RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2019, **9**, 36492

A collection of forcefield precursors for metal–organic frameworks†

Taoyi Chen and Thomas A. Manz *

A host of important performance properties for metal–organic frameworks (MOFs) and other complex materials can be calculated by modeling statistical ensembles. The principle challenge is to develop accurate and computationally efficient interaction models for these simulations. Two major approaches are (i) *ab initio* molecular dynamics in which the interaction model is provided by an exchange–correlation theory (*e.g.*, DFT + dispersion functional) and (ii) molecular mechanics in which the interaction model is a parameterized classical force field. The first approach requires further development to improve computational speed. The second approach requires further development to automate accurate forcefield parameterization. Because of the extreme chemical diversity across thousands of MOF structures, this problem is still mostly unsolved today. For example, here we show structures in the 2014 CoRE MOF database contain more than 8 thousand different atom types based on first and second neighbors. Our results showed that atom types based on both first and second neighbors adequately capture the chemical environment, but atom types based on only first neighbors do not. For 3056 MOFs, we used density functional theory (DFT) followed by DDEC6 atomic population analysis to extract a host of important forcefield precursors: partial atomic charges; atom-in-material (AIM) C_6, C_8, and C_{10} dispersion coefficients; AIM dipole and quadrupole moments; various AIM polarizabilities; quantum Drude oscillator parameters; AIM electron cloud parameters; *etc.* Electrostatic parameters were validated through comparisons to the DFT-computed electrostatic potential. These forcefield precursors should find widespread applications to developing MOF force fields.

Received 11th September 2019
Accepted 25th October 2019

DOI: 10.1039/c9ra07327b

rsc.li/rsc-advances

1. Introduction

Metal–organic frameworks (MOFs) are a kind of coordination network comprised of metal atoms connected by organic linkers.[1] In this work, we are interested in MOFs that are porous crystals. Because of their nanoporous structures, these materials attract much interest for gas storage, gas separation, catalysis, and other applications.[2–6] Many thousands of MOF crystal structures have been deposited in the Cambridge Structural Database (CSD).[1,7,8,115–117] Most often, these crystal structures were measured using X-ray diffraction crystallography (XRDC). Since hydrogen atoms contain no core electrons, they diffract X-rays extremely weakly.[9] This makes it much harder to refine hydrogen atom positions than positions for heavier atoms.[9–11] Consequently, hydrogen atom positions may be unresolved in some of the reported crystal structures. Other complications include the presence of disordered atoms, solvent molecules, and/or free ions in some of the reported MOF crystal structures.

In 2014, Chung *et al.* reported a Computation Ready Experimental (CoRE) MOF database that was constructed by first searching the CSD to identify MOFs and then partially cleaning these structures.[7] The searching step was designed to identify structures containing metal atoms bonded to non-metal atoms that form 3-dimensional networks. The cleaning process was intended to fix or discard structures containing disordered atoms and partial occupancies. The cleaning process was also intended to remove solvent molecules and other small adsorbates in the MOF's pores but to retain charge-balancing ions. Finally, missing hydrogen atoms were added to some of the structures. However, this cleaning process was imperfect resulting in some structures with errors.[12,13,115] This 2014 CoRE

Department of Chemical & Materials Engineering, New Mexico State University, Las Cruces, New Mexico, 88003-8001, USA. E-mail: tmanz@nmsu.edu

† Electronic supplementary information (ESI) available: A 7-zip format archive containing: classification of 5109 individual MOFs into calculated, uncalculated, and CSD groups with a list of rejected and non-rejected structures in each group; calculated forcefield precursors for 3056 MOFs; atom types with *XYZ* coordinates for 3056 calculated structures and 256 uncalculated structures; lists of individual atom type frequencies for calculated, uncalculated, and CSD structures; list of rejected structures and the itemized reason(s); means and standard deviations of forcefield precursors based on first-neighbor atom typing; means and standard deviations of forcefield precursors based on second-neighbor atom typing; list of atom types contained in each MOF for all non-rejected structures (calculated, uncalculated, and CSD); list of atom types appearing in uncalculated and CSD structures that did not appear in any calculated structures; list of atom types that appeared in only one calculated structure; list of atom types, NAC means, and NAC standard deviations for atom types appearing in both calculated MOF structures and small molecule database. See DOI: 10.1039/c9ra07327b

MOF database was the starting point for our study. It contains a total of 5109 structures, of which 4764 structures were modified during the cleaning process and 345 retained their original CSD structures.[7] Fig. 1 illustrates some of the chemical diversity within this database. Since organic compounds contain carbon, porous metal-linker networks that do not contain any carbon atoms are called metal–inorganic frameworks (MIFs).[14,15] The 2014 CoRE database contains some structures that are MIFs.[13]

Several follow-ups were subsequently made to the CoRE MOF database. In 2016, Nazarian *et al.* reported DDEC net atomic charges (NACs) for 2932 structures.[16] In 2017, Nazarian *et al.* reported DFT-optimized geometries for 838 structures including 502 with computed DDEC NACs.[17] Soon, a major revision of the CoRE MOF database will be published that expands the number of structures to approximately 15 thousand (Yongchul G. Chung, Emmanuel Haldoupis, Benjamin J. Bucior, Maciej Haranczyk, Seulchan Lee, Hongda Zhang, Konstantinos D. Vogiatzis, Marija Milisavljevic, Sanliang Ling, Jeff S. Camp, Ben Slater, J. Ilja Siepmann, David S. Sholl, Randall Q. Snurr, in revision). Altintas *et al.*[115] reported a comparative analysis between the 2014 CoRE MOF database and the CSD-derived MOF database of Moghadam *et al.*[116]

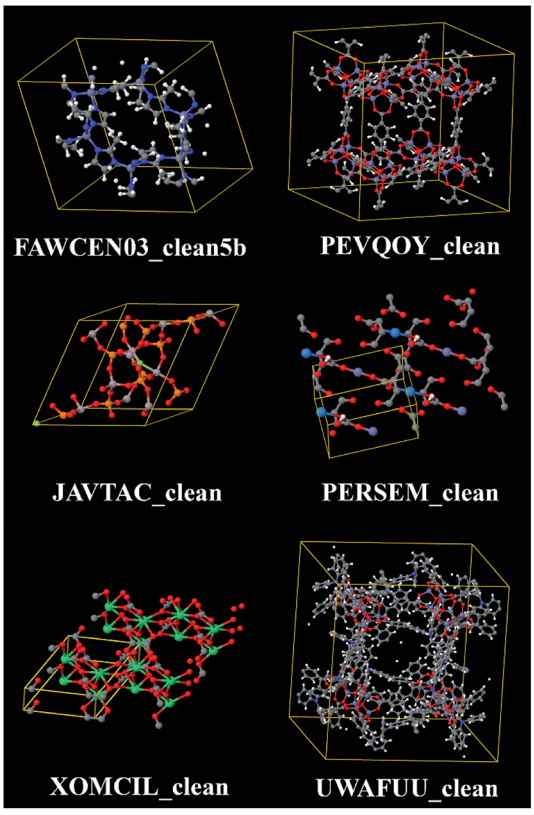

Fig. 1 Example structures from the 2014 CoRE MOF database. Top left: ZIF-8 is a zeolitic imidazolate framework containing zinc atoms and imidazolate linkers. Top right: IRMOF-3-AM4XL is an isoreticular metal–organic framework (IRMOF) containing interligand crosslinks.[51] Middle left: example metal–inorganic framework (MIF) containing no carbon or hydrogen atoms. Middle right: example MOF containing an actinide element (uranium). Bottom left: example MOF having a small unit cell, containing a lanthanide element (erbium), with no hydrogen atoms. Bottom right: example MOF having a large unit cell (1160 atoms).

Although many thousands of MOFs have been synthesized to date, this is only a tiny fraction of the number of MOFs that could potentially be made.[1] Owing to the large number of different metal atoms and organic linkers that could be combined in various ways, there is a nearly infinite number of potentially synthesizable MOFs.[1,18–20] Computational chemistry can be an efficient way to search this vast chemical space to help identify the most promising materials to later synthesize and experimentally test.[21] This will avoid unnecessary efforts to synthesize a large number of materials that do not perform well for the desired applications.

A host of important performance properties for MOFs and other complex materials can be calculated by modeling statistical ensembles (*e.g.*, constant NVE, NVT, NPT, μVT, μPT, NPH, *etc.*). Here, the properties can be roughly divided into thermodynamic properties that occur at equilibrium and dynamic properties that describe transient system response to non-equilibrium. Gas adsorption isotherms are often computed in the grand canonical (μVT) or Gibbs ensembles.[22,23] Gas diffusion constants are often computed in the canonical (NVT) or microcanonical (NVE) ensembles using equilibrium molecular dynamics.[24,25] System behaviors under extreme conditions, such as high pressures, low or high temperatures, applied electromagnetic fields, irradiation, high stresses, and corrosive conditions are of emerging interest.

When modeling statistical ensembles, an interaction model is required to compute the system's energy as a function of chemical configuration. Several different types of interaction models exist depending on the length scale at which the system is modeled. The smallest length scale considers individual electrons using an exchange–correlation theory (*e.g.*, DFT + dispersion functional[26,27]) to solve the Schrodinger equation for the material's electron distribution (*i.e.*, quantum chemistry). A somewhat larger length scale considers individual atoms interacting *via* a classical force field in atomistic simulations.[21,28,29] Coarse-grained and continuum models treat even larger length scales.[30,31]

Quantum chemistry using an exchange–correlation theory has different advantages and limitations compared to classical atomistic simulations using a force field. An exact exchange–correlation theory (*e.g.*, full configuration interaction) would describe all material types but is extremely computationally expensive. In practice, approximate exchange–correlation theories are used that describe an extremely broad range of different material types with acceptable accuracy and computational efficiency. Because exchange–correlation theories are not material specific, they should not have to be reparameterized for a new material type. In contrast, classical force fields have to be extensively parameterized for individual atoms or atom types.[29,32] Because computing the system's energy is usually much faster using a classical force field than an exchange–correlation theory, classical atomistic simulations can usually be performed faster and over larger length and time scales than quantum chemistry calculations.

Classical force fields typically contain bonded and non-bonded terms. Bonded terms describe flexibility (bond stretching, angle bending, torsion, and/or out-of-plane parameters) and have been incorporated into many flexible force fields for MOFs.[33–38] Non-bonded terms describe electrostatic interactions (NACs, atomic multipoles, charge penetration, polarizability[39–43]), London dispersion interactions, and/or short-range exchange-repulsion.[44,45] When optimizing the bonded parameter values, it is orders of magnitude more efficient to fit them to atom-in-material forces across multiple geometries than to only fit them to the system's total energy across multiple geometries.[46,47] Because each geometry yields many atom-in-material forces but only one total energy, fewer geometries are needed to fit the bonded parameter values to forces than to energies.

A classical force field must contain many-body dispersion and/or many-body polarization to accurately describe system properties over a wide range of conditions.[48] As shown in Fig. 2, dipolar interactions between two particles can be purely attractive, but dipolar interactions between many non-collinear particles are partly attractive and partly repulsive. Consequently, fitting two-body potentials (*e.g.*, Lennard-Jones) to two-particle interaction curves gives force fields that often overestimate liquid-phase densities.[49] If the parameters of the two-body potential are adjusted to yield correct liquid-phase densities, then the two-particle interaction curve will be incorrectly described.[45] This problem can be fixed by explicitly including many-body dispersion and/or many-body polarization in the force field. For example, Kiss and Baranyai showed polarizability must be included in a force field to describe liquid water accurately over a wide range of conditions.[50]

Currently, there is a bottleneck to effectively model statistical ensembles: the quantum chemistry calculations are computationally expensive while the classical force fields are tedious to accurately parameterize. To resolve this bottleneck, either the quantum chemistry calculations should be sped up by orders of magnitude or classical force field parameterization should be made orders of magnitude easier. Some progress was made, but much more work remains to be done. Car-Parrinello molecular dynamics (CPMD) is a quantum chemistry method in which the wavefunction does not have to be self-consistently solved at each time step, because it evolves *via* an effective Lagrangian. CPMD improves the computational efficiency of *ab initio* molecular dynamics, but CPMD is limited to materials containing a band gap (*i.e.*, insulators and semi-conductors).[52–54] VASP can perform *ab initio* molecular dynamics even for metals, but this requires computing self-consistent orbitals at each time step.[54,55] Recently, many studies focused on first-principles derived classical force fields whose parameters are extracted from quantum chemistry calculations.[22,35,40,44,45,49,56,57] Other studies parameterized force fields by fitting to experimental data[36,58] or using machine learning (ref. 59–62).

This work is a large-scale computation of atom types and non-bonded parameters for MOFs. Fig. 3 summarizes our project workflow. The 5109 structures from the 2014 CoRE MOF database included 345 structures directly from CSD plus 4764 structures modified by Chung *et al.*[7] We attempted DFT calculations on all of the modified structures except the largest ones containing >∼1700 atoms; however, some DFT calculations did not converge. DFT calculations converged for 4445 structures that formed the calculated group, while the unconverged and large structures comprised 319 uncalculated structures. The acceptance or rejection criteria described in Section 2.4 were then applied to the CSD, calculated, and uncalculated structures. Atom typing was then applied to all of the accepted structures. A total of 8607 different atom types were identified. Various forcefield precursors were computed and reported in the ESI† for 3056 accepted calculated structures.

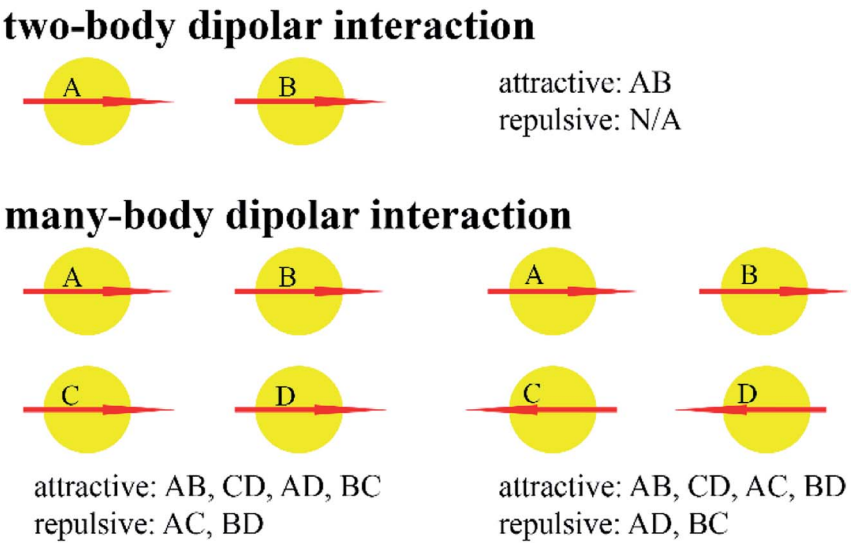

Fig. 2 Two-body (top) *versus* many-body (bottom) dipolar interactions. Two-body dipolar interactions can be purely attractive. As shown in the bottom left and bottom right, many-body dipolar interactions often contain a mixture of attractive and repulsive contributions.

2. Methods

2.1 Electron density calculation

The periodic quantum chemistry calculations of MOFs were performed using the PBE[63] exchange–correlation functional and the VASP[64,65] software. Frozen-core all-electron calculations were performed using the projector augmented wave[66,67] (PAW) method that uses a scalar-relativistic treatment of relaxed valence electrons and high-level relativistic treatment of frozen-core electrons.[68] The PAW potentials recommended on the VASP website were used for all these calculations. The planewave energy cutoff was 400 eV. Following prior recommendation,[69] the number of k-points along each lattice vector times the lattice vector length was ≥16 Å. If the *k*-points mesh was 1 × 1 × 1 then the Gamma point was used, otherwise we used Monkhorst–Pack[70] *k*-point grids. Real-space grids were chosen to avoid aliasing errors (*i.e.*, PREC = accurate).[69] The magnetic alignment of unpaired electron spins was optimized starting from the VASP default guess (which corresponds to a ferromagnetic alignment).

Calculations for the small organic molecules in Section 3.3 were performed using Gaussian16 (ref. 71) with the B3LYP[72,73] functional and def2QZVPPDD[74,75] basis set. Geometries of these molecules were optimized at this level of theory.

2.2 Computing forcefield precursors *via* atoms-in-material partitioning

Table 1 lists the computed forcefield precursors. The net atomic charge, atomic dipole, atomic quadrupole, electron cloud parameters, $\langle r^3 \rangle$ and $\langle r^4 \rangle$ radial moments, and atomic spin moment were computed in the Chargemol program using DDEC6 partitioning.[69,76,77]

The electron cloud parameters fit the electron density tail of each atom in the material to an exponential decay function. Specifically, the logarithm of the spherically averaged electron density assigned to atom A was fit to a straight line

$$\ln(\rho_A^{avg}(r_A)) \approx a - br_A \quad (1)$$

using least squares regression over the r_A values for which $10^{-4} \leq \rho_A^{avg}(r_A) \leq 10^{-2}$ e bohr^{-3}. r_A is the distance from position $\vec{r}$ to atom A's nuclear position. The fitted intercept a, slope b, and squared correlation coefficient R^2 are reported. The R^2 values were nearly always > 0.99 indicating nearly perfect linearity. These electron cloud parameters have several uses in molecular mechanics force fields. First, they are useful to describe charge penetration (also called cloud penetration).[76] Second, they are useful to describe short-range exchange-repulsion.[44] van Vleet *et al.* showed the Born-Mayer exponential term describing short-range exchange-repulsion has an effective exponent of ~0.84 × b.[44] Although their results were derived for iterated stockholder atom[78] (ISA) partitioning,[44] we expect the same relation to hold for DDEC6 partitioning. Third, these electron cloud parameters are useful to parameterize dispersion damping functions (*e.g.*, Tang–Toennies[79,80]) that keep the dispersion energy finite as the distance between two atoms approaches zero.[44,45]

The polarizabilities, dispersion coefficients, and quantum Drude oscillator (QDO) parameters were computed using the MCLF method[74] and software (ref. 81). The MCLF method yields several different types of polarizabilities and dispersion coefficients. The forcefield polarizability is a non-directionally screened polarizability that is suitable for use as an input parameter in polarizable force fields.[74] This avoids double counting the directional screening, because directional

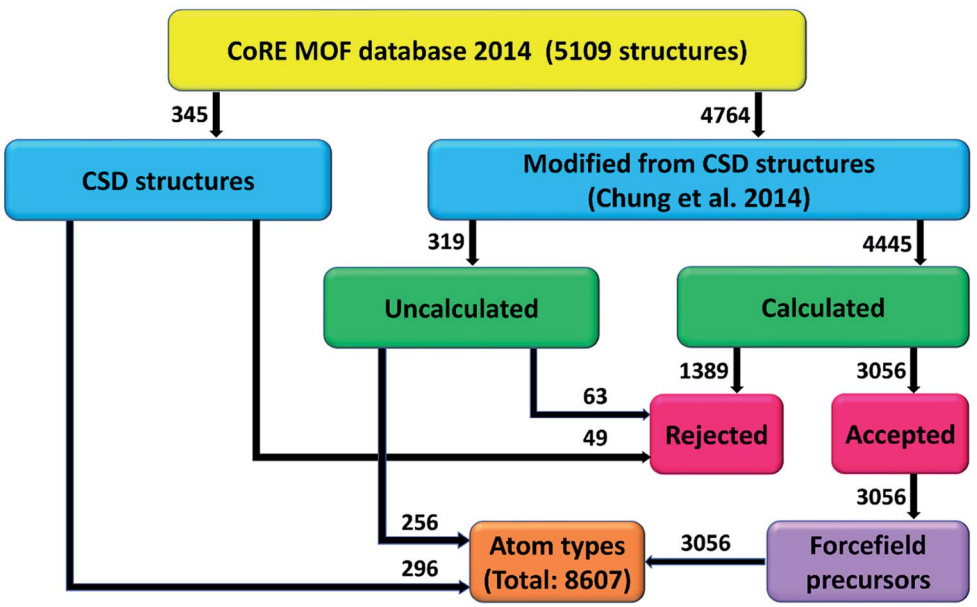

Fig. 3 Project flow diagram. Number of structures labeled next to each arrow.

screening naturally arises during the classical atomistic simulation when the force field is used.[74] The fluctuating polarizability is the polarizability associated with the fluctuating dipoles of the London dispersion interaction.[74] The static polarizability and static polarizability tensor quantify the system's response to an externally applied constant electric field.[74] Various dispersion coefficients are associated with different kinds of fluctuating multipoles: C_6 (dipole–dipole), C_8 (dipole–quadrupole), C_9 (dipole–dipole–dipole), C_{10} (dipole–octupole and quadrupole–quadrupole).[74,82,83] The C_9 coefficients were not printed to the forcefield precursors file, because they can be readily computed[74] from the printed forcefield polarizability and QDO parameters. The MCLF method includes convenient mixing rules (based on a QDO model) to easily calculate each of these dispersion coefficients between unlike atoms using only parameters of the individual atoms.[74]

Of key importance, the C_6 dispersion coefficient does not equal the r^{-6} coefficient of the Lennard-Jones force field.[84] Because the Lennard-Jones force field does not explicitly include higher-order dispersion (*e.g.*, C_8, C_{10}, *etc.*) terms, the Lennard-Jones r^{-6} coefficient must be made artificially higher than C_6 to effectively compensate for the neglected C_8 and C_{10} dispersion terms.[74,84] Near the minimum energy separation between two atoms, higher-order dispersion (*e.g.*, C_8, C_{10}, *etc.*) can contribute ~35% of the total dispersion energy.[79]

A QDO is a quantum harmonic oscillator containing a pseudoelectron attracted to a pseudonucleus.[85,86] Three QDO parameters were computed for each atom in the material: (a) the pseudoelectron's effective charge, (b) the effective QDO frequency, and (c) the QDO's reduced mass.[74,85] Force fields based on QDO models describe multibody dispersion and multibody polarization beyond the dipole approximation but require advanced simulation techniques.[85–87]

2.3 Quantifying electrostatic model accuracy

The root mean squared error (RMSE) of an electrostatic model quantifies its error compared to the quantum mechanically computed (*e.g.*, DFT) electrostatic potential over a chosen set of grid points:[88]

Table 1 List of computed forcefield precursors

Forcefield precursor	Method reference
Net atomic charge	69 and 76
Atomic dipole and quadrupole	76
Electron cloud parameters	This work
$\langle r^3 \rangle$ and $\langle r^4 \rangle$ radial moments	69
Atomic spin moment	69 and 77
C_6, C_8, C_{10} dispersion coefficients	74
QDO parameters	74
Forcefield, fluctuating, and static polarizability scalars	74
Static polarizability tensor	74

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{grid\_points}}} \sum_{\mathrm{grid\ \ points}} \left(V^{\mathrm{model}}\left(\vec{r}\right) + V_{\mathrm{offset}}^{\mathrm{model}} - V^{\mathrm{QM}}\left(\vec{r}\right) \right)^2} \tag{2}$$

$$V_{\mathrm{offset}}^{\mathrm{model}} = \frac{1}{N_{\mathrm{grid\_points}}} \left(\sum_{\mathrm{grid\ \ points}} \left(V^{\mathrm{QM}}\left(\vec{r}\right) - V^{\mathrm{model}}\left(\vec{r}\right) \right) \right) \tag{3}$$

where $N_{\mathrm{grid\_points}}$ is the number of grid points used to compute RMSE. $V_{\mathrm{offset}}^{\mathrm{model}}$ is the average difference between $V^{\mathrm{QM}}(\vec{r})$ and $V^{\mathrm{model}}(\vec{r})$ over these grid points.[89] These grid points are normally chosen to be outside the material's van der Waals surface.[90] For a MOF crystal, these grid points occur inside the MOF's pores.

In this study, the RMSE grid points were uniformly distributed in the volume of space that simultaneously met all three of the following criteria: (i) the material's total electron density, $\rho(\vec{r})$, was $\leq 10^{-4}$ e bohr^{-3}, (ii) the grid point was no closer than 5 bohr to any atom in the material, and (iii) the grid point was closer than 12 bohr to at least one atom in the material. If any one of these three criteria were not met, the grid point was not used.

The relative root mean squared error (RRMSE) is the ratio of the electrostatic model's RMSE to the RMSE of a null model for which $V^{\mathrm{null\_model}}(\vec{r}) = 0$.[91] Note that $V_{\mathrm{offset}}^{\mathrm{null\_model}}$ equals the average of $V^{\mathrm{QM}}(\vec{r})$ over the RMSE grid points, $V_{\mathrm{avg}}^{\mathrm{QM}}$. Therefore,

$$\mathrm{RRMSE} = \sqrt{\frac{\sum_{\mathrm{grid\ \ points}} \left(V^{\mathrm{model}}\left(\vec{r}\right) + V_{\mathrm{offset}}^{\mathrm{model}} - V^{\mathrm{QM}}\left(\vec{r}\right) \right)^2}{\sum_{\mathrm{grid\ \ points}} \left(V_{\mathrm{avg}}^{\mathrm{QM}} - V^{\mathrm{QM}}\left(\vec{r}\right) \right)^2}} \tag{4}$$

The RRMSE quantifies the fraction of electrostatic potential variations that are not described by the model. For example, an RRMSE = 0.1 means that 90% of the electrostatic potential variations are captured by the model and 10% are not. This RRMSE definition was used in several prior studies[76,88,91,92] co-authored by one of us and differs from the even earlier literature[89,93] by including the potential offsets in this RRMSE definition. These potential offsets must be included in the RMSE and RRMSE definitions, because in periodic materials the electrostatic potential has no obvious spatial position where it approaches zero value and the energy of a system having no net charge is invariant to a constant electrostatic potential shift.

In this work, the DFT electrostatic potential $V^{\mathrm{QM}}(\vec{r})$ was output from VASP using the keyword LVHAR = .true. We wrote an OpenMP parallelized Fortran program to compute RMSE and RRMSE. The RMSE and RRMSE were computed for the following four electrostatic models: (A) NACs only (aka monopole order), (B) NACs plus spherical cloud penetration, (C) NACs plus atomic dipoles (aka dipole order), and (D) NACs plus atomic dipoles plus spherical cloud penetration.

2.4 Structure acceptance or rejection

The 2014 CoRE MOF database was developed with goals of fixing disordered atoms, adding missing hydrogens, removing solvent molecules, discarding unfixable structures, and other cleaning protocols.[7] These goals were partially but not fully achieved.

Table 2 Reasons for rejecting MOF structures

Rejection reason	Identification method	Category
Unreasonable SBO for non-metal atom	DDEC6	Calculated
Negative NAC on metal atom	DDEC6	Calculated
RRMSE > 0.5	Direct calculation	Calculated
Isolated atom	Radii based connectivity	All
Unreasonable connectivity	Visual inspection	CSD and uncalculated

To partially address remaining structural deficiencies in the 2014 CoRE MOF database, we performed: (i) reasonability checks on our computed forcefield precursors for the calculated structures and (ii) connectivity checks for all structure categories. Table 2 lists reasons for rejecting a MOF structure.

The sum of bond orders (SBO) was used as a screening criterion for reasonableness. DDEC6 bond orders and SBO for each atom in a material were computed using the Chargemol program.[69,94] Table 3 lists accepted SBO ranges for non-metal elements appearing in the database. A MOF structure was rejected if the SBO for any atom in the material was outside these ranges. Since a carbon atom has four valence electrons available to share in covalent bonding, its SBO is normally expected to be ~4. The acceptable range of SBO for C atoms was set to [3.5, 4.75]. Since F and H have one electron to share in covalent bonding, their accepted SBO ranges were set to [0.5, 1.5]. The remaining halogens (Cl, Br, and I) had accepted SBO ranges of [0.5, 5.0] to accommodate situations in which these atoms were bonded to several O atoms (*e.g.*, ClO_4^-, BrO_4^-, IO_4^-) which can give relatively large SBOs. Oxygen atom has two electrons to share in covalent bonding but can also bond *via* lone-pairs (*e.g.*, Lewis acid–base interaction and hydrogen bonding); therefore, its accepted SBO range was set to [1.5, 3.0]. Boron and nitrogen atoms can exhibit variable bonding involving ~3 or ~4 covalent bonds (*e.g.*, BF_3, B_2H_6, NH_3, NH_4^+); therefore, accepted SBO ranges were set to [2.5, 4.5] for B and [2.5, 4.75] for N. The upper bound for N was slightly higher than for B to accommodate the formally higher SBO of N in nitrates than B in borates. Since Si prefers ~4 bonds, its accepted SBO range was set to [3.5, 4.5]. The accepted SBO ranges for S, Se, and Te were set to [1.5, 6.0] to accommodate their ability to form ~2 (*e.g.*, H_2S, H_2Se, H_2Te) to ~6 heuristic bonds (*e.g.*, SO_4^{2-}, SeO_4^{2-}, TeO_4^{2-}). The accepted SBO ranges for P and As were set to [3.0, 6.0] to accommodate their ability to form ~3 (*e.g.*, PF_3, AsF_3) to ~6 heuristic bonds (*e.g.*, $H_2PO_4^-$, $H_2AsO_4^-$). We do not claim these choices are perfect, but they can screen out some bad structures.

A calculated structure was rejected if any metal atom had negative NAC. In this article, the definition of metal atom included all elements not listed in Table 3 except rare gases (He, Ne, Ar, Kr, Xe, Rn) and At. The rational for this criterion is that metal atoms in MOFs are typically bound to more electronegative elements (*e.g.*, N, O, *etc.*).

A calculated structure was rejected if any one of the four RRMSE described in Section 2.3 was greater than 0.5. This criterion ensured each of our four electrostatic models described the electrostatic potential with an error less than half of a no charges model. In other words, each electrostatic model described the majority of electrostatic potential variations in the MOF's pores.

Structures were rejected from all categories if they contained any isolated atoms. An atom was considered isolated if it was not connected to any other atom based on the radii listed in Table 5.

For CSD and uncalculated structures, some structures were rejected based on strange connectivity identified by visual inspection. Because this visual inspection was not performed systematically, we do not claim all instances of bad connectivity were identified.

Table 4 lists the category breakdown for 2014 CoRE MOF structures. Calculated structures were those for which an electron density distribution was calculated using DFT. Uncalculated structures were non-CSD structures for which the electron density distribution was not calculated. This could be due to one of two factors: (i) the MOF structure was so large that DFT calculation was not practical within our computational budget or (ii) the DFT calculation did not readily converge. CSD structures were those for which the geometry came directly from the CSD database.[7] We do not report forcefield precursors for the

Table 3 Accepted ranges of SBOs for non-metal elements

H	[0.5, 1.5]	S	[1.5, 6.0]
B	[2.5, 4.5]	Cl	[0.5, 5.0]
C	[3.5, 4.75]	As	[3.0, 6.0]
N	[2.5, 4.75]	Se	[1.5, 6.0]
O	[1.5, 3.0]	Br	[0.5, 5.0]
F	[0.5, 1.5]	Te	[1.5, 6.0]
Si	[3.5, 4.5]	I	[0.5, 5.0]
P	[3.0, 6.0]		

Table 4 Category breakdown for 2014 CoRE MOF structures

	Total structures	Rejected	Accepted
Calculated	4445	1389	3056
Uncalculated	319	63	256
CSD	345	49	296
Total	5109	1501	3608

CSD structures, because CSD licensing terms may not allow us to publicly distribute their geometries.

The second column of Table 4 lists the total number of structures in each category. The third and fourth columns list the number of rejected and non-rejected structures in each category, respectively. The ESI† contains detailed lists of MOFs in each category. For each rejected structure, the ESI† lists the specific reason(s) that structure was rejected.

3. Results and discussion

3.1 Forcefield precursors

Files containing forcefield precursors for 3056 accepted calculated structures are contained in the ESI.† The data collected for each atom in a MOF includes: the coordinates; the net atomic charge (NAC); C_6, C_8 and C_{10} dispersion coefficients; fluctuating polarizability; forcefield polarizability; electron cloud parameters; $\langle r^3 \rangle$ and $\langle r^4 \rangle$ moments; QDO parameters; atomic dipole and quadrupole; atomic screened static polarizability tensor and isotropic polarizability; atomic spin moment. The data collected for the MOF's unit cell includes: the lattice vectors, the total spin magnetic moment, the net charge, and the RRMSE of the electrostatic potential for electrostatic models including (A) NACs only, (B) NACs plus spherical cloud penetration, (C) NACs plus atomic dipoles, and (D) NACs plus atomic dipoles and spherical cloud penetration. For each MOF, these data were written to a *xyz* file whose chemical structure can be visualized with the Jmol[95,96] program. In the future, these forcefield precursors could be used as building blocks to construct force fields for these materials.

3.2 RRMSE of electrostatic potential

Fig. 4 plots histograms of electrostatic potential RRMSE for the accepted calculated structures. At monopole order (with or without spherical cloud penetration), the histograms peak at RRMSE = (0.1, 0.2). At both monopole and dipole orders, spherical cloud penetration had negligible effect on the histograms. Dipole order showed dramatic improvement over monopole order.

Some additional comments are in order. Even though spherical cloud penetration had negligible effect on the RRMSE values, this does not necessarily mean charge penetration effects are not important for constructing accurate electrostatic models. In MOF's, the RRMSE probes volume in space that is likely occupied by an adsorbate atom's nuclear position. Because charge penetration extends over the volume occupied by an adsorbate atom's full electron density distribution, it affects the electrostatic potential over a much larger volume than that probed by RRMSE calculation. For this reason, charge penetration can be more important than indicated by the RRMSE values. Prior studies demonstrated intermolecular electrostatic interaction energies are substantially improved by including charge penetration.[97–99] Therefore, including

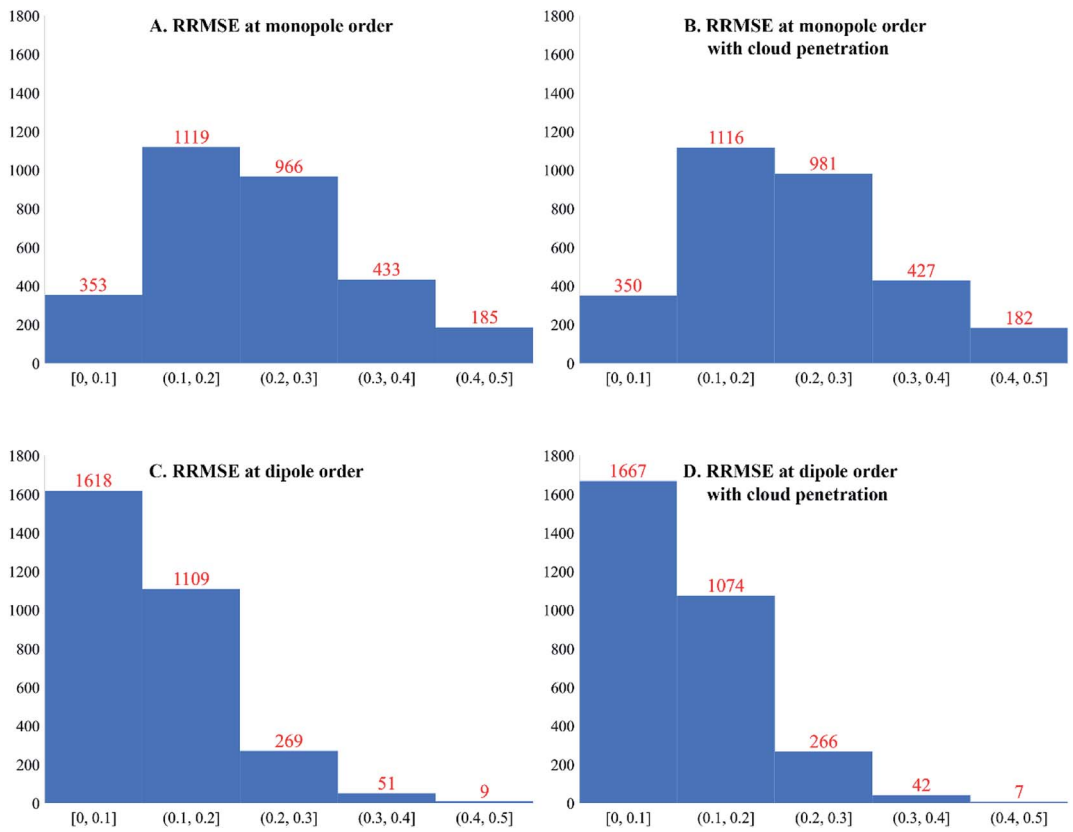

Fig. 4 Histograms of electrostatic potential RRMSE (dimensionless) for accepted calculated structures.

spherical cloud penetration (aka 'charge penetration') will likely improve the adsorbate–MOF and adsorbate–adsorbate electrostatic interaction energies even though its effects on RRMSE were negligible.

When constructing force fields, decisions have to be made about tradeoffs between simplicity and accuracy. For best accuracy, the force field should include NACs, atomic dipoles (and possibly atomic quadrupoles), and spherical charge penetration. However, the simplest force field would include NACs only and neglect atomic multipoles and charge penetration.

3.3 Atom types

Atom types are widely used in force fields.[29] The main idea of atom typing is to classify similar atoms into the same atom type to facilitate forcefield parameterization, where all atoms of the same atom type are normally assigned identical forcefield parameters.[32] Biomolecular force fields have used atom types for several decades.[100–106] The universal force field (UFF) and its extensions to MOFs (*e.g.*, UFF4MOF) focused on atom types for structural optimizations.[107–109]

The connectivity based atom contribution (CBAC) method defined atom types in porous materials (*e.g.*, MOFs and covalent organic frameworks) using first neighbors only.[110,111] Two significant limitations of the CBAC method are: (i) its atom types are based on first neighbors only which limits chemical transferability and (ii) the NACs for its atom types were obtained by CHELPG analysis of small fragment clusters but it is unclear how accurately these may reproduce the fully periodic electrostatic potential.[110,111]

In our work, atom types are defined based on both first and second neighbors leading to better chemical transferability. Also, the NACs for our atom types were calculated using the fully periodic electrostatic potential, rather than small fragment clusters. Fig. 5 illustrates our atom typing scheme. When multiple neighbor groups are present, they are sorted by the atomic number of the neighbor atoms from smallest to largest. If a first neighbor has no second neighbors, its lack of second neighbor is indicated by (0) in the atom type label. For example, the atom type 6[1-(0),1-(0),1-(0),6-(1,1,8)] indicates a central carbon atom with four first neighbors (H, H, H, C) where each of the first neighbor H atoms is not directly bonded to any second neighbors and the first-neighbor C atom is directly bonded to H, H, and O plus the central atom. As another example, the atom type 15[6-(6),6-(6,6),6-(6,6),29-(7,7,7)] demonstrates that 6-(6) is listed prior to 6-(6,6). As another example, the atom type 7[1-(0),6-(1,1,6),6-(6,8)] demonstrates that 6-(1,1,6) is listed prior to 6-(6,8).

During atom typing, two atoms were considered directly connected if the distance between them was no greater than the sum of the atom typing radii listed in Table 5. These radii have the following history. Starting with the atom connectivity radii of OpenBabel version 1.100.1, we added 0.18 Å (*i.e.*, a skin distance was incorporated into the radii). Second, we reduced some transition metal radii to eliminate unreasonable metal–metal bonds and excessively high coordination numbers.

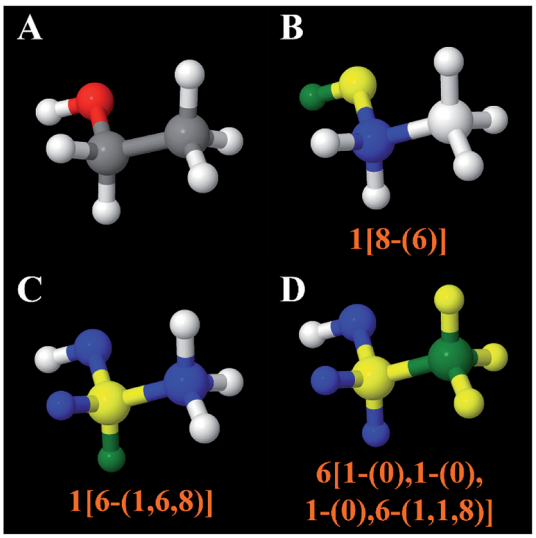

Fig. 5 Illustration of labeling atom types. Panel (A) shows an ethanol molecule with atoms colored by element: grey (C), red (O), white (H). Panels (B–D) illustrate the typing of different atoms in this molecule. In each panel, the chosen atom is colored green. Its first neighbors are colored yellow. Its second neighbors are colored blue. In Panel (B), the hydrogen in the hydroxy group is chosen as the central atom. It is connected to one oxygen that is connected to a carbon atom plus the central atom. In Panel (C), the α-H is chosen as the central atom. It is connected to one carbon that is connected to one hydrogen, one carbon, and one oxygen plus the central atom. In Panel (D), the β-C is chosen as the central atom. It is connected to three hydrogens and one carbon. The carbon is connected to two hydrogens and one oxygen plus the central atom.

Third, the carbon and oxygen radii were slightly increased to maintain carbon/oxygen–metal bonds after decreasing the metal radii. These radii values are unique only in an

Table 5 Radii (Å) used for atom typing. Two atoms were considered directly connected if the distance between them was no greater than the sum of these radii

H	0.38	Cr	1.53	Pd	1.68	Er	1.80
Li	0.86	Mn	1.53	Ag	1.56	Tm	1.84
Be	0.53	Fe	1.43	Cd	1.56	Yb	1.80
B	1.01	Co	1.31	In	1.53	Lu	1.86
C	0.88	Ni	1.33	Sn	1.64	Hf	1.73
N	0.86	Cu	1.31	Sb	1.64	W	1.33
O	0.89	Zn	1.41	Te	1.65	Re	1.29
F	0.82	Ga	1.40	I	1.58	Ir	1.50
Na	1.15	Ge	1.35	Cs	1.85	Pt	1.66
Mg	1.28	As	1.39	Ba	1.52	Au	1.68
Al	1.53	Se	1.40	La	1.91	Hg	1.88
Si	1.38	Br	1.39	Ce	1.98	Pb	1.72
P	1.28	Rb	1.65	Pr	1.75	Bi	1.72
S	1.20	Sr	1.30	Nd	1.92	Th	1.97
Cl	1.17	Y	1.84	Sm	1.89	U	1.76
K	1.44	Zr	1.73	Eu	1.83	Np	1.73
Ca	1.17	Nb	1.66	Gd	1.79	Pu	1.71
Sc	1.62	Mo	1.57	Tb	1.82		
Ti	1.65	Ru	1.58	Dy	1.79		
V	1.51	Rh	1.63	Ho	1.63		

approximate sense, because small increases or decreases in the radii values may be feasible. Large changes in these radii values are not feasible, because bad atom connectivity would result.

These atom type definitions gave a total of 1313 first-neighbor atom types and 7033 second-neighbor atom types for the 3056 accepted calculated MOFs. Of these, 783 first-neighbor and 3015 second-neighbor atom types appeared in more than one MOF.

Is second-neighbor-based atom typing optimal? To explore this question, the averages and standard deviations of 17 different forcefield precursors were computed and listed in the ESI† for each first-neighbor-based and each second-neighbor-based atom type: NAC; C_6, C_8, and C_{10} dispersion coefficients; fluctuating, forcefield, and static polarizabilities; $\langle r^3 \rangle$ and $\langle r^4 \rangle$ radial moments; a, b, and R^2 electron cloud parameters; atomic dipole moment magnitude; three QDO parameters; and atomic spin moment. Fig. 6 presents histograms of the standard

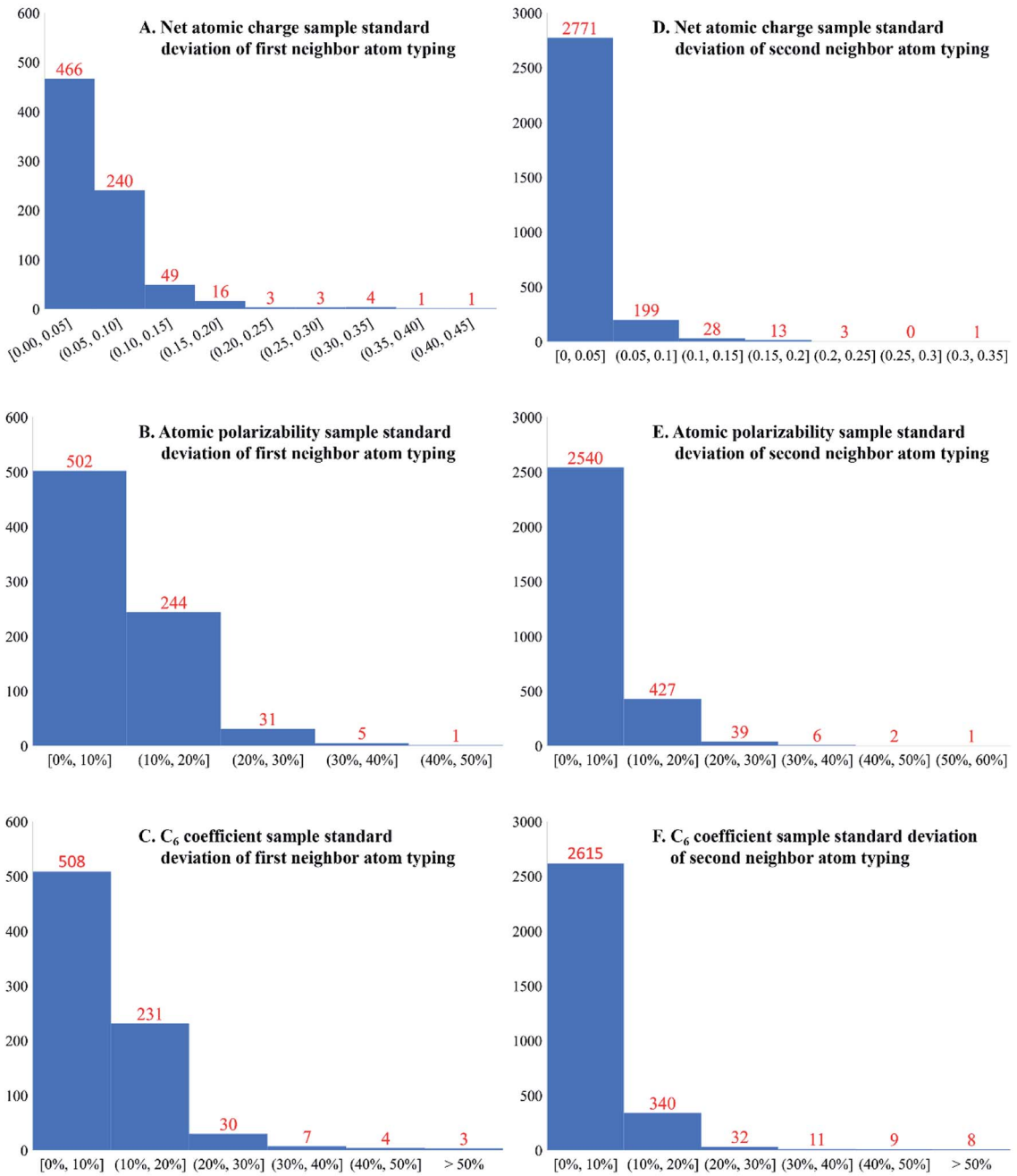

Fig. 6 Histograms of standard deviations of NACs, forcefield polarizabilities, and C_6 dispersion coefficients for first-neighbor-based and second-neighbor-based atom types. Second-neighbor-based atom types included both first and second neighbors. Each standard deviation was computed across different occurrences of the same atom type. The second-neighbor-based atom types showed much smaller standard deviations than the first-neighbor-based atom types.

deviations for NAC, forcefield polarizability, and C_6 across first-neighbor-based and second-neighbor-based atom types. Only atom types appearing in more than one accepted calculated structure were included in this analysis. Second-neighbor-based atom types showed remarkably better chemical transferability compared to first-neighbor-based atom types. For example, 91.9% of the second-neighbor-based atom types had NAC standard deviations $\leq 0.05e$, while only 59.5% of the first-neighbor-based atom types did.

This atom typing is theoretically justified, because in organic molecules electronic effects play a major role in group reactivity and induction contributes a significant portion of that effect. In organic chemistry, the inductive effect is the electron donating or withdrawing ability of a chemical group that can be transmitted to other parts of the molecule through chemical bonds.[112] The inductive effect usually decreases across each bond and is usually limited to 2 to 3 bonds lengths.[112] Our decision to base atom typing on first and second neighbors is a pragmatic one. As shown in Fig. 6, using only first neighbors leads to large standard deviations in the calculated forcefield parameters. The standard deviation is greatly reduced by incorporating second neighbors in the atom types. Incorporating third neighbors into the atom type definition would lead to an enormous number of different atom types, thus making forcefield parameterization too tedious. Therefore, second-neighbor-based atom types are optimal.

Fig. 7 explores an extreme case of third neighbor inductive effects. In malonic acid (Fig. 7B), the four highly electronegative oxygen groups withdraw electrons increasing the acidity of the circled hydrogen atom compared to the circled hydrogen in propane (Fig. 7A). Specifically, the circled hydrogen in Fig. 7B can be stripped away from the carbon atom by strong base while the one in Fig. 7A does not possess the same reactivity. The NACs of those hydrogens are 0.06 for Fig. 7A and 0.18 for Fig. 7B. If these hydrogens are replaced with a methyl group, the methyl carbon NACs are −0.37 and −0.34 respectively. These differences in NACs are within an acceptable range for atoms sharing the same atom type.

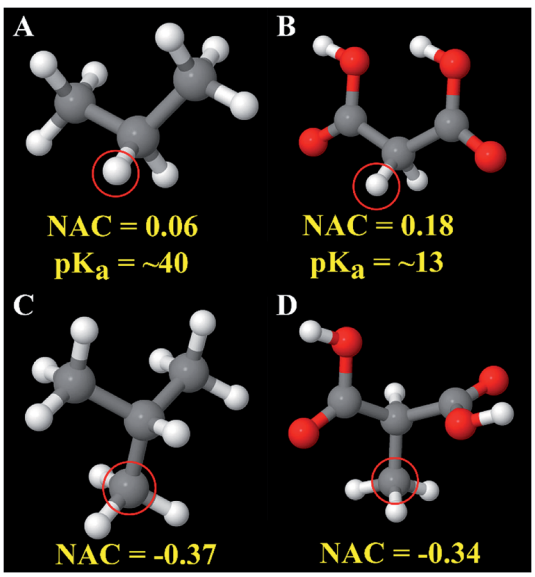

Fig. 7 Atoms of the same atom type can sometimes have different chemical reactivities. Panels (A and B) show hydrogen atoms having the same atom type but very different pK_a values. The approximate pK_a values are from ref. 113. Panels (C and D) show the good transferability of carbon atom NAC for analogous compounds in which the circled hydrogen atoms in Panels (A and B) were replaced by a methyl group.

Table 6 summarizes the breakdown of atom types in the 3608 accepted structures. A total of 8607 different atom types appeared in these structures. 3015 atom types appeared in more than one accepted calculated structure, while 4018 atom types appeared in only one accepted calculated structure. An additional 1574 atom types appeared only in the uncalculated and/or CSD structures. These results clearly show that many atom types appeared in only one MOF in the dataset. These results also show the nearly infinite chemical diversity of MOFs. For example, the 296 accepted CSD structures contained 320 new atom types that did not appear in either the calculated or uncalculated structures.

Do a relatively small number of atom types completely describe a large percentage of the accepted calculated structures? Table 7 summarizes calculations performed to explore this question. Starting with a 'frequency threshold', any MOF containing any atom type that appeared in fewer than 'frequency threshold' different MOFs (within the 3056 accepted calculated structures) was discarded. All atom types appearing in the remaining MOFs occurred in at least 'frequency threshold' different MOFs in the 3056 accepted calculated structures. Then, the number of distinct atom types appearing in the remaining MOFs was counted and reported in the second column of Table 7. The results show the chemical diversity of MOFs in this dataset is well-mixed. In other words, there was not a large segregated subset of MOFs in this dataset that are completely described by a few atom types.

Zn and Cu were the most common metal elements in the 2014 CoRE MOF database.[7] Among the 3056 accepted calculated structures, 785 structures were built entirely of the elements Cu, Zn, C, H, N, and O. These 785 structures contained 1314 different atom types.

Finally, we tested the transferability of atom typing across different classes of materials. Applying the second-neighbor atom typing procedure discussed above to the molecular test set of Bleiziffer *et al.*,[62] we obtained 32 530 atom types from

Table 6 Breakdown of atom types in accepted structures

Appeared in more than one calculated MOFs	3015
Appeared in only one calculated MOF	4018
Unique to uncalculated MOFs	1241
Unique to CSD MOFs	320
Appeared in both uncalculated and CSD MOFs but not calculated MOFs	13
Total	8607

Table 7 Exploration of most important atom types in accepted calculated structures. Each atom type appearing in the enumerated MOFs appeared in at least 'frequency threshold' different accepted calculated structures. The listed number of atom types is the number of different atom types in these enumerated MOFs

Frequency threshold	# atom types	# MOFs	Frequency threshold	# atom types	# MOFs
5	1167	1120	30	166	219
10	572	682	35	138	188
15	385	508	40	120	151
20	273	334	45	106	107
25	208	268	50	94	77

130 265 molecules in the dataset (aka '130k' dataset). (Bleiziffer *et al.*'s dataset contained 130 267 structures, but we did not include the two of these structures containing unbonded atoms.) The DDEC6 NACs for these molecules are obtained from Bleiziffer *et al.*'s data for a dielectric constant of 4.[62] Of the atom types from the 130k molecules and 3056 accepted calculated CoRE MOFs, 798 are shared in both datasets and 780 of these have more than one occurrence (in the same or multiple structures) in each dataset. Fig. 8 is a histogram of the absolute difference of the average DDEC6 NAC between the 130k molecules dataset and the 3056 accepted calculated MOF dataset for these 780 atom types. As shown in Fig. 8, most of these atom types had an average DDEC6 NAC absolute difference less than 0.1. This shows our atom typing scheme has good chemical transferability and can be applied to different material types. Also, Bleiziffer *et al.* trained a machine learning model on the 130k dataset to predict DDEC6 NACs with high accuracy.[62] The similar DDEC6 NAC between CoRE MOF and 130k molecules datasets for a shared atom type and Bleiziffer *et al.*'s successful machine learning model indicate it should be feasible to train a machine learning model to predict the properties of new atom types. Due to the extremely large number of distinct atom types we found in these datasets, a machine learning model could be highly useful to assign force-field parameter values for such a large number of atom types.

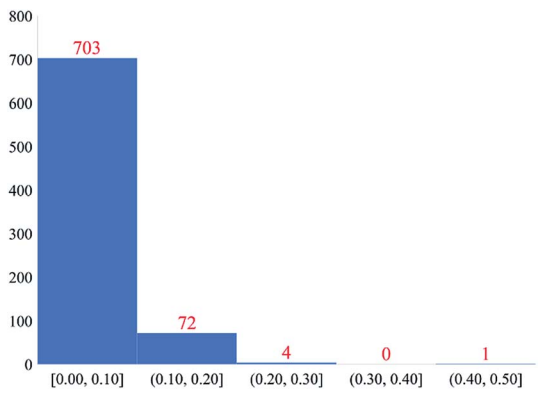

Fig. 8 Histogram of absolute difference of average DDEC6 NAC between 130k molecules dataset and 3056 accepted calculated MOFs dataset. This histogram includes 780 atom types that have more than one occurrence in each dataset.

4. Conclusions

In this article, we considered the problem of how to automatically extract forcefield precursors from quantum chemistry calculations using DDEC6 atomic population analysis. We focused on calculating non-bonded parameters for each atom in the material: net atomic charge, atomic dipole and quadrupole, electron cloud parameters, atomic spin moment, dispersion coefficients, various polarizabilities, and QDO parameters. Our first main result was calculated values of these atomistic descriptors for 3056 MOFs that will serve as building blocks to construct classical force fields for these materials.

Regarding the electrostatic parameters, our results confirm the general belief that NACs can often reproduce the electrostatic potential surrounding a material with acceptable accuracy. Including atomic dipoles dramatically reduced the electrostatic potential errors; therefore, it is preferable to include atomic dipoles in the electrostatic model. Spherical cloud penetration had negligible effect on the RRMSE histograms; however, charge penetration effects can still be important at short interatomic distances.

Our second main result was the practical assessment of atom typing for MOFs. We improved values of the atom typing radii that define whether two atoms in a material are directly bonded to each other. Our results showed that atom types based on both first and second neighbors adequately captures the chemical environment, but atom types based on only first neighbors does not. Specifically, the standard deviation of calculated forcefield precursors was relatively large across atoms sharing similar first neighbor environments but relatively small across atoms sharing similar first and second neighbor environments. Including third neighbors in the atom type definition would create an unnecessarily large and burdensome number of different atom types. Therefore, atom typing including both first and second neighbors is optimal.

Our results demonstrate the large chemical diversity of MOFs. 8607 different atom types were identified in the 3608 non-rejected MOFs. These atom types should be useful to develop future force fields for MOFs. Although the MOF to atom type ratio was lower than 1.0 in our study, there are two key reasons to be believe this ratio will improve in the future: (1) the number of chemical elements from which MOFs can be synthesized is practically limited to ~100, because heavier chemical elements undergo rapid radioactive decay. Therefore,

new MOFs will reuse many of the same chemical elements. (2) There is an almost infinite number of different ways to combine common metals and common organic linkers to form MOFs.[18,19] To date, only a tiny fraction of these hypothetical MOFs have been chemically synthesized.[18,19] As more and more of these hypothetical MOFs are synthesized, many atom types will be reused leading to a higher MOF to atom type ratio. Moreover, parameterizing only the popular atom types could describe a substantial percentage of MOFs with higher MOF to atom type ratio.

In summary, more research is urgently needed to develop accurate interaction models for MOFs. Because of the large chemical diversity of real and hypothetical MOFs, it is impractical to evaluate all possible MOFs experimentally. Therefore, computational assessment is required. The interaction model is a crucial ingredient of computational assessment. Two major types of interaction models are exchange–correlation theories (*e.g.*, DFT + dispersion) and classical force fields. The former needs orders of magnitude improvement in computational speed to yield rapid *ab initio* molecular dynamics. The latter requires extensive parameterization to yield accurate force fields. To actually develop working force fields, the non-bonded interaction terms studied herein will need to be combined with bonded interaction terms (*e.g.*, bond springs, angle springs, torsion and/or out-of-plane parameters) and further parameterization of short-range exchange-repulsion. However, in simulations using a rigid framework approximation (*e.g.*, Monte Carlo simulations of adsorption in rigid frameworks), the bonded interaction terms are not required.

Funding

National Science Foundation (NSF) CAREER Award DMR-1555376 provided financial support. Supercomputing resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE).[114] XSEDE is funded by NSF grant ACI-1548562. XSEDE project grant TG-CTS100027 provided allocations on the Stampede2 cluster at the Texas Advanced Computing Center and the Comet cluster at the San Diego Supercomputing Center.

Authors' contributions

T. C. performed the calculations and wrote Python scripts to prepare input files, analyze output files, perform atom typing, determine whether to accept or reject structures, and write files containing force-field precursors and atom type statistics. T. A. M. wrote the programs to compute the electrostatic potential RRMSE and electron cloud parameters. T. A. M. obtained funding for the project. Both authors designed the study, interpreted data, and wrote the manuscript.

Conflicts of interest

There are no conflicts of interest to declare.

References

1 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, The chemistry and applications of metal-organic frameworks, *Science*, 2013, **341**, 1230444.
2 Y. B. He, F. L. Chen, B. Li, G. D. Qian, W. Zhou and B. L. Chen, Porous metal-organic frameworks for fuel storage, *Coord. Chem. Rev.*, 2018, **373**, 167–198.
3 J. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S. T. Nguyen and J. T. Hupp, Metal-organic framework materials as catalysts, *Chem. Soc. Rev.*, 2009, **38**, 1450–1459.
4 K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T. H. Bae and J. R. Long, Carbon dioxide capture in metal-organic frameworks, *Chem. Rev.*, 2012, **112**, 724–781.
5 J. R. Li, J. Sculley and H. C. Zhou, Metal-organic frameworks for separations, *Chem. Rev.*, 2012, **112**, 869–932.
6 A. U. Czaja, N. Trukhan and U. Muller, Industrial applications of metal-organic frameworks, *Chem. Soc. Rev.*, 2009, **38**, 1284–1293.
7 Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl and R. Q. Snurr, Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals, *Chem. Mater.*, 2014, **26**, 6185–6192.
8 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, The Cambridge structural database, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
9 M. Woinska, S. Grabowsky, P. M. Dominiak, K. Wozniak and D. Jayatilaka, Hydrogen atoms can be located accurately and precisely by X-ray crystallography, *Sci. Adv.*, 2016, **2**, e1600192.
10 F. L. Hirshfeld, Can X-ray data distinguish bonding effects from vibrational smearing?, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1976, **32**, 239–244.
11 S. C. Capelli, H. B. Burgi, B. Dittrich, S. Grabowsky and D. Jayatilaka, Hirshfeld atom refinement, *IUCrJ*, 2014, **1**, 361–379.
12 A. Sturluson, M. T. Hyynh, A. R. Kaija, C. Laird, Y. Sunghyun, F. Hou, Z. Feng, C. E. Wilmer, Y. J. Colon, Y. G. Chung, D. W. Siderius and C. M. Simon, The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation, *Mol. Simul.*, 2019, **45**, 1082–1121.
13 S. Barthel, E. V. Alexandrov, D. M. Proserpio and B. Smit, Distinguising metal-organic frameworks, *Cryst. Growth Des.*, 2018, **18**, 1738–1747.
14 N. Ding, G. S. Armatas and M. G. Kanatzidis, Metal inorganic frameworks: dynamic flexible architecture with extended pore order built from $[Se_3]^{2-}$ linkers and $[Re_6Se_6Br_8]^{2-}$ clusters, *J. Am. Chem. Soc.*, 2010, **132**, 6728–6734.
15 S. Kumar, M. Samolia and T. J. D. Kumar, Hydrogen storage in Sc and Li decorated metal–inorganic framework, *ACS Appl. Energy Mater.*, 2018, **1**, 1328–1336.

16 D. Nazarian, J. S. Camp and D. S. Sholl, A comprehensive set of high-quality point charges for simulations of metal-organic frameworks, *Chem. Mater.*, 2016, **28**, 785–793.
17 D. Nazarian, J. S. Camp, Y. G. Chung, R. Q. Snurr and D. S. Sholl, Large-scale refinement of metal-organic framework structures using density functional theory, *Chem. Mater.*, 2017, **29**, 2521–2528.
18 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, Large-scale screening of hypothetical metal-organic frameworks, *Nat. Chem.*, 2012, **4**, 83–89.
19 B. J. Sikora, C. E. Wilmer, M. L. Greenfield and R. Q. Snurr, Thermodynamic analysis of Xe/Kr selectivity in over 137000 hypothetical metal-organic frameworks, *Chem. Sci.*, 2012, **3**, 2217–2223.
20 Y. J. Cui, Y. F. Yue, G. D. Qian and B. L. Chen, Luminescent functional metal-organic frameworks, *Chem. Rev.*, 2012, **112**, 1126–1162.
21 Q. Y. Yang, D. H. Liu, C. L. Zhong and J. R. Li, Development of computational methodologies for metal-organic frameworks and their application in gas separations, *Chem. Rev.*, 2013, **113**, 8261–8323.
22 E. Haldoupis, J. Borycz, H. L. Shi, K. D. Vogiatzis, P. Bai, W. L. Queen, L. Gagliardi and J. I. Siepmann, *Ab initio* derived force fields for predicting CO_2 adsorption and accessibility of metal sites in the metal-organic frameworks M-MOF-74 (M = Mn, Co, Ni, Cu), *J. Phys. Chem. C*, 2015, **119**, 16058–16071.
23 N. Lamia, M. Jorge, M. A. Granato, F. A. A. Paz, H. Chevreau and A. E. Rodrigues, Adsorption of propane, propylene and isobutane on a metal-organic framework: molecular simulation and experiment, *Chem. Eng. Sci.*, 2009, **64**, 3246–3259.
24 A. I. Skoulidas and D. S. Sholl, Self-diffusion and transport diffusion of light gases in metal-organic framework materials assessed using molecular dynamics simulations, *J. Phys. Chem. B*, 2005, **109**, 15760–15768.
25 S. Keskin, J. C. Liu, J. K. Johnson and D. S. Sholl, Testing the accuracy of correlations for multicomponent mass transport of adsorbed gases in metal-organic frameworks: diffusion of H_2/CH_4 mixtures in CuBTC, *Langmuir*, 2008, **24**, 8254–8261.
26 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, Dispersion-corrected mean-field electronic structure methods, *Chem. Rev.*, 2016, **116**, 5105–5154.
27 J. Hermann, R. DiStasio and A. Tkatchenko, First-principles models for van der waals interactions in molecules and materials: concepts, theory, and applications, *Chem. Rev.*, 2017, **117**, 4714–4758.
28 I. G. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*, John Wiley & Sons, West Sussex, England, 2006.
29 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
30 T. S. Gates, G. M. Odegard, S. J. V. Frankland and T. C. Clancy, Computational materials: multi-scale modeling and simulation of nanostructured materials, *Compos. Sci. Technol.*, 2005, **65**, 2416–2434.
31 T. Murtola, A. Bunker, I. Vattulainen, M. Deserno and M. Karttunen, Multiscale modeling of emergent materials: biological and soft matter, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1869–1892.
32 B. L. Bush and R. P. Sheridan, PATTY - a programmable atom typer and language for automatic classification of atoms in molecular databases, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 756–762.
33 J. K. Bristow, D. Tiana and A. Walsh, Transferable force field for metal-organic frameworks from first-principles: BTW-FF, *J. Chem. Theory Comput.*, 2014, **10**, 4644–4652.
34 L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier and V. Van Speybroeck, *Ab initio* parametrized force field for the flexible metal-organic framework MIL-53(Al), *J. Chem. Theory Comput.*, 2012, **8**, 3217–3231.
35 L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier and V. Van Speybroeck, QuickFF: a program for a quick and easy derivation of force fields for metal-organic frameworks from *ab initio* input, *J. Comput. Chem.*, 2015, **36**, 1015–1027.
36 D. Dubbeldam, K. S. Walton, D. E. Ellis and R. Q. Snurr, Exceptional negative thermal expansion in isoreticular metal-organic frameworks, *Angew. Chem., Int. Ed.*, 2007, **46**, 4496–4499.
37 J. Heinen and D. Dubbeldam, On flexible force fields for metal-organic frameworks: recent developments and future prospects, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1363.
38 S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy and R. Schmid, MOF-FF: a flexible first-principles derived force field for metal-organic frameworks, *Phys. Status Solidi B*, 2013, **250**, 1128–1141.
39 M. M. Ghahremanpour, P. J. van Maaren, C. Caleman, G. R. Hutchison and D. van der Spoel, 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.
40 K. M. Visscher and D. P. Geerke, Deriving force-field parameters from first principles using a polarizable and higher order dispersion model, *J. Chem. Theory Comput.*, 2019, **15**, 1875–1883.
41 J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, An empirical polarizable force field based on the classical Drude oscillator model: development history and recent applications, *Chem. Rev.*, 2016, **116**, 4983–5013.
42 J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, Current status of the AMOEBA polarizable force field, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.
43 T. A. Halgren and W. Damm, Polarizable force fields, *Curr. Opin. Struct. Biol.*, 2001, **11**, 236–242.

44 M. J. Van Vleet, A. J. Misquitta, A. J. Stone and J. R. Schmidtt, Beyond Born-Mayer: improved models for short-range repulsion in *ab initio* force fields, *J. Chem. Theory Comput.*, 2016, **12**, 3851–3870.

45 S. Vandenbrande, M. Waroquier, V. Van Speybroeck and T. Verstraelen, The monomer electron density force field (MEDFF): a physically inspired model for noncovalent interactions, *J. Chem. Theory Comput.*, 2017, **13**, 161–179.

46 L. P. Wang, J. H. Chen and T. Van Voorhis, Systematic parametrization of polarizable force fields from quantum chemistry data, *J. Chem. Theory Comput.*, 2013, **9**, 452–460.

47 B. Waldher, J. Kuta, S. Chen, N. Henson and A. E. Clark, ForceFit: a code to fit classical force fields to quantum mechanical potential energy surfaces, *J. Comput. Chem.*, 2010, **31**, 2307–2316.

48 C. Desgranges and J. Delhommelle, Many-body effects on the thermodynamics of fluids, mixtures, and nanoconfined fluids, *J. Chem. Theory Comput.*, 2015, **11**, 5401–5414.

49 J. G. McDaniel and J. R. Schmidt, First-principles many-body force fields from the gas phase to liquid: a "universal" approach, *J. Phys. Chem. B*, 2014, **118**, 8042–8053.

50 P. Kiss and A. Baranyai, Testing the recent charge-on-spring type polarizable water models. II Vapor-liquid equilibrium, *J. Chem. Phys.*, 2012, **137**, 194103.

51 C. A. Allen, J. A. Boissonnault, J. Cirera, R. Gulland, F. Paesani and S. M. Cohen, Chemically crosslinked isoreticular metal-organic frameworks, *Chem. Commun.*, 2013, **49**, 3200–3202.

52 R. Car and M. Parrinello, Unified approach for molecular dynamics and density-functional theory, *Phys. Rev. Lett.*, 1985, **55**, 2471–2474.

53 K. Laasonen, A. Pasquarello, R. Car, C. Lee and D. Vanderbilt, Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 10142–10153.

54 G. Kresse and J. Hafner, Abinitio molecular-dynamics for liquid-metals, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.

55 G. Kresse and J. Hafner, Ab-initio molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251–14269.

56 J. T. Horton, A. E. A. Allen, L. S. Dodda and D. J. Cole, QUBEKit: automating the derivation of force field parameters from quantum mechanics, *J. Chem. Inf. Model.*, 2019, **59**, 1366–1381.

57 P. Xu, E. B. Guidez, C. Bertoni and M. S. Gordon, Perspective: *ab initio* force field methods derived from quantum mechanics, *J. Chem. Phys.*, 2018, **148**, 090901.

58 B. Chen and J. I. Siepmann, Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of normal alkanes, *J. Phys. Chem. B*, 1999, **103**, 5370–5379.

59 S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schutt and K. R. Muller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.*, 2017, **3**, e1603015.

60 P. Maxwell, N. di Pasquale, S. Cardamone and P. L. A. Popelier, The prediction of topologically partitioned intra-atomic and inter-atomic energies by the machine learning method kriging, *Theor. Chem. Acc.*, 2016, **135**, 195.

61 V. Botu, R. Batra, J. Chapman and R. Ramprasad, Machine learning force fields: construction, validation, and outlook, *J. Phys. Chem. C*, 2017, **121**, 511–522.

62 P. Bleiziffer, K. Schaller and S. Riniker, Machine learning of partial charges derived from high-quality quantum-mechanical calculations, *J. Chem. Inf. Model.*, 2018, **58**, 579–590.

63 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

64 G. Kresse and J. Furthmuller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.

65 G. Kresse and J. Furthmuller, Efficiency of *ab initio* total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

66 G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.

67 P. E. Blochl, Projector augmented-wave method, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.

68 J. Hafner, Ab-initio simulations of materials using VASP: density-functional theory and beyond, *J. Comput. Chem.*, 2008, **29**, 2044–2078.

69 N. Gabaldon Limas and T. A. Manz, Introducing DDEC6 atomic population analysis: part 4. Efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more, *RSC Adv.*, 2018, **8**, 2678–2707.

70 H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1976, **13**, 5188–5192.

71 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma,

O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision B.01*, Gaussian, Inc., Wallingford CT, 2016.
72 A. D. Becke, Density-functional thermochemistry 3. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
73 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
74 T. A. Manz, T. Chen, D. J. Cole, N. Gabaldon Limas and B. Fiszbein, New scaling relations to compute atom-in-material polarizabilities and dispersion coefficients: part 1. Theory and accuracy, *RSC Adv.*, 2019, **9**, 19297–19324.
75 D. Rappoport and F. Furche, Property-optimized Gaussian basis sets for molecular response calculations, *J. Chem. Phys.*, 2010, **133**, 134105.
76 T. A. Manz and N. Gabaldon Limas, Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology, *RSC Adv.*, 2016, **6**, 47771–47801.
77 T. A. Manz and D. S. Sholl, Methods for computing accurate atomic spin moments for collinear and noncollinear magnetism in periodic and nonperiodic materials, *J. Chem. Theory Comput.*, 2011, **7**, 4146–4164.
78 T. C. Lillestolen and R. J. Wheatley, Redefining the atom: atomic charge densities produced by an iterative stockholder approach, *Chem. Commun.*, 2008, 5909–5911.
79 K. T. Tang and J. P. Toennies, An improved simple-model for the van der Waals potential based on universal damping functions for the dispersion coefficients, *J. Chem. Phys.*, 1984, **80**, 3726–3741.
80 K. T. Tang and J. P. Toennies, The damping function of the van der Waals attraction in the potential between rare-gas atoms and metal surfaces, *Surf. Sci.*, 1992, **279**, L203–L206.
81 T. A. Manz and T. Chen, New scaling relations to compute atom-in-material polarizabilities and dispersion coefficients: part 2. Linear-scaling computational algorithms and parallelization, *RSC Adv.*, 2019, **9**, 33310–33336.
82 G. Starkschall and R. G. Gordon, Calculation of coefficients in the power series expansion of the long-range dispersion force between atoms, *J. Chem. Phys.*, 1972, **56**, 2801–2806.
83 B. M. Axilrod and E. Teller, Interaction of the van der Waals type between three atoms, *J. Chem. Phys.*, 1943, **11**, 299–300.
84 E. T. Walters, M. Mohebifar, E. R. Johnson and C. N. Rowley, Evaluating the London dispersion coefficients of protein force fields using the exchange-hole dipole moment model, *J. Phys. Chem. B*, 2018, **122**, 6690–6701.
85 A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield and G. J. Martyna, Quantum Drude oscillator model of atoms and molecules: many-body polarization and dispersion interactions for atomistic simulation, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 144103.
86 T. W. Whitfield and G. J. Martyna, Low variance energy estimators for systems of quantum Drude oscillators: treating harmonic path integrals with large separations of time scales, *J. Chem. Phys.*, 2007, **126**, 074104.
87 M. Sadhukhan and F. R. Manby, Quantum mechanics of Drude oscillators with full Coulomb interaction, *Phys. Rev. B*, 2016, **94**, 115106.
88 T. A. Manz and D. S. Sholl, Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials, *J. Chem. Theory Comput.*, 2012, **8**, 2844–2867.
89 C. Campana, B. Mussard and T. K. Woo, Electrostatic potential derived atomic charges for periodic systems using a modified error functional, *J. Chem. Theory Comput.*, 2009, **5**, 2866–2878.
90 S. R. Cox and D. E. Williams, Representation of the molecular electrostatic potential by a net atomic charge model, *J. Comput. Chem.*, 1981, **2**, 304–323.
91 T. Watanabe, T. A. Manz and D. S. Sholl, Accurate treatment of electrostatics during molecular adsorption in nanoporous crystals without assigning point charges to framework atoms, *J. Phys. Chem. C*, 2011, **115**, 4824–4836.
92 N. Gabaldon Limas and T. A. Manz, Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials, *RSC Adv.*, 2016, **6**, 45727–45747.
93 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
94 T. A. Manz, Introducing DDEC6 atomic population analysis: part 3. Comprehensive method to compute bond orders, *RSC Adv.*, 2017, **7**, 45552–45581.
95 R. M. Hanson, Jmol - a paradigm shift in crystallographic visualization, *J. Appl. Crystallogr.*, 2010, **43**, 1250–1260.
96 Jmol: an open-source Java viewer for chemical structures in 3D, http://www.jmol.org, accessed August 2019.
97 T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier and P. W. Ayers, Minimal basis iterative stockholder: atoms in molecules for force-field development, *J. Chem. Theory Comput.*, 2016, **12**, 3894–3912.
98 B. Wang and D. G. Truhlar, Including charge penetration effects in molecular modeling, *J. Chem. Theory Comput.*, 2010, **6**, 3330–3342.
99 M. A. Freitag, M. S. Gordon, J. H. Jensen and W. J. Stevens, Evaluation of charge penetration between distributed multipolar expansions, *J. Chem. Phys.*, 2000, **112**, 7300–7306.
100 K. Vanommeslaeghe and A. D. MacKerell, Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing, *J. Chem. Inf. Model.*, 2012, **52**, 3144–3154.
101 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
102 J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov,

A. D. MacKerell and R. W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, *J. Phys. Chem. B*, 2010, **114**, 7830–7843.

103 C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, Lipid14: the Amber lipid force field, *J. Chem. Theory Comput.*, 2014, **10**, 865–879.

104 P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik and J. Vondrasek, Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and *ab initio* Hartree-Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree-Fock results, *J. Comput. Chem.*, 1997, **18**, 1136–1150.

105 B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, CHARMM: the biomolecular simulation program, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

106 R. Salomon-Ferrer, D. A. Case and R. C. Walker, An overview of the Amber biomolecular simulation package, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 198–210.

107 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, UFF: a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

108 D. E. Coupry, M. A. Addicoat and T. Heine, Extension of the universal force field for metal-organic frameworks, *J. Chem. Theory Comput.*, 2016, **12**, 5215–5225.

109 M. A. Addicoat, N. Vankova, I. F. Akter and T. Heine, Extension of the universal force field to metal-organic frameworks, *J. Chem. Theory Comput.*, 2014, **10**, 880–891.

110 Q. Xu and C. L. Zhong, A general approach for estimating framework charges in metal-organic frameworks, *J. Phys. Chem. C*, 2010, **114**, 5035–5042.

111 C. C. Zheng and C. L. Zhong, Estimation of framework charges in covalent organic frameworks using connectivity-based atom contribution method, *J. Phys. Chem. C*, 2010, **114**, 9945–9951.

112 A. R. Katritzky and R. D. Topsom, The sigma- and pi-inductive effects, *J. Chem. Educ.*, 1971, **48**, 427–431.

113 L. O. Smith and S. J. Cristol, *Organic Chemistry*, Reinhold Publishing Corporation, New York, 1966, p. 212.

114 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkins-Diehr, XSEDE: accelerating scientific discovery, *Comput. Sci. Eng.*, 2014, **16**, 62–74.

115 C. Altintas, G. Avci, H. Daglar, A. N. V. Azar, I. Erucar, S. Velioglu and S. Keskin, An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH_4 and H_2 adsorption, *J. Mater. Chem. A*, 2019, **7**, 9593–9608.

116 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, Development of a Cambridge Structural Database subset: a collection of metal-organic frameworks for past, present, and future, *Chem. Mater.*, 2017, **29**, 2618–2625.

117 C. Altintas, G. Avci, H. Daglar, A. N. V. Azar, S. Velioglu, I. Erucar and S. Keskin, Database for CO_2 separation performances of MOFs based on computational materials screening, *ACS Appl. Mater. Interfaces*, 2018, **10**, 17257–17268.