those with
  - nonzero formal charges
  - strong internal interactions
  - conjugated central rotatable bonds (>10 kcal/mol rotational barrier)
  - halogens

The extracted charged molecules were further subdivided into groups based on the type of charged functional group. Here are the SMARTS patterns of the subgroups which include,

- functional groups with a charge of −1
  - phenolate: c[O-]
  - carboxamide: C(=O)[N-]

- aromatic carbons connected to a nitrogen with a negative charge: c[N-]c
- sulfonamide: S(=O)(=O)[N-]
- sulfinamide: S(=[N-])(=O)
- functional groups with a charge of +1:
  - protonated amine with carbon bond: [NH+,nH+](=,:[C,c])[C,c]
  - protonated amine: [NH+]([*])[*]
  - protonated primary amine: [NH2+]([*])[*]
  - protonated ammonium: [NH3+][*]

We then performed one-dimensional torsion scans on one selected molecule per cluster. We selected one molecule per cluster by picking the centroid using the MACCS keys fingerprint method.[56] Following selection, we performed one-dimensional torsion scans by rotating around the central bond in the chosen dihedral. In total, 59 torsions were scanned in 36 unique molecules, shown in Figure 1 with the dihedral atoms highlighted, in this benchmark study.

2.2. Computational Details. Torsion scans were performed with a 15-degree grid spacing, using Torsiondrive,[57] resulting in a total of 24 points on the grid within the range of [-180°, 180°]. More details on the compute infrastructure were provided in SI section 1.2.

2.2.1. Choice of Reference Theory. The benchmark geometries were the final geometries from the torsion scan at MP2/heavy-aug-cc-pVTZ level of theory. Whereas, the benchmark relative energies were obtained using coupled cluster with single, double, and perturbative triple excitations at the complete basis set limit, CCSD(T)/CBS. It is to be noted that MP2 is a wave function-based method that treats electron correlation explicitly, and does not rely on the same approximations as DFT

Table 1. Overall RMSE and MUE (in kcal/mol) in Torsion Profile Energies of the Molecule Set with Respect to the Reference Level CCSD(T)/[haTQZ; δ:haDZ]//MP2/heavy-aug-cc-pVTZ Level of Theory[a]

	metric (kcal/mol)		
specification	RMSE	MUE	RMSE-TRLR
B97-D3(BJ)/def2-TZVP	$0.5309_{0.4981}^{0.5591}$	$0.3490_{0.3264}^{0.3694}$	$0.4524_{0.4136}^{0.4872}$
M05-2X-D3/DZVP	$0.5089_{0.4770}^{0.5418}$	$0.3214_{0.3024}^{0.3423}$	$0.3846_{0.3516}^{0.4152}$
M06-2X-D3/DZVP	$0.4492_{0.4214}^{0.4778}$	$0.2898_{0.2734}^{0.3084}$	$0.3582_{0.3326}^{0.3846}$
M08-HX-D3/DZVP	$0.5197_{0.4888}^{0.5507}$	$0.3423_{0.3222}^{0.3632}$	$0.4119_{0.3824}^{0.4424}$
ωB97X-D3BJ/DZVP	$0.4408_{0.4125}^{0.4681}$	$0.2808_{0.2635}^{0.2986}$	$0.3421_{0.3131}^{0.3716}$
ωB97M-D3BJ/DZVP	$0.4141_{0.3880}^{0.4397}$	$0.2618_{0.2452}^{0.2788}$	$0.3177_{0.2895}^{0.3466}$
ωB97M-V/DZVP	$0.4219_{0.3935}^{0.4503}$	$0.2694_{0.2530}^{0.2870}$	$0.3320_{0.3013}^{0.3639}$
PW6B95-D3(BJ)/DZVP	$0.4341_{0.4086}^{0.4577}$	$0.2809_{0.2642}^{0.2994}$	$0.3310_{0.3080}^{0.3539}$
PW6B95-D3/DZVP	$0.4371_{0.4120}^{0.4621}$	$0.2878_{0.2719}^{0.3053}$	$0.3311_{0.3098}^{0.3514}$
MP2/aug-cc-pVTZ	$0.6947_{0.6487}^{0.7380}$	$0.4520_{0.4230}^{0.4798}$	$0.5914_{0.5314}^{0.6504}$
MP2/heavy-aug-cc-pVTZ	$0.6341_{0.5941}^{0.6773}$	$0.4198_{0.3943}^{0.4444}$	$0.5487_{0.4942}^{0.6069}$
DSD-BLYP-D3(BJ)/heavy-aug-cc-pVTZ	$0.4986_{0.4759}^{0.5211}$	$0.3316_{0.3130}^{0.3520}$	$0.3709_{0.3488}^{0.3951}$
B3LYP-D3(BJ)/DZVP	$0.5148_{0.4834}^{0.5475}$	$0.3230_{0.3016}^{0.3446}$	$0.3964_{0.3612}^{0.4301}$
B3LYP-D3MBJ/DZVP	$0.5332_{0.5021}^{0.5642}$	$0.3335_{0.3111}^{0.3570}$	$0.4040_{0.3728}^{0.4405}$
B3LYP-NL/DZVP	$0.5379_{0.5056}^{0.5686}$	$0.3370_{0.3157}^{0.3586}$	$0.4038_{0.3697}^{0.4351}$
B3LYP-D3(BJ)/def2-SV(P)	$1.0702_{1.0054}^{1.1362}$	$0.6691_{0.6270}^{0.7158}$	$0.7852_{0.7209}^{0.8528}$
B3LYP-D3(BJ)/def2-SVP	$1.1343_{1.0650}^{1.1950}$	$0.7160_{0.6699}^{0.7658}$	$0.8170_{0.7507}^{0.8811}$
B3LYP-D3(BJ)/def2-TZVP	$0.4866_{0.4596}^{0.5170}$	$0.3077_{0.2886}^{0.3276}$	$0.3922_{0.3625}^{0.4228}$
B3LYP-D3(BJ)/def2-TZVPD	$0.4609_{0.4320}^{0.4897}$	$0.2961_{0.2777}^{0.3142}$	$0.3853_{0.3549}^{0.4200}$
B3LYP-D3(BJ)/def2-TZVPP	$0.4940_{0.4616}^{0.5239}$	$0.3133_{0.2943}^{0.3329}$	$0.3974_{0.3661}^{0.4276}$
B3LYP-D3(BJ)/def2-TZVPPD	$0.4621_{0.4332}^{0.4925}$	$0.2966_{0.2775}^{0.3152}$	$0.3847_{0.3516}^{0.4184}$
B3LYP-D3(BJ)/def2-QZVP	$0.4603_{0.4323}^{0.4917}$	$0.2950_{0.2771}^{0.3136}$	$0.3813_{0.3511}^{0.4149}$
B3LYP-D3(BJ)/6-31G*	$0.8069_{0.7603}^{0.8560}$	$0.5071_{0.4776}^{0.5390}$	$0.5888_{0.5482}^{0.6305}$
B3LYP-D3(BJ)/6-31+G**	$0.5140_{0.4861}^{0.5424}$	$0.3321_{0.3134}^{0.3516}$	$0.4157_{0.3839}^{0.4485}$
B3LYP-D3(BJ)/6-311+G**	$0.4659_{0.4389}^{0.4948}$	$0.3014_{0.2830}^{0.3201}$	$0.3895_{0.3605}^{0.4193}$

[a]The 95% confidence intervals, calculated with cinnabar, are presented as subscripts and superscripts. Furthermore, the last column includes the RMSE within the thermally relevant low-energy region (TRLR) with energies of <5 kcal/mol, averaged over all of the molecules, which serves as a metric for assessing accuracy in low-energy regions. The best performer on this set of molecules is ωB97M-D3(BJ)/DZVP, and our current choice of theory level, B3LYP-D3(BJ)/DZVP, lags behind it by only 0.1 kcal/mol in RMSE.

functionals. Hence, MP2 geometries offer a more neutral and consistent basis for benchmarking devoid of the specific biases inherent to individual DFT functionals. For the complete basis set calculation, Helgaker's 2-point extrapolation scheme was used as implemented in Psi4.[58−60] Helgaker extrapolation scheme here includes a reference total energy from Hartree−Fock, correlation correction which includes correlation effects beyond HF with the MP2 method, and a delta correction, which gives a highly accurate correlation calculation with CCSD(T), accounting for the error in MP2. The difference between CCSD(T) and MP2 converges quickly with increase in basis functions and hence a smaller basis set can be used for this part of calculation. Psi4 performs these calculations in stages and the treatment follows this equation:[60]

$$E_{\text{total}}^{\text{CBS}} = \mathcal{F}_{\text{scf\_scheme}}(E_{\text{total,SCF}}^{\text{scf\_basis}}) + \mathcal{F}_{\text{corl\_scheme}}(E_{\text{corl,corl\_wfn}}^{\text{corl\_basis}}) + \delta_{\text{delta\_wfn\_lesser}}^{\text{delta\_wfn}} \tag{1}$$

where $\mathcal{F}$ is an energy or energy extrapolation scheme. And, in our case this translates to

$$E_{\text{total}}^{\text{CBS}} = E_{\text{HF}}^{\text{heavy-aug-cc-pVQZ}} + (E_{\text{MP2}}^{\text{heavy-aug-cc-pV[TQ]Z}} - E_{\text{HF}}^{\text{heavy-aug-cc-pV[TQ]Z}}) + (E_{\text{CCSD(T)}}^{\text{heavy-aug-cc-pVDZ}} - E_{\text{MP2}}^{\text{heavy-aug-cc-pVDZ}}) \tag{2}$$

Although heavy-aug-cc-pVDZ (or haDZ) is computationally affordable it may fall short of the gold standard reference level of theory often used in the community.[61,62] It is expected that the delta correction error with haDZ would fall somewhere in between 0.1 and 0.25 kcal/mol, the errors observed with aug-cc-

Table 2. Comparison between Neutral and Charged Molecule Subsets[a]

	RMSE (kcal/mol)		
specification	whole set	neutral subset	charged subset
B97-D3(BJ)/def2-TZVP	$0.5309_{0.4981}^{0.5591}$	$0.5278_{0.4853}^{0.5697}$	$0.5341_{0.4894}^{0.5826}$
M05-2X-D3/DZVP	$0.5089_{0.4770}^{0.5418}$	$0.5665_{0.5159}^{0.6154}$	$0.4414_{0.3929}^{0.4875}$
M06-2X-D3/DZVP	$0.4492_{0.4214}^{0.4778}$	$0.4982_{0.4554}^{0.5404}$	$0.3922_{0.3549}^{0.4314}$
M08-HX-D3/DZVP	$0.5197_{0.4888}^{0.5507}$	$0.5541_{0.5125}^{0.5942}$	$0.4815_{0.4399}^{0.5252}$
ωB97X-D3(BJ)/DZVP	$0.4408_{0.4125}^{0.4681}$	$0.4638_{0.4228}^{0.5080}$	$0.4156_{0.3783}^{0.4563}$
ωB97M-D3BJ/DZVP	$0.4141_{0.3880}^{0.4397}$	$0.4325_{0.3944}^{0.4698}$	$0.3941_{0.3538}^{0.4334}$
ωB97M-V/DZVP	$0.4219_{0.3935}^{0.4503}$	$0.4335_{0.3917}^{0.4722}$	$0.4096_{0.3688}^{0.4506}$
PW6B95-D3(BJ)/DZVP	$0.4341_{0.4086}^{0.4577}$	$0.4288_{0.3970}^{0.4608}$	$0.4394_{0.4008}^{0.4788}$
PW6B95-D3/DZVP	$0.4371_{0.4120}^{0.4621}$	$0.4503_{0.4160}^{0.4829}$	$0.4230_{0.3899}^{0.4585}$
MP2/aug-cc-pVTZ	$0.6947_{0.6487}^{0.7380}$	$0.6818_{0.6086}^{0.7586}$	$0.7079_{0.6595}^{0.7597}$
MP2/heavy-aug-cc-pVTZ	$0.6341_{0.5941}^{0.6773}$	$0.6382_{0.5692}^{0.7076}$	$0.6297_{0.5875}^{0.6673}$
DSD-BLYP-D3(BJ)/heavy-aug-cc-pVTZ	$0.4986_{0.4759}^{0.5211}$	$0.4735_{0.4415}^{0.5058}$	$0.5233_{0.4917}^{0.5530}$
B3LYP-D3(BJ)/DZVP	$0.5148_{0.4834}^{0.5475}$	$0.4461_{0.4088}^{0.4830}$	$0.5773_{0.5282}^{0.6230}$
B3LYP-D3MBJ/DZVP	$0.5332_{0.5021}^{0.5642}$	$0.4690_{0.4302}^{0.5052}$	$0.5924_{0.5451}^{0.6419}$
B3LYP-NL/DZVP	$0.5379_{0.5056}^{0.5686}$	$0.4605_{0.4268}^{0.4930}$	$0.6077_{0.5566}^{0.6555}$
B3LYP-D3(BJ)/def2-SV(P)	$1.0702_{1.0054}^{1.1362}$	$0.9048_{0.8330}^{0.9744}$	$1.2180_{1.1184}^{1.3157}$
B3LYP-D3(BJ)/def2-SVP	$1.1343_{1.0650}^{1.1950}$	$0.9492_{0.8777}^{1.0188}$	$1.2984_{1.1952}^{1.4037}$
B3LYP-D3(BJ)/def2-TZVP	$0.4866_{0.4596}^{0.5170}$	$0.4797_{0.4370}^{0.5197}$	$0.4938_{0.4527}^{0.5370}$
B3LYP-D3(BJ)/def2-TZVPD	$0.4609_{0.4320}^{0.4897}$	$0.4970_{0.4491}^{0.5450}$	$0.4203_{0.3857}^{0.4546}$
B3LYP-D3(BJ)/def2-TZVPP	$0.4940_{0.4616}^{0.5239}$	$0.4937_{0.4516}^{0.5399}$	$0.4944_{0.4522}^{0.5335}$
B3LYP-D3(BJ)/def2-TZVPPD	$0.4621_{0.4332}^{0.4925}$	$0.4978_{0.4537}^{0.5395}$	$0.4219_{0.3878}^{0.4584}$
B3LYP-D3(BJ)/def2-QZVP	$0.4603_{0.4323}^{0.4917}$	$0.4857_{0.4400}^{0.5305}$	$0.4325_{0.3964}^{0.4708}$
B3LYP-D3(BJ)/6-31G*	$0.8069_{0.7603}^{0.8560}$	$0.6701_{0.6209}^{0.7162}$	$0.9274_{0.8540}^{1.0008}$
B3LYP-D3(BJ)/6-31+G**	$0.5140_{0.4861}^{0.5424}$	$0.5298_{0.4818}^{0.5742}$	$0.4971_{0.4625}^{0.5331}$
B3LYP-D3(BJ)/6-311+G**	$0.4659_{0.4389}^{0.4948}$	$0.5019_{0.4558}^{0.5480}$	$0.4254_{0.3926}^{0.4569}$

[a]The overall RMSE in torsion profile energies of the molecule set with respect to the reference level CCSD(T)/[haTQZ; δ:haDZ]//MP2/heavy-aug-cc-pVTZ level of theory for the whole benchmark set and subsets of neutral and charged molecules. The 95% confidence intervals, calculated with cinnabar, are presented as subscripts and superscripts. The RMSE of neutral molecules for OpenFF's default theory level, B3LYP-D3(BJ)/DZVP, is comparable to the best functional.

pVDZ and cc-pVDZ on either end of the range.[63,64] To assess how good our choice of reference theory level is we have performed energy calculations with the gold standard reference theory level for a subset of 7 molecules from the benchmark set[61] at CCSD(T)/CBS, where for the extrapolation to CBS the correlation basis is aug-cc-pV[TQ]Z, and delta basis is aug-cc-pVTZ. The gold standard energy is calculated as,[60]

$$E_{\text{gold\_std}}^{\text{CBS}} = E_{\text{HF}}^{\text{aug−cc−pVQZ}} + \left(E_{\text{MP2}}^{\text{aug−cc−pV[TQ]Z}} - E_{\text{HF}}^{\text{aug−cc−pV[TQ]Z}}\right) + \left(E_{\text{CCSD(T)}}^{\text{aug−cc−pVTZ}} - E_{\text{MP2}}^{\text{aug−cc−pVTZ}}\right)$$

(3)

Our reference theory level employed a computationally efficient smaller basis set (haDZ) for delta correction when compared to a larger basis (aTZ) used in the gold standard reference. And, the RMSE in relative energies for the 7 molecule subset (7 × 24 grid points) of our reference theory with respect to the gold standard was $0.0761_{0.0609}^{0.0904}$ kcal/mol (the subscript and superscript are the 95% confidence intervals). So, the difference with respect to the gold standard was one tenths of a kcal/mol for our choice of reference theory level, which was accurate enough and quite affordable for our study.

2.2.2. Choice of Density Functionals to Benchmark. We have chosen a smaller pool of density functional approximations (DFA) that are cost-effective from the get go and we are not looking into a comprehensive evaluation of all available DFAs. The choice of DFAs include those commonly used in developing force fields and charge models, and better performing ones from other benchmark studies. From our prior studies DZVP has been proven cost friendly and all of the DFAs were tested with this basis set. And, within our current choice of DFA used in developing OpenFF force fields, B3LYP-D3BJ, we tested Ahlrichs def2 basis sets incrementing them systematically in size, and the Pople basis sets in increasing size as they are
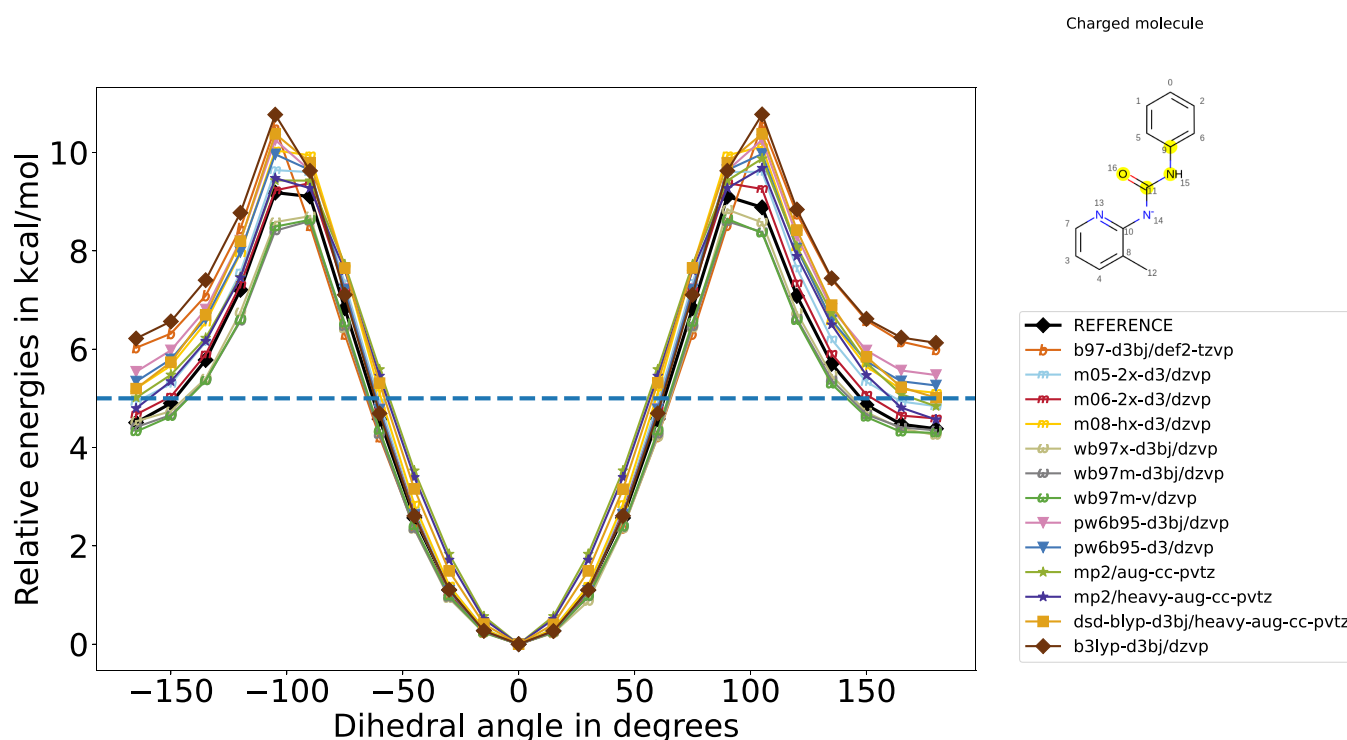
Figure 2. Torsion profile that corresponds to a molecule with a high RMSE at relative energies of >1.0 kcal/mol for most of the methods (row 23 from Table S1). The horizontal dashed line represents 5 kcal/mol, and if we truncate the energies above it, we can see that the low-energy regions were captured better than were the high-energy regions.

commonly used in practice in industry. Also, one of the purposes of the study was to determine which DFAs are useful at medium sized basis sets. The rankings and differences might vary from the benchmarks at larger basis sets. This selection represents different rungs on the Jacob's ladder of chemical accuracy.

- B97 (GGA), with D3(BJ) dispersion correction and with def2-TZVP basis set
- B3LYP hybrid family (hybrid-GGA), with[65]
  - D3(BJ) dispersion correction, and following basis sets[55,66−75]
    - DZVP
    - def2-SV(P)
    - def2-SVP
    - def2-TZVP
    - def2-QZVP
    - def2-TZVPP
    - def2-TZVPD
    - def2-TZVPPD
    - 6-31G*
    - 6-31+G**
    - 6-311+G**
  - D3MBJ[76] dispersion correction and DZVP basis
  - NL dispersion correction and DZVP basis
- Berkeley range separated functionals (meta-GGA),[77] with DZVP basis and dispersion corrections,
  - ωB97M-V, ωB97M-D3(BJ)
  - ωB97X-D3(BJ)
- Minnesota functionals (meta-GGA),[78−81] with DZVP basis and dispersion corrections
  - M05-2X-D3
  - M06-2X-D3
  - M08-HX-D3

- PW6B95 (meta-GGA),[82] with DZVP and dispersion corrections
  - D3(BJ)
  - D3
- Double hybrid DSD-BLYP-D3(BJ)[83] with haTZ basis

3. RESULTS AND DISCUSSION

3.1. Comparison of Different Levels of Theory. In the context of force field fitting, apart from conformer energies and optimized geometries, torsion energy profiles are another important source of molecular geometries/energetics data. The energy barriers observed in torsion energy profiles are a measure of the strength of steric hindrance or strong intramolecular interactions that prevent certain conformations. Thus, accurately capturing torsion profiles is relevant for understanding molecular recognition, binding, and other interactions that occur in complex systems. Our aim here is to pick an accurate and computationally efficient QM level of theory to train the valence parameters in a general small molecule force field. For this purpose, single point energies were evaluated at different levels of theory for comparison at the benchmark geometries, and the RMSE in relative energies was tabulated. Single point energies were evaluated at the same geometry to ensure parity between the methods since performing a geometry optimization with each of the methods will result in minor differences in final geometries, and sometimes TorsionDrive may push them to a completely different minima. SI Table S1 lists the RMSE in torsion profile energies for each of the molecules considered in this benchmark set with respect to the reference level of theory, CCSD(T)/ [haTQZ; δ:haDZ].

(a) All molecules   (b) Neutral subset   (c) Charged subset

(d) All molecules   (e) Neutral subset   (f) Charged subset
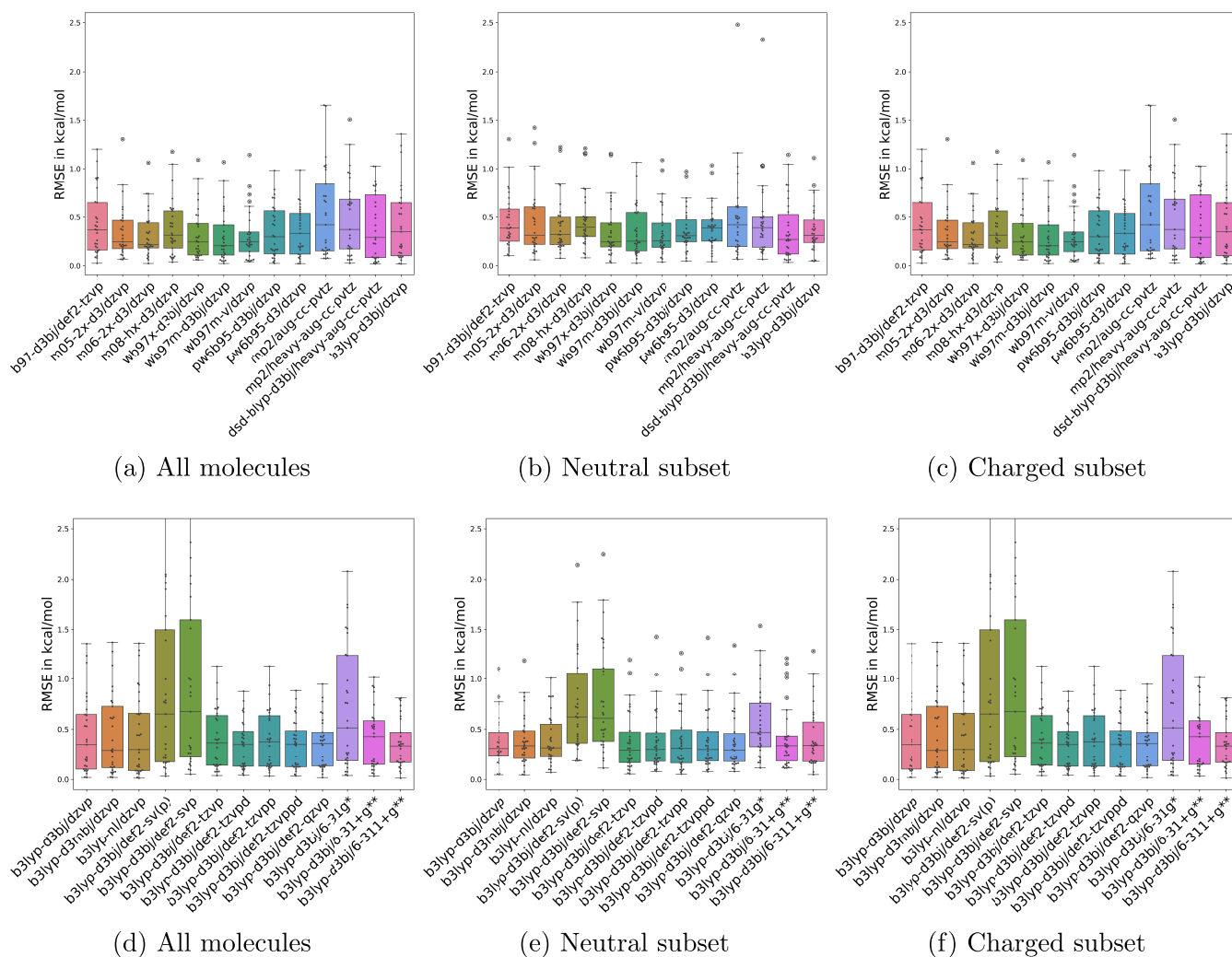
Figure 3. RMSE on the entire benchmark data set and on subsets of neutral and charged molecules. The top row shows the RMSE in torsion profile energies with different functional and basis set combinations. The bottom row shows the RMSE within the B3LYP functional.

The relative energies were calculated by subtracting the minimum energy for each of the methods. And, the RMSE was calculated with respect to the reference theory level.

$$E_{ref}(\mathbf{x}_i) = E'_{ref}(\mathbf{x}_i) - E'_{ref}(\mathbf{x}_0)$$

$$E_{method}(\mathbf{x}_i) = E'_{method}(\mathbf{x}_i) - E'_{method}(\mathbf{x}_0)$$

$$RMSE_{TRLR}(\mathbf{x}_{ref}, \mathbf{x}_{method}) = \sqrt{\frac{\sum_{i=1}^{n} (\mathbf{x}_{ref_i} - \mathbf{x}_{method_i})^2}{n}} \quad \forall \, i$$

$$\ni \mathbf{x}_{ref_i}$$

$$< 5 kcal/mol$$

(4)

where E' represents the absolute energies, x_0 represents the minimum energy point. And, the RMSE and MUE were evaluated with the relative energies. For calculating the RMSE of thermally relevant low-energy region (TRLR), only the relative energies below 5 kcal/mol on the reference energy surface were chosen. The cutoff of 5 kcal/mol was chosen to favor low-energy regions in state space.[3,28,84]

From Table 1 we can see that RMSE values of relative energies for most of the functional and basis set combinations studied

here fall in the range of 0.4−0.7 kcal/mol with respect to CCSD(T)/[haTQZ; δ:haDZ], the reference level of theory. ωB97M-D3(BJ)/DZVP is the most accurate among those tested here with an RMSE in torsion profile energies of 0.41 kcal/mol. This range-separated hybrid functional has been one of the top performers in several other recent studies that were done with larger basis sets.[43,44] Our current choice of theory, B3LYP-D3(BJ)/DZVP, has an RMSE of 0.52 kcal/mol and it closely trails by a difference of 0.1 kcal/mol when compared to the best functional in this study. Almost half of the of the other tested levels of theory were within 0.1 kcal/mol of the best functional. No method is consistently the most accurate across all of the subsets, highlighting the importance of considering a large and diverse benchmark set.

The RMSEs on the whole set and the subsets of neutral and charged molecule sets were depicted in Figure 3 and tabulated in Table 2. We can see from Figure 3(f), and from Table 2, that the accuracy of B3LYP-D3(BJ) functional increases for charged molecules with addition of polarization and diffuse functions from DZVP to def2-TZVP and higher. Within B3LYP-D3(BJ) it is to be noted that the high quality DZVP basis set yields comparable results to the larger triple-ζ level basis set, def2-TZVP. PW6B95-D3(BJ)/DZVP is the best among the tested methods for neutral molecules with an RMSE of 0.43 kcal/mol.
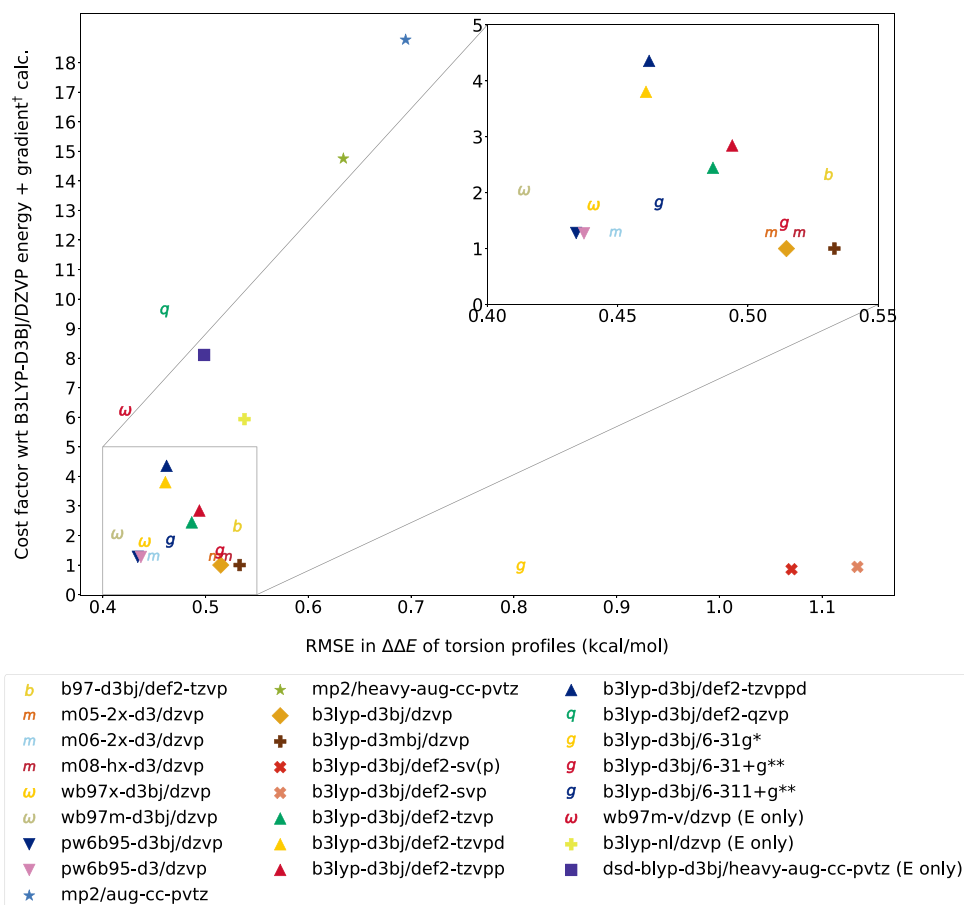
Figure 4. Cost factor of a single point energy and gradient calculations within each method, averaged over 10 runs of different conformations of the same molecule (with 16 heavy atoms) for all of the methods. All of the run times of each method were normalized with the time taken for a B3LYP-D3(BJ)/DZVP calculation. All of the calculations were run on an Intel(R) Xeon(R) Gold 6148 CPU @ 2.40 GHz CPU using 16 threads and 48 GB of memory. The best functional, ωB97M-D3(BJ)/DZVP, is indicated with an $\boldsymbol{\omega}$, while OpenFF's default, B3LYP-D3(BJ)/DZVP, is indicated with a ◆.

ωB97M-D3(BJ)/DZVP is also on the same level with statistically insignificant difference. And, B3LYP-D3(BJ)/DZVP, has an RMSE of 0.45 kcal/mol and it closely trails these two functionals by a negligible difference of 0.02 kcal/mol for neutral molecules.

Some of the larger errors were seen in the subset of charged molecules, when compared to the neutral molecules, among the whole set of molecules used in this benchmark. ωB97M-D3(BJ)/DZVP is the best among the tested methods for the charged subset of molecules. And, OpenFF's default level of theory, B3LYP-D3(BJ)/DZVP, has an RMSE error 0.18 kcal/mol worse than the best method, with an overall RMSE of 0.58 kcal/mol. The difference in RMSE between charged molecules and the neutral molecules is slightly higher for B3LYP-D3(BJ)/DZVP level of theory, and addition of basis functions helped drive this error down. The addition of basis functions from DZVP to def2-TZVPD basis set shows a significant improvement, with B3LYP-D3(BJ) functional, and can be a choice for fine-tuning the force field with more accurate data.

Despite observing higher errors in certain molecules, the methodology used to construct the torsion profile target data, which prioritizes the match to low energy regions, has the potential to mitigate some of these discrepancies. During the training of OpenFF force fields we construct a torsion profile target and optimize the force field using ForceBalance,[85,86] and the objective function in terms of relative energies is defined as follows:[3]

$$E_{QM}(\mathbf{x}_i) = E'_{QM}(\mathbf{x}_i) - E'_{QM}(\mathbf{x}_0)$$

$$E_{MM}(\mathbf{x}_i) = E'_{MM}(\text{OptMM}(\mathbf{x}_i; \theta)) - E'_{MM}(\text{OptMM}(\mathbf{x}_0; \theta)) \quad (5)$$

where the primes indicate the absolute energies at each grid point and the weighted differences in relative energy profiles serve as the loss function to minimize:

$$L(\theta) = \left(1/d_E^2\right)\left\{\left[\sum_{i \in N(\text{gridpoints})} w(E_{QM}(\mathbf{x}_i))\right.\right.$$

$$\left. (E_{QM}(\mathbf{x}_i) - E_{MM}(\mathbf{x}_i; \theta))^2\right]$$

$$\left. \left/ \left[\sum_{i \in N(\text{gridpoints})} w(E_{QM}(\mathbf{x}_i))\right]\right\}\right. \quad (6)$$

where $\mathbf{x}_i$ represents the coordinates of ith conformer, the 0th conformer is the minimum energy conformer in respective potential energy landscapes, θ is the force field parameter set at that iteration, and $\text{OptMM}(\mathbf{x}_i, \theta)$ corresponds to the MM energy obtained via constrained minimization and $d_E = 1$ kcal/mol is a conversion factor to make the sum over deviations dimensionless. The applied weights $w(E_{QM})$ in eq 6 prioritize matching the

Table 3. Cost Factors of a Single Energy + Gradient Calculation of Various Methods with Respect to a B3LYP-D3(BJ) Calculation[a]

specification	mean wall time ± standard deviation (s)	cost factor
B97-D3(BJ)/def2-TZVP	49.41 ± 1.10	2.33
M05-2X-D3/DZVP	27.05 ± 0.42	1.28
M06-2X-D3/DZVP	27.26 ± 0.35	1.29
M08-HX-D3/DZVP	27.06 ± 0.49	1.28
ωB97X-D3(BJ)/DZVP	37.49 ± 0.84	1.77
ωB97M-D3(BJ)/DZVP	43.17 ± 0.82	2.04
PW6B95-D3(BJ)/DZVP	27.11 ± 0.39	1.28
PW6B95-D3/DZVP	26.98 ± 0.33	1.27
MP2/aug-cc-pVTZ	397.87 ± 15.91	18.78
MP2/heavy-aug-cc-pVTZ	312.62 ± 12.16	14.75
B3LYP-D3(BJ)/DZVP	21.19 ± 0.24	1.00
B3LYP-D3MBJ/DZVP	21.21 ± 0.21	1.00
B3LYP-D3(BJ)/def2-SV(P)	18.21 ± 0.36	0.86
B3LYP-D3(BJ)/def2-SVP	19.89 ± 0.17	0.94
B3LYP-D3(BJ)/def2-TZVP	51.80 ± 0.63	2.44
B3LYP-D3(BJ)/def2-TZVPD	80.60 ± 0.63	3.80
B3LYP-D3(BJ)/def2-TZVPP	60.27 ± 0.95	2.84
B3LYP-D3(BJ)/def2-TZVPPD	92.33 ± 1.18	4.36
B3LYP-D3(BJ)/def2-QZVP	203.17 ± 3.71	9.59
B3LYP-D3(BJ)/6-31G*	19.41 ± 0.32	0.92
B3LYP-D3(BJ)/6-31+G**	30.22 ± 0.49	1.43
B3LYP-D3(BJ)/6-311+G**	37.98 ± 0.46	1.79
Energy Calculation Only		
ωB97M-V/DZVP	66.97 ± 0.61	6.21
B3LYP-NL/DZVP	64.02 ± 0.47	5.94
DSD-BLYP-D3(BJ)/heavy-aug-cc-pVTZ	87.43 ± 1.13	8.11
B3LYP-D3(BJ)/DZVP	10.78 ± 0.17	1.00

[a]All of the wall times were averaged over 10 different runs of 10 different conformers of the same molecule, molecule 1 in Table 1, which contains 16 heavy atoms. For a few methods for which timings for an energy calculation were reported, they were scaled with respect to the cost of a B3LYP-D3(BJ) energy calculation.

torsion profile at energy minima since Boltzmann sampling favors low-energy regions in state space.[84]

$$w(E_{QM}) = \begin{cases} 1 & E_{QM} < 1.0 \text{ kcal/mol} \\ (1 + (E_{QM} - 1)^2)^{-1/2} & 1.0 \le E_{QM} < 5.0 \text{ kcal/mol} \\ 0 & E_{QM} \ge 5.0 \text{ kcal/mol} \end{cases}$$

$$(7)$$

In the context of fitting OpenFF force fields, the torsion profile energy loss function defined above in eq 6 may further mitigate the differences between the methods since we applied a hard cutoff of 5 kcal/mol to exclude the higher energy regions from torsion fits for the OpenFF Parsley and Sage line of force fields.[3,28]

The RMSE on thermally relevant low-energy region (TRLR), only considering the energies less than 5 kcal/mol with respect to the minima on the torsion profile, were tabulated in the last column of Table 1. We can see that ωB97M-D3(BJ)/DZVP still holds its place as the most accurate functional with torsion target score as well, and the differences between various methods drop drastically in the low energy regions. In a sense, the modeling of low-energy, thermally relevant regions was quite accurate, while it was in the high-energy regions where the influence of

stereoelectronic and steric effects became prominent, leading to discernible differences between various methods.

The error in the subset of neutral molecules (30 data points) is small for OpenFF's default of B3LYP-D3(BJ) and close to the best functional, ωB97M-D3(BJ). However, the error is larger for the charged subset of molecules (29 data points) with OpenFF's default, compared to the best functional, and yet remains accurate in low-energy regions. Most of the large deviations come from high energy regions which were (and usually are) excluded in fitting to torsion profile energies as they are thermally irrelevant. In this sense, our default method may remain appealing given its low computational cost and relatively low error in key regions of torsion profiles. For charged molecules, addition of more basis functions would help as seen in Figure 3, subplot (f).

Referring to SI Table S1, when we examine the molecule with the largest error using OpenFF's default, B3LYP-D3(BJ)/DZVP, which exhibits an RMSE in relative energies of 1.36 kcal/mol, we observe that the majority of discrepancies arise from the high-energy region exceeding 5 kcal/mol. The torsion profile for this molecule with respect to different basis sets was shown in Figure 2. It is worth reiterating that we exclude this high-energy region during the training process.

3.2. Comparison of Timings and Accuracy. The computation time of a single point (energy + gradient) calculation provides a rough approximation of the method's cost for a torsion scan or geometry optimization. When examining the OpenFF QM data sets of small molecules, it is observed that a geometry optimization typically requires a median of 42 gradient calculations (based on data from approximately 8000 geometry optimizations). Additionally, a 1D torsion scan with 24 grid points generally costs around 788 gradient calculations (from data on roughly 4000 torsion scans). As a rule of thumb, a geometry optimization calculation is 40x costlier than a single point energy and gradient calculation, and a 1D torsion scan is 800x costlier if executed serially, or nearly 72-fold for parallelized torsion scans (considering the median of maximum number of optimization steps taken among all grid points). So, the differences scale up pretty quickly with the type of calculation. To provide a reference point, we present timing data for a molecule containing 16 heavy atoms. The timings were normalized with the time for a B3LYP-D3(BJ)/DZVP calculation. The cost factor versus RMSE plot is shown in Figure 4, and the most accurate functional is almost twice the cost of a B3LYP-D3(BJ)/DZVP calculation, despite only a modest accuracy benefit. It is to be noted that Psi4, as of v1.4.1, does not yet have analytic gradients for NL and VV10 dispersion terms, and also for DSD-BLYP method, so for these methods only the costs of an energy calculation were reported and scaled with respect to the cost of a B3LYP-D3(BJ) energy calculation (Table 3).

B3LYP-D3(BJ) with def2-SVP and def2-SV(P) were lower in accuracy but cheaper, and by adding more basis functions to improve accuracy the cost increases, and it can be seen from Figure 4 that Salahub's DZVP offers a perfect balance in cost and accuracy for this functional. In the basis set study using B3LYP, 6-311+G** is already very close to the basis set limit (def2-QZVP) in terms of accuracy, in fact it is closer than def2-TZVP and def2-TZVPP, yet it is much cheaper than all of those. When comparing the Pople basis sets, perhaps the largest improvement in accuracy for B3LYP came from adding the first set of diffuse functions (from 6-31g* to 6-31+g*). In the Minnesota functionals, M06-2X-D3 performs better than M05-2X-D3
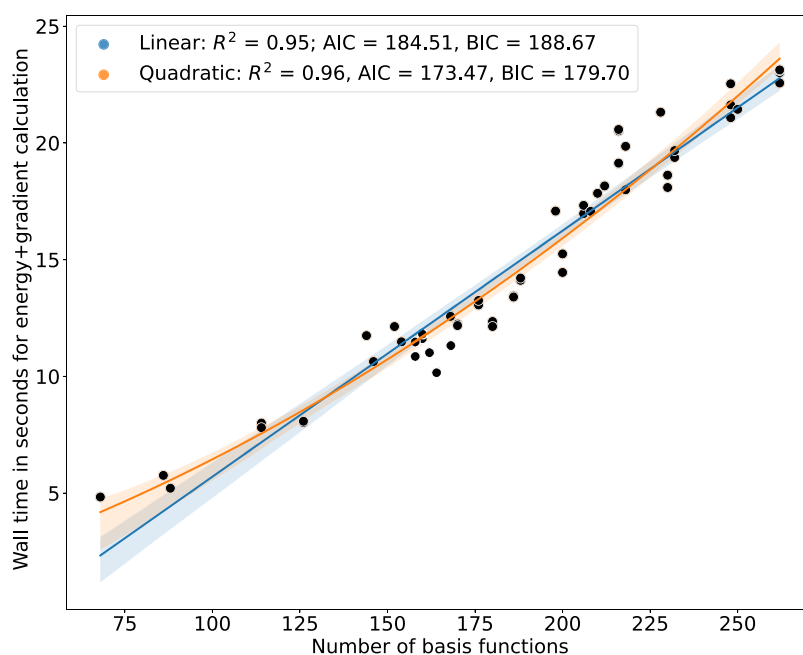
Figure 5. Scaling of B3LYP-D3(BJ)/DZVP single point energy plus gradient calculation with an increase in the number of basis functions, which in turn represents an increase in the number of atoms. The time of calculation for the dispersion correction term is not included as it is negligible, on the order of a fraction of a second. The quadratic (orange) and linear (blue) fits to the data are shown along with the metrics R^2, Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The quadratic scaling model seems appropriate from these metrics.

Table 4. Comparison of SQM, DFT-3c, and ML Methods on This Benchmark Set[a]

specification	RMSE (kcal/mol)		
	whole set	neutral subset	charged subset
GFN2-XTB	$1.2437_{1.1742}^{1.3163}$	$1.3367_{1.2276}^{1.4589}$	$1.1395_{1.0604}^{1.2106}$
AIMNET2	$0.5518_{0.5176}^{0.5891}$	$0.5737_{0.5343}^{0.6157}$	$0.5281_{0.4697}^{0.5877}$
MACE-OFF23	$1.1013_{1.0049}^{1.2011}$	$0.4061_{0.3778}^{0.4336}$	$1.5156_{1.3669}^{1.6542}$
HF-3c	$1.8223_{1.7147}^{1.9320}$	$1.6124_{1.4597}^{1.7677}$	$2.0166_{1.8458}^{2.1811}$

[a]The overall RMSE in torsion profile energies of the molecule set with respect to the reference level CCSD(T)/[haTQZ; δ:haDZ]// MP2/heavy-aug-cc-pVTZ level of theory for the whole benchmark set and subsets of neutral and charged molecules. The 95% confidence intervals, calculated with cinnabar, are presented as subscripts and superscripts.

and M08-HX-D3. In both the B3LYP and ωB97 M tests, the ab initio nonlocal dispersion correction actually gives larger RMSEs than the empirical dispersion.

Psi4's efficient implementation results in quadratic scaling of DFT calculation time with an increase in molecule size.[87] Both the quadratic and linear fits of the data yielded a similar R^2, which prompted considering the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to differentiate the models better. Higher R^2, lower AIC, and lower BIC values indicate a better fit to the quadratic model. From Figure 5 we can see that with an increase in number of atoms represented by an increase in number of basis functions and the scaling of computation cost for a B3LYP-D3(BJ)/DZVP single point energy plus gradient calculation.

3.3. Faster Parametrization with SQM/ML Potentials.
Creating bespoke force fields on-the-fly adds a lot of new torsion parameters, which in turn requires generation of new QM reference data for training them.[88] Generation of new QM data

is time-consuming, and a possible alternative for faster parametrization, without too much loss in accuracy, can be a semiempirical method, or a couterpoise corrected method with minimal basis, or machine learning potentials.[21,89−96] Here we checked the performance of the semiempirical method GFN2-XTB,[95] Grimme's 3-corrected Hartree−Fock method (HF-3c),[93] and two recent machine learning potentials, AIMNET2[96] and MACE-OFF23,[21] the ML potentials which have demonstrated accuracy very close to the level of DFT that they were trained on. Only single point energies were evaluated here, as in the comparisons above.

Table 4 summarizes the performance of the four methods across the whole benchmark set, and also broken down into neutral and charged subsets (see also Table 2). GFN2-XTB shows reasonable accuracy, but is not competitive with the DFT methods. AIMNet2 shows very good accuracy, close to DFT, consistently across both neutral and charged species. MACE-OFF23 shows remarkable accuracy for the neutral subset, in fact more accurate than any of the DFT methods studied here. Note that MACE-OFF23 was not trained on charged species, and hence the error is much higher for this subset, but this will be addressed in future models. HF-3c does not perform well here. Thus, machine learning potentials such as these are a reasonable alternative to DFT, particularly for high-throughput, bespoke parametrization work.[88]

4. CONCLUSION

We conducted a benchmark of QM levels of theory that strike a balance between accuracy and computational efficiency for generating large QM data sets with diverse chemistries to train the valence parameters in a general small molecule force field. The benchmark set of molecules included charged molecules, biaryls, complex hypervalent sulfur chemistry, and complex nitrogen chemistry. This benchmark study is an extension to an earlier work[2] on benchmarking conformer energies, which

suggested B3LYP-D3(BJ)/DZVP as the level of theory to generate OpenFF force field training and validation data. In the context of force field development, aside from conformer energies and optimized geometries, torsion energy profiles represent another indispensable source of molecular interaction data. Achieving accuracy in torsion profile energies relative to a highly accurate QM level of theory reflects in the trained force field. For this purpose, relative energies were compared against CCSD(T)/[haTQZ; δ:haDZ]//MP2/haTZ level of theory for different functional and basis set combinations. And, among the tested methods, ωB97M-D3(BJ) outperforms the others even within a smaller basis set of DZVP, boasting an RMSE in torsion profile energies of just 0.41 kcal/mol. This range-separated hybrid functional has consistently ranked among the top performers in various recent studies[43,44] which were done with a larger basis set. OpenFF's choice of B3LYP-D3(BJ)/DZVP closely follows it with an RMSE in relative energies of 0.52 kcal/mol. And, the computational cost of B3LYP-D3(BJ)/DZVP for a single gradient is only half of the best functional, but depending on the type of data set, geometry optimization or torsion scans, the cost would scale up with the number of steps taken during the calculation. Within a subset of neutral molecules the RMSE in relative energies with B3LYP-D3(BJ)/DZVP is comparable to the most accurate method. And, the larger errors appear to originate from molecules with charges. However, in practice, the distinctions between levels of theory become evident in the high-energy regions, which are typically excluded during force field training with torsion profiles as they are thermally irrelevant.

■ ASSOCIATED CONTENT

Data Availability Statement

The scripts to download the structure data from QCArchive and the downloaded energy and dipole data are available at https://github.com/MobleyLab/qm-theory-benchmark/tree/master.

Ⓢ Supporting Information

Supporting Information includes tables on individual errors in each molecule, details on compute infrastructure used, dipole moments comparison, correlation between energy and delta energy difference wrt reference, error estimates in barriers. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.4c03167.

(PDF)

■ AUTHOR INFORMATION

Corresponding Author

Pavan Kumar Behara — *Center for Neurotherapeutics, University of California, Irvine, California 92697, United States;* ⓘ orcid.org/0000-0001-6583-2148; Email: pbehara@uci.edu

Authors

Hyesu Jang — *Chemistry Department, University of California at Davis, Davis, California 95616, United States; OpenEye Scientific Software, Santa Fe, New Mexico 87508, United States*

Joshua T. Horton — *School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.*

Trevor Gokey — *Department of Chemistry, University of California, Irvine, California 92697, United States*

David L. Dotson — *The Open Force Field Initiative, Open Molecular Software Foundation, Davis, California 95616, United States; Datryllic LLC, Phoenix, Arizona 85003, United States*

Simon Boothroyd — *Boothroyd Scientific Consulting Ltd., London WC2H 9JQ, U.K.*

Christopher I. Bayly — *OpenEye Scientific Software, Santa Fe, New Mexico 87508, United States*

Daniel J. Cole — *School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.*

Lee-Ping Wang — *Chemistry Department, University of California at Davis, Davis, California 95616, United States;* ⓘ orcid.org/0000-0003-3072-9946

David L. Mobley — *Center for Neurotherapeutics, University of California, Irvine, California 92697, United States; Department of Chemistry, University of California, Irvine, California 92697, United States;* ⓘ orcid.org/0000-0002-1083-5533

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.4c03167

Notes

The authors declare the following competing financial interest(s): D.L.M. serves on the scientific advisory boards of Anagenex and OpenEye Scientific Software, Cadence Molecular Sciences, and is an Open Science Fellow with Psivant.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(2) Wang, L. P. Parameterization perspective I: Parameter optimization methodology. 2019. https://zenodo.org/records/3405539 (last accessed 2024-02-12).

(3) Boothroyd, S.; Behara, P. K.; Madin, O. C.; Hahn, D. F.; Jang, H.; Gapsys, V.; Wagner, J. R.; Horton, J. T.; Dotson, D. L.; Thompson, M. W.; Maat, J.; Gokey, T.; Wang, L.-P.; Cole, D. J.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Shirts, M. R.; Mobley, D. L. Development and Benchmarking of Open Force Field 2.0.0: The Sage Small Molecule Force Field. *J. Chem. Theory Comput.* **2023**, *19*, 3251−3275.

(4) 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.

(5) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(6) Maerzke, K. A.; Schultz, N. E.; Ross, R. B.; Siepmann, J. I. TraPPE-UA Force Field for Acrylates and Monte Carlo Simulations for Their Mixtures with Alkanes and Alcohols. *J. Phys. Chem. B* **2009**, *113*, 6415−6425.

(7) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. 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.

(8) Bayly, C. I.; McKay, D.; Truchon, J.-F. An Informal AMBER Small Molecule Force Field: parm@Frosst. http://www.ccl.net/cca/data/parm_at_Frosst/ (last accessed 2024-01-04).

(9) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023−4039.

(10) 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.

(11) 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.

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

(13) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291−4300.

(14) 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.

(15) 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.

(16) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y.; De Fabritiis, G.; Markland, T. E. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *Sci. Data* **2023**, *10*, 11.

(17) 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.

(18) Takaba, K.; Pulido, I.; Behara, P. K.; Cavender, C. E.; Friedman, A. J.; Henry, M. M.; Opeskin, H. M.; Iacovella, C. R.; Nagle, A. M.; Payne, A. M.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Wang, Y. Machine-learned molecular mechanics force field for the simulation of protein-ligand systems and beyond. *arXiv* **2023**, DOI: 10.48550/arXiv.2307.07085.

(19) Lehner, M. T.; Katzberger, P.; Maeder, N.; Schiebroek, C. C.; Teetz, J.; Landrum, G. A.; Riniker, S. DASH: Dynamic Attention-Based Substructure Hierarchy for Partial Charge Assignment. *J. Chem. Inf. Model.* **2023**, *63*, 6014−6028.

(20) Isert, C.; Atz, K.; Jiménez-Luna, J.; Schneider, G. QMugs, quantum mechanical properties of drug-like molecules. *Sci. Data* **2022**, *9*, 273.

(21) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Kapil, V.; Witt, W. C.; Magdău, I.-B.; Cole, D. J.; Csányi, G. MACE-OFF23: Transferable Machine Learning Force Fields for Organic Molecules. *arXiv* **2023**, DOI: 10.48550/arXiv.2312.15211.

(22) Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. Model.* **2017**, *57*, 1265−1275.

(23) Ehrlich, S.; Göller, A. H.; Grimme, S. Towards full Quantum-Mechanics-based Protein−Ligand Binding Affinities. *ChemPhysChem* **2017**, *18*, 898−905.

(24) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969−1981.

(25) Lehtola, S.; Dimitrova, M.; Fliegl, H.; Sundholm, D. Benchmarking Magnetizabilities with Recent Density Functionals. *J. Chem. Theory Comput.* **2021**, *17*, 1457−1468.

(26) Zhou, A.; Schauperl, M.; Nerenberg, P. S. Benchmarking Electronic Structure Methods for Accurate Fixed-Charge Electrostatic Models. *J. Chem. Inf. Model.* **2020**, *60*, 249−258.

(27) Hickey, A. L.; Rowley, C. N. Benchmarking Quantum Chemical Methods for the Calculation of Molecular Dipole Moments and Polarizabilities. *J. Phys. Chem. A* **2014**, *118*, 3678−3687.

(28) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.

(29) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. J. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(30) 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.

(31) Rai, B. K.; Sresht, V.; Yang, Q.; Unwalla, R.; Tu, M.; Mathiowetz, A. M.; Bakken, G. A. TorsionNet: A Deep Neural Network to Rapidly Predict Small-Molecule Torsional Energy Profiles with the Accuracy of Quantum Mechanics. *J. Chem. Inf. Model.* **2022**, *62*, 785−800.

(32) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R.; Miller, T. F. OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *J. Chem. Phys.* **2021**, *155*, 204103.

(33) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2001**, *577*, 1−20.

(34) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(35) Lee, K.; Murray, D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B* **2010**, *82*, 081101.

(36) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103.

(37) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105−5154.

(38) 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.

(39) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(40) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

(41) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn−Sham global-hybrid exchange−correlation density functional with broad accuracy for multi-reference and single-reference systems and non-covalent interactions. *Chem. Sci.* **2016**, *7*, 5032−5051.

(42) 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.

(43) 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.

(44) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, 115, 2315−2372.

(45) Mehta, N.; Fellowes, T.; White, J. M.; Goerigk, L. CHAL336 Benchmark Set: How Well Do Quantum-Chemical Methods Describe Chalcogen-Bonding Interactions? *J. Chem. Theory Comput.* **2021**, 17, 2783−2806.

(46) 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.

(47) Yuan, Y.; Mills, M. J. L.; Popelier, P. L. A.; Jensen, F. Comprehensive Analysis of Energy Minima of the 20 Natural Amino Acids. *J. Phys. Chem. A* **2014**, 118, 7876−7891.

(48) 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.

(49) 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.

(50) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, 16, 2355−2368.

(51) Folmsbee, D.; Hutchison, G. Assessing conformer energies using electronic structure and machine learning methods. *Int. J. Quantum Chem.* **2021**, 121, No. e26381.

(52) Siiskonen, A.; Priimagi, A. Benchmarking DFT methods with small basis sets for the calculation of halogen-bond strengths. *J. Mol. Model.* **2017**, 23, 50.

(53) Nam, S.; Cho, E.; Sim, E.; Burke, K. Explaining and Fixing DFT Failures for Torsional Barriers. *J. Phys. Chem. Lett.* **2021**, 12, 2796−2804.

(54) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, 12, 444−454.

(55) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chem.* **1992**, 70, 560−571.

(56) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 1273−1280.

(57) Qiu, Y.; Smith, D. G. A.; Stern, C. D.; Feng, M.; Jang, H.; Wang, L.-P. Driving torsion scans with wavefront propagation. *J. Chem. Phys.* **2020**, 152, 244116.

(58) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, 106, 9639−9646.

(59) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-set convergence of the energy in molecular Hartree−Fock calculations. *Chem. Phys. Lett.* **1999**, 302, 437−446.

(60) Burns, L. A.; Smith, D. G. A.; Kraus, P. Complete Basis Set. 2023. https://psicode.org/psi4manual/master/cbs.html.

(61) 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.

(62) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, 151, 070901.

(63) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta(MP_2)(CCSD(T))$: 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.

(64) Marshall, M. S.; Sears, J. S.; Burns, L. A.; Brédas, J.-L.; Sherrill, C. D. An Error and Efficiency Analysis of Approximations to Møller Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2010**, 6, 3681−3687.

(65) Becke, A. D. A new mixing of Hartree−Fock and local density-functional theories. *J. Chem. Phys.* **1993**, 98, 1372−1377.

(66) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* basis set for third-row atoms. *J. Comput. Chem.* **2001**, 22, 976−984.

(67) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, 77, 3654−3665.

(68) Dill, J. D.; Pople, J. A. Self-consistent molecular orbital methods. XV. Extended Gaussian-type basis sets for lithium, beryllium, and boron. *J. Chem. Phys.* **1975**, 62, 2921−2923.

(69) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, 56, 2257−2261.

(70) Hellweg, A.; Rappoport, D. Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations. *Phys. Chem. Chem. Phys.* **2015**, 17, 1010−1017.

(71) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, 7, 3297−3305.

(72) Dunning, T. H., Jr Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, 90, 1007−1023.

(73) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H., Jr Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.* **1999**, 110, 7667−7676.

(74) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H−Kr. *J. Chem. Phys.* **2003**, 119, 12753−12762.

(75) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, 59, 4814−4820.

(76) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, 7, 2197−2203.

(77) Mardirossian, N.; Head-Gordon, M. B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, 144, 214110.

(78) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, 2, 364−382.

(79) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J. Chem. Phys.* **2005**, 123, 161103.

(80) Zhao, Y.; Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2008**, 4, 1849−1868.

(81) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new

functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(82) Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.

(83) Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *J. Phys. Chem. C* **2010**, *114*, 20801−20808.

(84) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.

(85) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.

(86) 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.

(87) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F., III; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E., III; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.

(88) Horton, J. T.; Boothroyd, S.; Wagner, J.; Mitchell, J. A.; Gokey, T.; Dotson, D. L.; Behara, P. K.; Ramaswamy, V. K.; Mackey, M.; Chodera, J. D.; Anwar, J.; Mobley, D. L.; Cole, D. J. Open Force Field BespokeFit: Automating Bespoke Torsion Parametrization at Scale. *J. Chem. Inf. Model.* **2022**, *62*, 5622−5633.

(89) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(90) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.

(91) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.

(92) Witte, J.; Neaton, J. B.; Head-Gordon, M. Effective empirical corrections for basis set superposition error in the def2-SVPD basis: gCP and DFT-C. *J. Chem. Phys.* **2017**, *146*, 234105.

(93) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.

(94) Müller, M.; Hansen, A.; Grimme, S. ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double- basis set. *J. Chem. Phys.* **2023**, *158*, 014103.

(95) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(96) Anstine, D.; Zubatyuk, R.; Isayev, O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *chemRxiv* **2023**, DOI: 10.26434/chemrxiv-2023-296ch.