Supporting information for: On the Force Field Prediction of Materials Properties in Metal Organic Frameworks

Peter G. Boyd,[†] Seyed Mohamad Moosavi,[†] Matthew Witman,[‡] and Berend Smit[*,†,‡]

[†]*Laboratory of Molecular Simulation, Institut des sciences et ingénierie chimiques, École polytechnique fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Valais, Switzerland*

[‡]*Department of Chemical and Biomolecular Engineering, University of California, Berkeley 94720, USA*

E-mail: berend.smit@epfl.ch

Phone: +41 21 693 0079

Contents

S.1 Atomic Partial Charge Assignment

Three charge assignment schemes were studied for those force fields which do not explicitly provide charges for their atom types. The following subsections present the details of charge assignment used in UFF,[S1] UFF4MOF[S2] and DREIDING[S3] force fields.

S.1.1 Charge Equilibration (QEq) method

The charge equilibration method[S4] was used to assign partial atomic charges to the atoms of IRMOF-1, IRMOF-10, HKUST-1 and UiO-66. For this purpose, the QEQ implementation in LAMMPS was used, with a relative convergence tolerance of 10^{-6}, a global cutoff of 12 Åfor the interactions, and a maximum number of iterations of 200. This calculation was performed only once with the initial crystal structure of each material, so that a comparison of the charges could be made with the other charge generation methods using the same atomic

configuration. The other methods require expensive periodic DFT calculations, making them impractical for repeated dynamic charge determination during a simulation.

The QEq parameters for each atom type A, the so-called 'idempotential' J^0_{AA} and 'electronegativity' χ_A terms were taken from the UFF values found in the GULP simulation package.[S5]

S.1.2 REPEAT and DDEC charge methods

The *ab initio* electrostatic potential and charge densities were determined with the Vienna Ab initio Simulation Package (VASP) version 5.3.5.[S6–S8] The Projector Augmented Wave[S9] method was used to model core electrons. The brillouin zone was modelled using the gamma point. A kinetic energy cutoff of 520 eV was used for the determination of the number of plane waves in the basis set. The electronic ground state was considered converged when the difference in energy was less than 10^{-5} eV between self-consistent loops. In all cases the PBE functional was used for exchange and correlation.[S10,S11] The REPEAT[S12] and DDEC6 methods[S13] were used to compute the partial atomic charges for IRMOF-1, IRMOF-10, HKUST-1, and UiO-66 using the default parameters and the electrostatic potentials and charge densities found from the VASP calculations.

S.2 Computing the bond topologies in MOFs

Bonding in each MOF, if not provided in the .cif file, was calculated from the sum of covalent radii between each atom. If two atoms lie within

$$R_{IJ} \leq \frac{(\sigma_I + \sigma_J)}{\lambda} \tag{S1}$$

where R_{IJ} is the distance between atoms I and J and σ_I, σ_J are the covalent radii of these atoms as reported by Cordero *et al.*[S14] λ represents a scaling constant which was set to $\lambda = 0.9$ for all organic interactions, and $\lambda = 0.85$ for all potential oxygen - metal bonds.

Each initial bond that is found this way is given an order of 1, and subsequent evaluations of the chemical environment are performed to assign bonds with higher bond orders if necessary. We note that in UFF4MOF, certain metal - oxygen bond orders and metal - metal bond orders are less than 1.[S2] These special cases are recognized when the force field is requested, further details are provided in Sections S.3.4 and S.8 below.

The evaluation of bond order procceeds through a series of molecular recognition cases. At the outset, each atom is initially 'typed' based on it's coordination environment. For example carbon atoms involved in four bonds will be classified as **sp3**, three as **sp2**, and two as **sp**. Following this, if two **sp2** hybridized carbon or nitrogen atoms are bonded, and share a ring of 6 **sp2** hybridized atoms or less, the bond order will be set to 1.5. If they are on separate rings, the bond order will remain 1 to indicate a biphenyl bridge, for example. In all cases, we have set the carboxylate carbon - oxygen bond orders to 1.5, so as to mimic an evenly distributed electron delocalized system.

Recognition of aliphatic unsaturated hydrocarbon moieties follow a rudimentary procedure, where double bonds are found if the distance between **sp2** carbon atoms is 5% shorter than their covalent radii. Alternating single and double bonds are found based on this initial criteria. This condition was set in the python interface for general purpose use, as none of the materials studied in this work possess aliphatic chains. Likewise, recognition of esters, amines, and ether groups follow a procedure of determining neighbouring atoms and their hybridizations.

S.3 Force Field Implementations in LAMMPS

Provided below are a description of the potentials used in each force field studied in this work, and how they were implemented in LAMMPS. What is **not** described are the actual parameter values and their associated atom types as, for the most part, these are described in the original articles. As a consolation, we have provided a LAMMPS input for each MOF

and each force field that accompanies this text and the main article.

S.3.1 UFF

The potentials defined in the original paper by Rappé *et al.*[S1] are available in the LAMMPS package, thus one can produce a faithful adaptation without the need to resort to any approximate functions, as is the case for the UFF implementation in GROMACS.[S15] We note that in many cases, the coordination environment found around metal ions in MOFs are not properly described by UFF. For example the Cu parameters in UFF are designed for a tetrahedral-like coordination, which does not support the square planar geometries found in HKUST-1. We have therefore included the option to 'fix' the metal geometries where the potential terms describing metal atoms and its neighbors are adjusted from their default values prescribed by the UFF recipe. Specifically, the equilibrium angles and distances for the bond stretching, angle bending, and torsional terms are changed to match the crystallographic geometries of the atoms, while the coefficient values remain the same. This seemingly *ad-hoc* adjustment of the UFF parameters has been shown to provide more accurate minimum energy structures of a range of MOFs.[S15] Unless explicitly stated, all of the values computed with the UFF parameters in this work have added this fix for the metal ions.

To achieve a description of the UFF in LAMMPS, one must ensure that the LAMMPS version is compiled with 'USER_MISC' turned on. This feature contains the angle functions needed to describe accurately the bend energy in the UFF force field. The potentials are discussed below.

S.3.1.1 Bond stretching

$$E_{stretch} = \frac{K_{IJ}}{2}\left(R_{IJ} - R_0\right)^2 \tag{S2}$$

Where R_{IJ} is the distance between atoms I and J, K_{IJ} is the bond strength and R_0 is the equilibrium bond distance. In LAMMPS this functional form is provided as the 'harmonic'

bond style. The only difference between the lammps 'harmonic' function and Equation S2 is that in LAMMPS, the K value is *not* divided by 2 internally. Thus we follow the recipe provided by Rappé *et al.*[S1] to derive the R_0 and K values, and divide K by 2 before writing them to the LAMMPS input file.

S.3.1.2 Angle bending

For atoms I and K bonded to the same atom J, the following cosine fourier expansions in θ hold for UFF:

$$E_{bend} = \frac{K_{IJK}}{n^2}\left[1 - cos\left(n\theta\right)\right] \tag{S3}$$

$$E_{bend} = K_{IJK}\left[C_0 + C_1 cos\left(\theta\right) + C_2 cos\left(2\theta\right)\right] \tag{S4}$$

Where Equation S3 applies to atoms in linear, trigonal-planar, square-planar, or octahedral arrangement, and Equation S4 can be used for the general non-linear case. Equations S3 and S4 are available in LAMMPS as the angle styles 'cosine/periodic' and 'fourier', respectively. When deriving the coefficients C_0, C_1, and C_2 in Equation S4, one can define any equilibrium geometry given an equilibrium angle, θ_0, and deriving the coefficients with the following recipe, detailed in the original UFF article,[S1]

$$\begin{aligned} C_2 &= \frac{1}{4sin\left(\theta_0\right)} \\ C_1 &= -4C_2 cos\left(\theta_0\right) \\ C_0 &= C_2\left(2cos\left(\theta_0\right)^2 + 1\right) \end{aligned} \tag{S5}$$

Equations S4 and S5 were used in this study to fix the metals to their crystallographic geometries, where equilibrium angles were determined between all atoms with a central metal ion using the following formula.

$$\theta_0 = cos^{-1}\left(\frac{\vec{r}_{JI} \cdot \vec{r}_{JK}}{|\vec{r}_{JI}| \cdot |\vec{r}_{JK}|}\right) \tag{S6}$$

where the $\vec{r}_{JI}$ and $\vec{r}_{JK}$ terms denote the vectors pointing away from the central J atom to the I,K atoms forming the angle (adjusting for periodic boundary conditions). The factors affecting the amplitude were otherwise unadjusted from the original UFF parameters. We recognize that this is an *ad hoc* adjustment to an empirical force field, however performing this adjustment permitted an expansion of UFF to a more diverse range of metal organic framework materials.

S.3.1.3 Dihedral torsion

The dihedral potential in UFF is of the form,

$$E_{torsion} = \frac{V_{IJKL}}{2}\left[1 - cos\left(n\phi_0\right) cos\left(n\phi\right)\right] \tag{S7}$$

for atoms I, J, K, and L, with the torsion angle ϕ being determined between the planes defined by atoms I, J, K, and J, K, L. This function is available in LAMMPS as the dihedral style 'harmonic'. Note that UFF specifically states there should be no dihedral potentials defined for metal ions in the J, or K positions, thus *ad hoc* adjustments to the dihedral potentials are unnecessary to fix the metal ions in metal organic frameworks to their crystallographic positions.

S.3.1.4 Improper torsion

The improper out-of-plane bending potential in UFF is defined by the following function,

$$E_{inversion} = K_{IJKL}\left[C_0 + C_1 cos\left(\omega\right) + C_2 cos\left(2\omega\right)\right] \tag{S8}$$

for atoms I, K, L, all connected to the same central atom, J. This is available in LAMMPS

as the improper style 'fourier'. No adjustments to the values of the parameters were necessary to fix the metal geometries in their crystallographic positions, as no inversion potentials were parameterized for the transition metals in UFF.

S.3.1.5 Non-bonded interactions

$$E_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^{6} \right] \tag{S9}$$

The non-bonded van der Waals interactions are modelled with the Lennard-Jones potential function available in LAMMPS, where each of the ϵ and σ parameters were mixed with the Lorentz-Berthelot mixing rules. A tail correction was applied to the van der Waals term to ensure continuity past the non-bonded cutoff (set to 12.5 Å in this work) where it is assumed that the system reverts to a bulk homogeneous liquid.

Where appropriate, if charges were included in the simulation, an Ewald summation scheme was used to account for long range coulombic interactions. The precision was set to 1×10^{-6}.

Rappé *et al.* mention that bonded atoms and atoms bonded to a common neighbour should not interact via non-bonded forces.[S1] This can be interpreted that 1-4 non-bonded interactions should be turned 'on', which was done in this study.

S.3.2 DREIDING

While the DREIDING force field does not span the same breath of the periodic table as does UFF, it contains parameters for tetrahedrally coordinated Zn ions, and thus should be used only with the IRMOF series. However, it was somewhat straight-forward to extend the parameterization to support square planar Cu containing MOFs due to simple philosophy of the DREIDING parameterization.

The DREIDING Cu atom type has a bond radius of $R_1 = 1.302$, taken directly from the UFF Cu3+1 parameters. The Lennard-Jones ϵ and σ parameters were also adopted from the

UFF Cu3+1 parameters to describe non-bonded dispersion interactions in this force field.

S.3.2.1 Bond stretching

In the current study, bonding was modeled with the harmonic form of Equation S2. DREIDING bonds are extremely straight-forward to implement, where the force constants are determined solely from the bond order, ignoring any differences in chemistry between atoms. The spring constant is derived as follows,

$$K = O_{bond} \times 700.0 \left(\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}\right) \tag{S10}$$

where the bond order, O_{bond} is 1 for single, 1.5 for aromatic, 2 for double, and 3 for triple bonds.

We note that the authors of the DREIDING paper[S3] provide a choice of using either a harmonic bond potential or a more accurate Morse potential, along with the suggestion that the Morse potential be used for more refined calculations near energetic minima. In the bulk modulus calculations reported in the main text, we have used the harmonic form of the bonding potential, as the Morse potential yielded poor fits to the Murnaghan equation of state.

S.3.2.2 Angle bending

The angle bending term between atoms I, J, and K is described by a harmonic cosine function,

$$E_{bend} = \frac{C_{IJK}}{2} \left[cos\left(\theta\right) - cos\left(\theta_0\right)\right]^2 \tag{S11}$$

where the equilibrium angle, θ_0, is determined from the DREIDING atom type for the central atom J. The values for θ_0 are listed in the original article,[S3] whereas the value for C_{IJK} is obtained from the formula

$$C_{IJK} = \frac{100}{2sin\left(\theta_0\right)^2} \tag{S12}$$

which, with the exception of θ_0, is generic for all angles, regardles of atomic identity. Equation S11 is available in LAMMPS as the angle style 'cosine squared'.

$$E_{angle} = \frac{K_{IJK}}{n^2} \left[1 - B\left(-1\right)^n cos\left(n\theta\right)\right] \tag{S13}$$

To extend DREIDING to apply to HKUST-1, the Cu angle potential was switched to an alternate cosine function (Equation S13) from the harmonic function (Equation S11) described in the original article,[S3] which can possesses minima at both 90° and 180°, ideal for Cu paddlewheel configurations. To achieve a minima at 90° and 180°, the parameters in Equation S13 were set to $B = 1$, and $n = 4$.

We have chosen to keep the general force constant the same for the cosine function in Equation S13, which effectively reduces the potential barrier by roughly 1/5, while preserving the curvature of the function near the minima. The value of K was set to 100 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}/n^2$. LAMMPS supports the potential described by Equation S13 as the 'cosine periodic' dihedral style. We note that LAMMPS divides K by n^2 internally, so the 'raw' K value in the input file is 100 in all cases.

Due to the markedly different coordination environment of the Zr ions in UiO-66, the DREIDING force field was not extended to this material.

S.3.2.3 Dihedral torsion

The angle ϕ between planes defined by atoms I and L adjacent to bonded atoms J and K is of the form,

$$E_{angle} = \frac{K_{IJKL}}{2} \left[1 - cos\left(n\left(\phi - \phi_0\right)\right)\right] \tag{S14}$$

where K_{IJKL} is determined from a series of cases outlined in the article by Mayo *et al.*[S3] A similar functional form to Equation S14 is available in LAMMPS as the dihedral style 'charmm',

$$E_{angle}^{charmm} = K_{IJKL}\left[1 + cos\left(n\phi - d\right)\right] \tag{S15}$$

Here the value for d is computed as $n \cdot \phi_0$ and shifted by $+\pi$, while K_{IJKL} is divided by 2 in order to achieve the intended functional form of the DREIDING authors.[S3] No dihedral potentials are defined for metallic elements, thus no additional considerations were required for the modified Cu parameters in the DREIDING force field.

S.3.2.4 Improper torsion

In the DREIDING force field, the inversion potential for a central atom J bonded to three neighbours I, K, L is defined by

$$E_{inversion} = \frac{K_{IJKL}}{2}\left[cos\left(\omega\right) - cos\left(\omega_0\right)\right]^2 \tag{S16}$$

for an angle ω formed between vector $\vec{JI}$ and the plane defined by JKL. This potential is provided in the LAMMPS package as the improper style 'umbrella'.

S.3.2.5 Non-bonded interactions

Like UFF, the non-bonded interactions are treated with the Lennard-Jones potential presented in Equation S9 and the coulombic potential determined from an Ewald summation. The same cutoff and precisions were used for DREIDING as described in Section S.3.1.5. In all simulations with the DREIDING force field, we have turned the nonbonded interactions on for atoms in a 1-4 configuration. All closer interactions were turned off, as is suggested in the DREIDING paper.[S3]

S.3.3 BTW-FF

Bristow *et al.*[S16] recently introduced BTW-FF, a full atom forcefield parametrized specifically to study the structural and mechanical properties of MOFs. In that work, periodic *ab initio*

calculations of six MOFs, including MOF-5, IRMOF-10, HKUST-1, UIO-66 and UIO-67 are used to fit the forcefield parameters in functions similar to those in the MM3 force field.[S17] In fact, the original MM3 force field was used as a jumping-off point for their parameterization of their MOF force field, using the pre-defined terms and functions available in TINKER.[S18]

TINKER is one of the few packages that includes all the MM3 complex formalism entirely. Since LAMMPS lacks some of the functional forms of MM3, transferring the parameters and functionals from TINKER to LAMMPS requires a few approximations which will be discussed in the following subsections. The original parameter values were reported with energy units of mdyne. They were converted to kcal/mol by multiplication of the factor 143.88 to recover the MM3 potentials in LAMMPS, explaining this values presence in several equations below.

S.3.3.1 Bond strecthing

The bond stretching potential in BTW-FF was adopted from the MM3 force field,

$$E_{stretch}^{MM3} = (143.88)\,\frac{k_s}{2}\,(R_{ij} - R_0)^2\left[1 - 2.55\,(R_{IJ} - R_0) + (7/12)\,(2.55)^2\,(R_{ij} - R_0)^2\right] \quad \text{(S17)}$$

The potential can be found in LAMMPS as the 'class 2' bond style which has a slightly different form

$$E_{stretch}^{class2} = K_2\,(R_{ij} - R_0)^2 + K_3\,(R_{ij} - R_0)^3 + K_4\,(R_{ij} - R_0)^4 \quad \text{(S18)}$$

for atoms I and J separated at a distance R_{IJ}, with equilibrium bond distance R_0. The K_2, K_3, and K_4 values determine the curvature of the quadratic, cubic, and quartic terms of the potential and are extracted from a single coefficient value, k_s, provided in the BTW force

field. The original BTW potential is restored from the coefficient k_s as follows;

$$\begin{aligned} K_2 &= \frac{k_s}{2}\,(143.88) \\ K_3 &= \frac{k_s}{2}\,(143.88)\,(-2.55) \\ K_4 &= \frac{k_s}{2}\,(143.88)\,(2.55)^2\,(7/12) \end{aligned} \tag{S19}$$

This ensures that the root of this polynomial equation is solely located at R_0.

S.3.3.2 Angle bending

To describe angle bending, the BTW force field uses a sixth order polynomial

$$\begin{aligned} E_{bend}^{MM3} = 0.021914\,(k_\theta)\,(\theta-\theta_0)^2\,[1 - 0.014\,(\theta-\theta_0) + \left(5.6\times10^{-5}\right)(\theta-\theta_0)^2 - \\ \left(7.0\times10^{-7}\right)(\theta-\theta_0)^3 + \left(9.0\times10^{-10}\right)(\theta-\theta_0)^4] \end{aligned} \tag{S20}$$

With a force constant k_θ between atoms I,J, and K, with an equilibrium angle, θ_0. LAMMPS supports polynomials for angle bending up to the fourth order with the 'class2' angle style,

$$E_{bend}^{class2} = K_2^{IJK}\,(\theta-\theta_0)^2 + K_3^{IJK}\,(\theta-\theta_0)^3 + K_4^{IJK}\,(\theta-\theta_0)^4 \tag{S21}$$

As a result, the fifth and sixth order terms were ignored in our implementation. Ignoring these terms may seem like a severe approximation, however the relative error in the angle bending energy is less than 10% for bending up to 45° and less than 1% up to 20° for each term. Further discussion is presented in Section S.3.3.6.

This approximation will clearly break-down in simultions of more extreme conditions using the LAMMPS implementation of BTW-FF, where strong deviations from equilibrium angles is expected. In any case, the forcefield parameters were derived from ground state calculations, thus it will be challenging for the original BTW force field implementation to produce accurate data at extreme temperatures or pressures.

The force constants K_n^{IJK} are recovered from the single force constant, k_θ, provided by

the BTW force field using the following conversions. Note, the original values are provided in units of mdyn and degrees, while the LAMMPS force constants are reported in kcal and radians.

$$\begin{aligned} K_2^{IJK} &= \frac{k_\theta}{2}\,(143.88) \\ K_3^{IJK} &= \frac{k_\theta}{2}\,(143.88)\,(-0.014)\,(\pi/180)^{-1} \\ K_4^{IJK} &= \frac{k_\theta}{2}\,(143.88)\left(5.6\times10^{-5}\right)(\pi/180)^{-2} \end{aligned} \tag{S22}$$

When implementing this force field, we discovered that the angle bending terms are not well described for square planar coordination environments, such as the copper paddlewheel ($Cu_2(O_2C)_4$). To accurately describe the 90° angle between coordinating oxygens in these species, two minima must exist at 90° and 180°. A polynomial with a single minima at 90°, such as that in Equation S21 where $\theta_0 = 90$, will not be able to reproduce this and results in additional forces on the oxygen atoms 180° from each other. As a consequence, *only for the Cu atoms in HKUST-1* we have replaced Equation S21 with a fourier type angle bending term ideal for square-planar coordination environments, found in Equation S13. Additionally, the force constant k_θ for the Cu paddlewheel is 7.191 kcal/mol/rad^2 in the BTW force field, however we suspect this value is small due to the large amount of error observed when this potential is in the range of more physically motivated values, thus we have we adopted the coefficient from MOF-FF,[S19] which used the same potential as in Equation S13, where $n = 4$. This value of K_{IJK} was set to 126.64 kcal/mol for the BTW O-Cu-O angle bending potential in LAMMPS.

Exceptionally, the MM3 force field includes special energy functions called 'bend-stretch' and 'torsion-stretch' terms. These allow for the synchronious movements of atoms bonded to a common neighbour, as observed in molecular vibrational spectra. The authors of the BTW force field did not report any parameters for these extra terms. After comparing calculations between TINKER and LAMMPS, we discovered that the coefficients for these extra functions were set to 0, thus they were not included in the BTW implementation in

LAMMPS.

S.3.3.3 Dihedral torsion

Aside from the bond function, the dihedral potential is the only other MM3 intramolecular potential that is faithfully supported in LAMMPS, under the dihedral style 'fourier'. For the angle ϕ between planes formed by atoms I, J, K, and J, K, L the potential is

$$E_{inversion} = \frac{K_1^{IJKL}}{2}(1 + cos\phi) + \frac{K_2^{IJKL}}{2}(1 - cos2\phi) + \frac{K_3^{IJKL}}{2}(1 + cos\phi) \tag{S23}$$

The only non-zero force constant reported in the BTW force field is for K_2^{IJKL}, meaning only the second term in Equation S23 is expressed in energy evaluations. In LAMMPS this value is not divided by 2 internally, so the force constants are halved before writing the necessary input files.

S.3.3.4 Improper torsion

Like the angle bending term, the improper torsion terms in BTW-FF were described with a 6-order polynomial identical to that described in Equation S20. Our implementation uses a harmonic term, again because LAMMPS does not support 6-order polynomials for improper torsions. In all cases examined this improper torsion term is used to correct for the anti-planar behaviour of aromatic systems during a simulation. Thus replacing the polynomial with a harmonic term, while severe, will have minimal effect on near-equilibrium simulations. In simulations of extreme conditions or strain, replacing a sixth-order term with a harmonic function will introduce significant energy deviations, and is a potential source of error in our calculations.

$$E_{inversion}^{class2} = K_{IJKL}\left[\left(\frac{\omega_{IJKL} + \omega_{KJIL} + \omega_{LJKI}}{3}\right) - \omega_0\right]^2 \tag{S24}$$

where the first term inside the brackets is an average over all out-of-plane angles formed between atom J's neighbours; I, K, and L. When predicting the movement of neighbours around a central atom, the MM3 force field introduced an additional interaction term titled 'Bend-Bend'. This involves correlated movements of the angles, and is characterized by the potential

$$\begin{aligned} E_{bend-bend} =& M_1 \left(\theta^{IJK} - \theta_0^{IJK}\right)\left(\theta^{KJL} - \theta_0^{KJL}\right) + \\ & M_2 \left(\theta^{IJK} - \theta_0^{IJK}\right)\left(\theta^{IJL} - \theta_0^{IJL}\right) + \\ & M_3 \left(\theta^{IJL} - \theta_0^{IJL}\right)\left(\theta^{KJL} - \theta_0^{KJL}\right) \end{aligned} \tag{S25}$$

Where the θ_0^{IJK} terms describe equilibrium angle values for the angle between atoms I, J, and K, for example. The M_n terms describe the degree of correlation between the pairwise angles specified on each line in Equation S25. BTW-FF includes non-zero Bend-Bend terms, thus we have implemented them in LAMMPS under the 'angle-angle' function provided as part of the improper style 'class2'. Some of these terms were not included in the BTW article but were implicitly used in the TINKER program, so we have provided a comprehensive list in table S1 of force field parameters used in our LAMMPS implementation of the force field. The greyed out values in table S1 were not provided by Bristow *et al.*[S16] but were extracted from the original MM3 parameters.[S17]

The atom labels provided in the first column of table S1 can be described as follows. C_{benz} are aromatic carbons bonded to a hydrogen and two other aromatic carbons, C_α are aromatic carbons bonded to a carboxylate carbon, C_γ are aromatic carbons bonded to three aromatic carbons, one of which is on a different phenyl ring and bonded by a single bond (e.g. the carbon atoms that bridge benzene rings in biphenyl), H_{benz} are hydrogen atoms bonded to aromatic carbons, C_{carbox} are carboxylate carbons bonded to two oxygens, O_{carbox} are carboxylate oxygens bonded to a carbon and one metal ion, finally O_{inorg} are oxygen atoms bonded only to to Zr ions.

The constants provided in Table S1 have been converted from the original mdyne to

Table S1: Force Field Parameters Used in Inversion and Angle-Angle Terms for BTW-FF.

Atoms	BTW type	K_{IJKL}[a]	$\omega_0(rad)$	M_1[b]	M_2[b]	M_3[b]
C_α-C_{benz}-C_{benz}-H_{benz}	902-912-912-915	0.0	0.0	-17.265	-21.582	0.0
C_α-C_{benz}-H_{benz}-C_{benz}	902-912-915-912	0.0	0.0	-17.265	-21.582	0.0
C_α-C_{benz}-C_α-H_{benz}	902-912-902-915	0.0	0.0	-17.265	-21.582	0.0
C_α-C_{benz}-H_{benz}-C_α	902-912-915-902	0.0	0.0	-17.265	-21.582	0.0
C_γ-C_{benz}-H_{benz}-C_{benz}	903-912-915-912	14.388	0.0	-17.265	-21.582	0.0
C_γ-C_{benz}-C_{benz}-H_{benz}	903-912-912-915	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_{benz}-H_{benz}-C_γ	912-912-915-903	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_{benz}-C_γ-H_{benz}	912-912-903-915	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_{benz}-C_α-H_{benz}	912-912-902-915	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_{benz}-H_{benz}-C_α	912-912-915-902	14.388	0.0	-17.265	-21.582	0.0
H_{benz}-C_{benz}-C_{benz}-C_α	915-912-912-902	7.9134	0.0	-17.265	-21.582	0.0
H_{benz}-C_{benz}-C_α-C_{benz}	915-912-902-912	7.9134	0.0	-17.265	-21.582	0.0
H_{benz}-C_{benz}-C_α-C_α	915-912-902-902	7.9134	0.0	-17.265	-21.582	0.0
H_{benz}-C_{benz}-C_{benz}-C_γ	915-912-912-903	7.9134	0.0	-17.265	-21.582	0.0
H_{benz}-C_{benz}-C_γ-C_{benz}	915-912-903-912	7.9134	0.0	-17.265	-21.582	0.0
O_{carbox}-C_{carbox}-O_{carbox}-C_α	170-913-170-902	107.91	0.0	0.00	0.00	0.0
O_{carbox}-C_{carbox}-C_α-O_{carbox}	170-913-902-170	107.91	0.0	0.00	0.00	0.0
C_α-C_{carbox}-O_{carbox}-O_{carbox}	902-913-170-170	0.0	0.0	0.00	0.00	0.0
C_{benz}-C_α-C_{benz}-C_{carbox}	912-902-912-913	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_α-C_{carbox}-C_{benz}	912-902-913-912	14.388	0.0	-17.265	-21.582	0.0
C_{carbox}-C_α-C_{benz}-C_{benz}	913-902-912-912	14.388	0.0	-17.265	-21.582	0.0
C_γ-C_γ-C_{benz}-C_{benz}	903-903-912-912	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_γ-C_{benz}-C_γ	912-903-912-903	14.388	0.0	-17.265	-21.582	0.0
C_{benz}-C_γ-C_γ-C_{benz}	912-903-903-912	14.388	0.0	-17.265	-21.582	0.0
Zr-O_{inorg}-Zr-Zr	192-171-192-192	143.88	0.0	0.00	0.00	0.0

a - $kcal \cdot mol^{-1} \cdot deg^{-2}$
b - $kcal \cdot mol^{-1} \cdot \text{Å}^{-1}$

kcal/mol using the multiplicative constant 143.88. In addition, when listing coefficients in the LAMMPS input file, each of the values listed in table S1 for M_1, M_2, and M_3 were divided by a factor of 3 to eliminate the effects of triple-counting angles in this equation (to be clear, the values in table S1 are *not* divided by 3).

S.3.3.5 Non-bonded interactions

The van der Waals interactions are modelled with a Buckingham potential in BTW-FF

$$E_{buck} = Ae^{-R_{IJ}/\rho} - \frac{C}{R_{IJ}^6} \tag{S26}$$

where the non-bonded energy between atoms I and J at a distance R_{IJ} is determined by the coefficients A, ρ and C. This is provided in the LAMMPS program as the pair style 'buck/coul/long'. We have also included coulombic interactions via the Ewald summation in LAMMPS with a precision of 10^{-6}, and the dielectric constant was set to 1.5 in accordance with their force field. Tail corrections were included past the cutoff value of 12.5 Å, and 1-4 bonded atom interactions were included in the non-bonded energy calculations.

We have discovered a complication when implementing these non-bonded terms, where the MM3 force field shifts the van der Waals interaction site of hydrogen. Here, all bonded parameters for C-H interactions remain the same, however hydrogen's non-bonded interaction site is shifted to a point along the bond from the hydrogen to carbon, resulting in a smaller inter-molecular radius of organic molecules. Such adjustments are done automatically when the MM3 force field is requested in TINKER (and, by association, the BTW force field). We note that Bureekaew *et al.*[S19] suggested a remedy of this complexity by simply reducing the vdW radii of hydrogen. However, in this work we have ignored this change and keep the original vdW radii of hydrogen. This has had a negligible effect on the results, including cell parameter and thermal expansion coefficient calculations.

S.3.3.6 BTW differences and justification

To justify and have an estimate on the consequences of the aforementioned approximations, we compared, for IRMOF-1, the various contributions to the total energy in LAMMPS with the BTW-FF implementation in TINKER. The reference values were obtained by simulating the desired properties using DYNAMIC, MINIMIZE and ANALYZE packages of TINKER and the reported input files of TINKER by Bristow *et al.*[S16] Table S2 shows the differences in energies after a minimization of the geometry from each code.

Table S2: Energetic Contributions Between BTW Implementations in LAMMPS and BTW-FF for IRMOF-1. Units are kcal/mol.

Program	Bond	Angle	Dihedral	Improper	Non-bonded	Total
LAMMPS	537.2	53.8	259.0	3.93	-23579.3	-22650.2
TINKER	552.7	56.7	258.8	0.0	-23605.9	-22793.7

The reported individual energy terms between LAMMPS and TINKER shows good quantitative agreement, suggesting there is no cancellation of errors between the two implementations. Most pronounced are the differences between the non-bonded interactions, where the BTW force field in TINKER includes charge-dipole and dipole-dipole interactions. No terms were provided for dipole interactions in the original article, and as such they were not included in the LAMMPS implementation of BTW-FF. These terms contribute rougly 170 kcal/mol to the non-bonded energy in TINKER, however we note that they do not contribute significantly to the behaviour of IRMOF-1. As table S3 shows, the dipole terms constitue 0.7% of the overall electrostatic interactions and do not change by more than 13% after minimization. In addition the values reported in the main text for linear coefficients for thermal expansion and bulk modulus agree well with the BTW author's data.[S16]

Table S3: Energetic Contributions for IRMOF-1 in TINKER.

Potential	Initial Energy (kcal/mol)	Post-optimized Energy (kcal/mol)	ΔE (% change)
Bond	863.30	552.70	−36.00
Angle	54.00	56.70	5.00
Angle-Angle	0.03	−0.04	−233.00
Torsion	258.90	258.30	−0.23
van der Walls	125.50	119.30	−4.90
Charge-Charge	−23 562.50	−23 605.90	−0.18
Charge-Dipole	−214.20	−219.10	−2.30
Dipole-Dipole	48.80	43.80	−10.20

We consider the approximation of the BTW force field in LAMMPS to be faithful when considering the approximations made with some of its functional forms. For example Figure S1 shows the angle deviation distributions of all atoms in IRMOF-1 over a 2 ns NPT simulation. It is clear that even in the extreme cases the energetic deviation is no more than 2% of the total angle energy between the original MM3-type angle (green) and the 4^{th} order potential shown in Equation S21 (blue). The distribution of IRMOF-1 angles is represented in the blue gaussian-like distribution in the centre of the image.

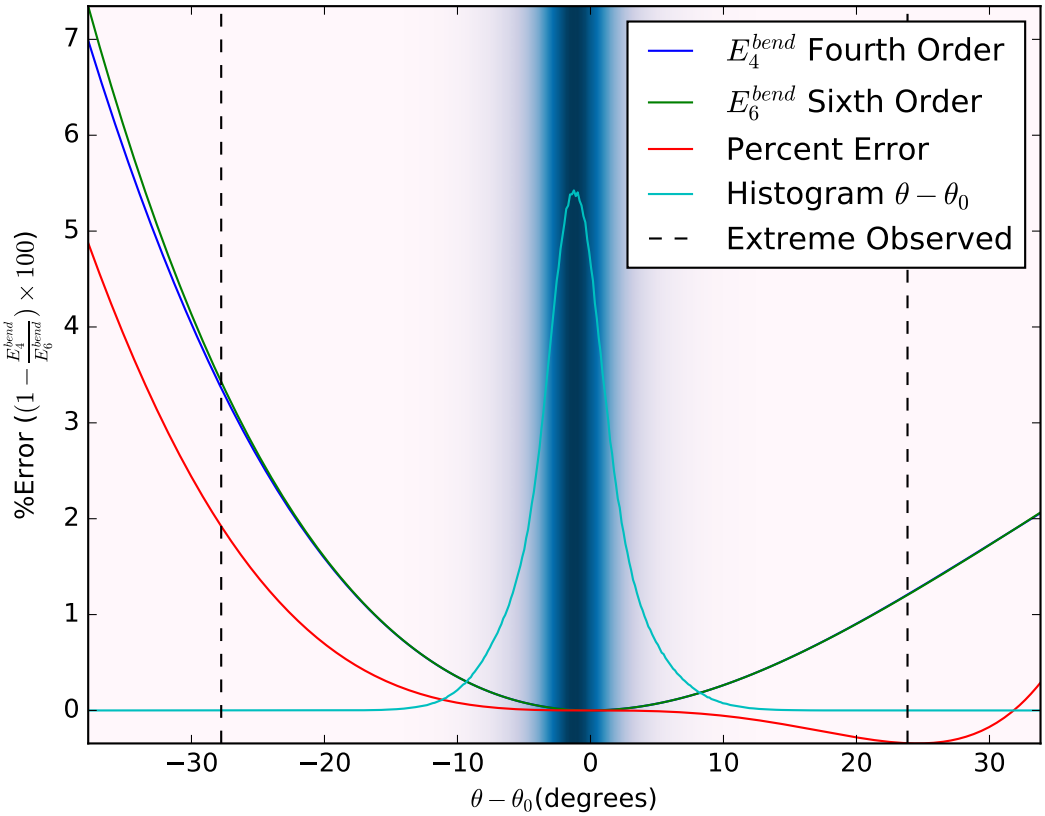

Figure S1: Angle distribution from a 2 ns NPT simulation in LAMMPS for IRMOF-1 using the BTW-FF. The blue curve plots the angle potential as represented in LAMMPS, while green is the original MM3 angle shown in Equation S20. The red curve shows the % error between these two potentials as a function of the angle deviation.

S.3.4 UFF4MOF

Addicoat *et al.* recently ammended the UFF parameters with the explicit purpose of being able to reproduce experimental MOF crystallographic data.[S2] They adjusted the covalent radii parameters of the relevant metal atoms such that, upon optimization, metal-ligand bond distances matched reference DFT data. While the covalent radii were fit to the DFT-derived bond distances, these radii are also included in the determination of the coefficients for the bond angles in UFF (and by association UFF4MOF). It is therefore interesting that they observed good agreement with select angles in relevant MOF structures. While the

primary purpose of their work was to be able to optimize MOFs to their minimum energy structure, it is of interest to see if this force field can be used to reproduce dynamic and bulk properties of MOFs, particularly due to it's expansive coverage of the periodic table.

The implementation is idential to that of UFF metioned in Section S.3.1, with the exception that some of the metal-oxygen and metal-metal bond orders were adjusted to those values prescribed by Addicoat *et al.*. Here we will only comment that in IRMOF-1 and IRMOF-10, we adjusted the zinc - carboxylate oxygen bond order to 0.5, while in HKUST-1 the bond order of Cu-Cu was set to 0.25, and Cu-O was set to 0.5. The secondary building units (SBUs), their special bond orders, and atom types were assigned through a molecular recognition algorithm involving subgraph clique detection. The details of this algorithm are provided below in Section S.8.

S.3.5 IRMOF Force Field by Dubbeldam, Walton, Ellis and Snurr (DWES)

A force field for the IRMOF-series of materials was developed for the purpose of not only characterising material properties, but also accurately reproducing gas adsorption isotherms in the isoreticular series of MOFs discovered by the Yaghi group.[S20] Exceptionally, this force field contains no explicit bonds between Zn and coordinating oxygens, as their interactions are modeled with isotropic non-bonded interactions.

To recognize and assign particular force field parameters to the various atom types in the DWES force field, molecular recognition of the aromatic groups biphenyl and benzene were performed. For example, in this study there were 5 separate carbon atom types for the biphenyl linker of IRMOF-10, Ca, Cb, Cc, Cd, and Ce (see the article by Dubbeldam *et al.* for further details[S21]). Thus we have created template molecules for each linker with pre-assigned force field types, and upon recognition of the template in the MOF, the types are automatically assigned to the correct atoms. In the current study only IRMOF-1 and IRMOF-10 were considered, so recognition was limited to the Zn_4O cluster, biphenyl, and

benzene. However this type of force field assignment can easily be extended to other types of molecular fragments.

This force field can be described accurately using available potentials in LAMMPS, and it's implementation is described in the sections below.

S.3.5.1 Bond stretching

As with UFF and DREIDING, the DWES force field uses a harmonic bond potential described in Equation S2. Prior to setting the parameters for harmonic stretching of each bond, each of the Zn-O bonds found initially with the bond detection scheme reported in Section S.2, were deleted. The parameters for harmonic bond stretching K and R_0 were found in the supporting information of the DWES article.[S21]

Note that the force constants provided in Ref S21 are given in units of Kelvins and were thus converted to kcal/mol using the conversion factor of 0.00198588.

S.3.5.2 Angle bending

The angle potential is described with a harmonic term,

$$E_{bend} = K_{IJK} \left(\theta - \theta_0\right)^2 \tag{S27}$$

for a force constant K_{IJK} between atoms I, J, and K and equilibrium angle θ_0. This is available in LAMMPS as the angle style 'harmonic'.

Note that the force constants provided in Ref S21 are given in units of Kelvins/degree^2 and were thus converted to kcal/rad^2 using the conversion factor of $0.00198588/\left(\pi/180\right)^2$.

S.3.5.3 Dihedral torsion

For bonded atoms J, and K, a dihedral potential is defined for their respective neighbors I and L is formed with the function,

$$E_{torsion} = K_{IJKL}\left[1 + cos\left(m\phi - \phi_0\right)\right] \tag{S28}$$

which is found in LAMMPS under the dihedral style 'charmm'. Each force constant K_{IJKL} was multiplied by a factor of 0.00198588 to convert from the reported Kelvin units to kcal/mol.

S.3.5.4 Improper torsion

Inversion potentials were modeled using the 'cvff' improper style in LAMMPS, which resembles the DWES improper function;

$$E^{DWES}_{inversion} = K_{IJKL}\left[1 + cos\left(m\omega - \omega_0\right)\right] \tag{S29}$$

with the exception that there is no out of phase shift (i.e. no ω_0 term). The cvff potential

$$E^{cvff}_{inversion} = K_{IJKL}\left[1 + dcos\left(n\omega\right)\right] \tag{S30}$$

was able to reproduce Equation S29 by setting $d = -1$, since all ω_0 values in the DWES force field were set to 180°. The conversion from Kelvin to kcal/mol energy units was accomplished by multiplying K_{IJKL} by 0.00198588.

S.3.5.5 Non-bonded interactions

Non-bonded interactions, particularly important in the DWES force field for their description of Zn - O coordination bonds, were modelled using an Ewald summation for charges and the Lennard-Jones function in Equation S9 for van der Waals interactions. Each ϵ value was

converted to kcal/mol from Kelvins via multiplication of the constant 0.00198588.

All 1-4 neighbour interactions were turned on, as was the case for UFF, UFF4MOF, DREIDING and BTW-FF, and a tail correction was included to account for the non-bonded cutoff value of 12.5 Å.

S.4 Energy Minimization of MOFs

Prior to collecting statistical data from the ensemble, each framework was relaxed using a conjugate gradient algorithm. Because of the ill defined nature of relaxing both the cell shape and the atoms at the same time, a self-consistent cycle was used where at each step, the cell shape was relaxed, followed by atomic positions. This was performed until the energy difference between subsequent cycles was below 1×10^{-11} kcal/mol. For each MOF, the number of cycles and minimization steps per cycle varied. Cycles are represented as green lines in the minimization trajectories in Figure S2 for the BTW force field. The root-mean-squared deviation (RMSD) is averaged over all atoms in the framework, where the deviation is defined from their initial x-ray refined positions.

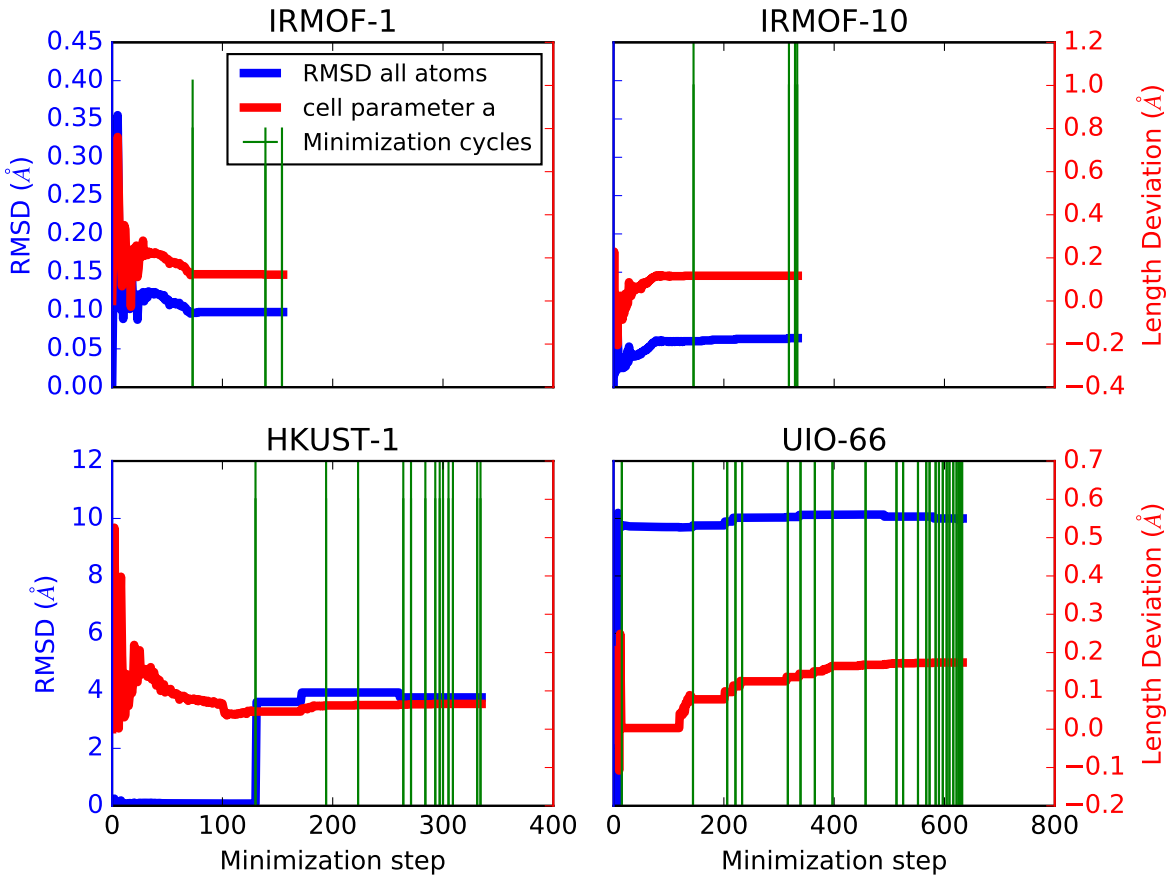

Figure S2: Trajectory of the root mean squared deviation (RMSD - blue) and the deviation of cell parameter 'a' (red) for the MOFs modeled with BTW-FF. The right y-axis corresponds to the cell parameter deviation, while the left corresponds to the atomic deviation. The green bars represent the end of a cycle, consisting of a relaxation of the atomic coordinates, followed by the cell shape.

The remaining cases are presented in the tables below. The relatively large RMSD values reported for UiO-66 are a result of aromatic carbon atoms adjusting from their 'puckered' symmetry-assigned placements to more 'flat' orientations. It is noteworthy that the two DFT-derived charge sets yield significant deviations in cell length when coupled to UFF and UFF4MOF. The reasons for this are currently unclear.

Table S4: Minimization Information For **BTW-FF**.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.098	+0.122	155	3
IRMOF-10	0.064	+0.117	334	4
HKUST-1	3.773	+0.066	335	13
UiO-66	10.00	+0.174	633	25

Table S5: Minimization Information For **UFF**.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.541	+0.086	1329	11
IRMOF-10	0.704	-0.015	1340	13
HKUST-1	4.404	+0.326	1575	18
UiO-66	9.869	+0.205	1110	11

Table S6: Minimization Information For **UFF4MOF**.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.285	+0.221	1628	14
IRMOF-10	0.773	+0.079	2719	18
HKUST-1	3.812	+0.667	775	6

Table S7: Minimization Information For **DREIDING**.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.263	+0.544	184	2
IRMOF-10	0.560	+0.370	1547	12
HKUST-1	2.187	+0.335	204	8

Table S8: Minimization Information For **DWES**.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.086	+0.174	54	4
IRMOF-10	0.098	+0.087	130	3

S.4.0.1 Minimization with Charge Equilibration (QEq) Charges

Table S9: Minimization Information For **UFF** with QEq charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.128	-0.129	137	3
IRMOF-10	0.061	+0.020	166	4
HKUST-1	3.782	+0.182	239	7
UiO-66	9.730	-0.087	864	11

Table S10: Minimization Information For **UFF4MOF** with QEq charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.109	+0.022	139	6
IRMOF-10	0.076	+0.098	220	2
HKUST-1	3.799	+0.556	238	7

Table S11: Minimization Information For **DREIDING** with QEq charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.199	+0.413	72	3
IRMOF-10	0.239	+0.503	143	3
HKUST-1	3.831	+0.186	175	7

S.4.0.2 Minimization with REPEAT Charges

Table S12: Minimization Information For **UFF** with REPEAT charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	2.309	-4.113	1940	8
IRMOF-10	2.367	-4.271	4239	16
HKUST-1	3.758	+0.075	109	4
UiO-66	9.046	-3.133	5423	25

Table S13: Minimization Information For **UFF4MOF** with REPEAT charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	2.118	-3.984	1757	27
IRMOF-10	2.216	-4.150	2693	52
HKUST-1	3.790	+0.469	282	10

Table S14: Minimization Information For **DREIDING** with REPEAT charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.080	+0.100	115	3
IRMOF-10	0.089	+0.134	169	6
HKUST-1	3.068	+0.073	208	4

S.4.0.3 Minimization with DDEC Charges

Table S15: Minimization Information For **UFF** with DDEC charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.324	-0.578	151	4
IRMOF-10	0.281	-0.510	84	2
HKUST-1	3.764	+0.154	203	7
UiO-66	9.046	-2.324	5980	17

Table S16: Minimization Information For **UFF4MOF** with DDEC charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	2.036	-3.721	3629	56
IRMOF-10	2.364	-4.145	7460	71
HKUST-1	3.628	+0.530	67	4

Table S17: Minimization Information For **DREIDING** with DDEC charge assignment.

MOFname	RMSD (Å)	ΔCellA (Å)	N steps	Number of cycles
IRMOF-1	0.111	+0.211	134	5
IRMOF-10	0.063	+0.066	121	4
HKUST-1	2.651	+0.145	297	10

S.5 Thermal Expansion Coefficients

The thermal expansion coefficients were determined from a isothermal-isobaric ensemble, where a Parrinello-Rahaman barostat[S22] was coupled to all six degrees of freedom of the unit cell. A pressure of 1 bar was applied uniformly to each degree of freedom in the unit cell, resulting in what was recently called the $NP\left(\sigma_a\right)T$ to distinguish between ensembles that control various elements of the unit cell geometry and stress tensor.[S23] The pressure

was relaxed every 1000 fs. The barostat and thermostat were coupled to three chains, as is the default in LAMMPS. The temperature was scaled in 20 equal increments from 80 to 500 Kelvin, where at each temperature the minimum energy structure was subjected to a 200 ps NVT simulation with particle velocities coupled to a Langevin thermostat, a 200 ps $NP\left(\sigma_a\right)T$ simulation with a Nosé-Hoover thermostat and Parrinello-Rahman barostat, and finally an 800 ps production run in the $NP\left(\sigma_a\right)T$ ensemble with the same thermostat and barostat. The thermostats were set to a 100 ps relaxation time.

Linear fits to the temperature - volume and temperature - lattice parameter curves provided the coefficients for linear (Eqn S31) and volumetric (Eqn S32) thermal expansion coefficients, using the slopes in the equations below.

$$\alpha = \left(\frac{1}{a_0}\right)\left(\frac{\partial a}{\partial T}\right) \tag{S31}$$

$$\beta = \left(\frac{1}{V_0}\right)\left(\frac{\partial V}{\partial T}\right) \tag{S32}$$

Where a_0 and V_0 are the averaged values for the lattice parameter a and cell volume at 298 K.

S.5.1 UFF Thermal Expansion Coefficients.

Table S18: Thermal Expansion Coefficients For **IRMOF-1** with UFF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-13.0	-14.8	+3.6	+10.6
Volumetric β	-39.9	-44.0	+1.3	+31.0

Table S19: Thermal Expansion Coefficients For **IRMOF-10** with UFF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-15.2	-27.2	-14.0	-8.9
Volumetric β	-46.3	-81.4	-38.5	-31.2

Table S20: Thermal Expansion Coefficients For **HKUST-1** with UFF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-11.8	-12.5	-12.9	-12.5
Volumetric β	-34.6	-34.8	-38.0	-36.6

Table S21: Thermal Expansion Coefficients For **UiO-66** with UFF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-8.4	-2.0	-3.4	-0.10
Volumetric β	-26.3	-4.8	-23.7	+9.7

S.5.2 DREIDING Thermal Expansion Coefficients.

Table S22: Thermal Expansion Coefficients For **IRMOF-1** with DREIDING. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-10.6	-11.1	-10.1	-0.09
Volumetric β	-32.2	-33.4	-27.8	-0.16

Table S23: Thermal Expansion Coefficients For **IRMOF-10** with DREIDING. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-16.0	-17.4	-47.1	-62.2
Volumetric β	-45.8	-52.1	-129.6	-187.7

Table S24: Thermal Expansion Coefficients For **HKUST-1** with DREIDING. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-4.0	-4.0	-3.9	-3.9
Volumetric β	-11.7	-10.9	-11.5	-11.2

S.5.3 UFF4MOF Thermal Expansion Coefficients.

Table S25: Thermal Expansion Coefficients For **IRMOF-1** with UFF4MOF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-26.2	-17.9	-6.0	-0.05
Volumetric β	-77.9	-53.6	-13.6	-12.1

Table S26: Thermal Expansion Coefficients For **IRMOF-10** with UFF4MOF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-16.9	-34.2	-14.8	-9.4
Volumetric β	-53.0	-100.8	-44.5	-30.8

Table S27: Thermal Expansion Coefficients For **HKUST-1** with UFF4MOF. Units in 10^{-6} K^{-1}.

	No Charge	QEq	REPEAT	DDEC
Linear α	-11.1	-8.7	-12.2	-11.5
Volumetric β	-29.8	-30.5	-34.3	-32.9

S.6 Bulk Modulus Calculations

Bulk modulus calculations were performed starting from the minimum energy structure as determined by the minimization scheme presented in Section S.4. For each different charge scheme, the minima of each MOF was re-calculated. The structures were isotropically expanded and contracted to ± 1% of their unit cell length, in increments of 0.05%. At each expansion or contraction, the atomic positions were relaxed to their strained minimum energy using a damped dynamics algorithm (fire[S24]) in LAMMPS, using an energy convergence criteria of 10^{-15} kcal/mol, a force cutoff of 10^{-15} kcal/mol/Å or a maximum of 10000 iterations.

The resulting energy vs volume plot was fit to the Murnaghan equation of state using a least squares fitting procedure. The Figures S4, S5, and S6 show the curve of best fit to Equation S33, along with the data points determined from each force field (points).

It is found that, in general, providing charges to these systems results in very unpredictable behaviour with respect to their bulk modulus, where poor fits to Eqn. S33 are observed in at least one case for each of the three charge generation methods used in this work (QEq, REPEAT, and DDEC).

$$E = E_0 + \frac{B_0 V}{B'} \left[\frac{\left(\frac{V_0}{V}\right)^{B'}}{(B'-1)} + 1 \right] \tag{S33}$$

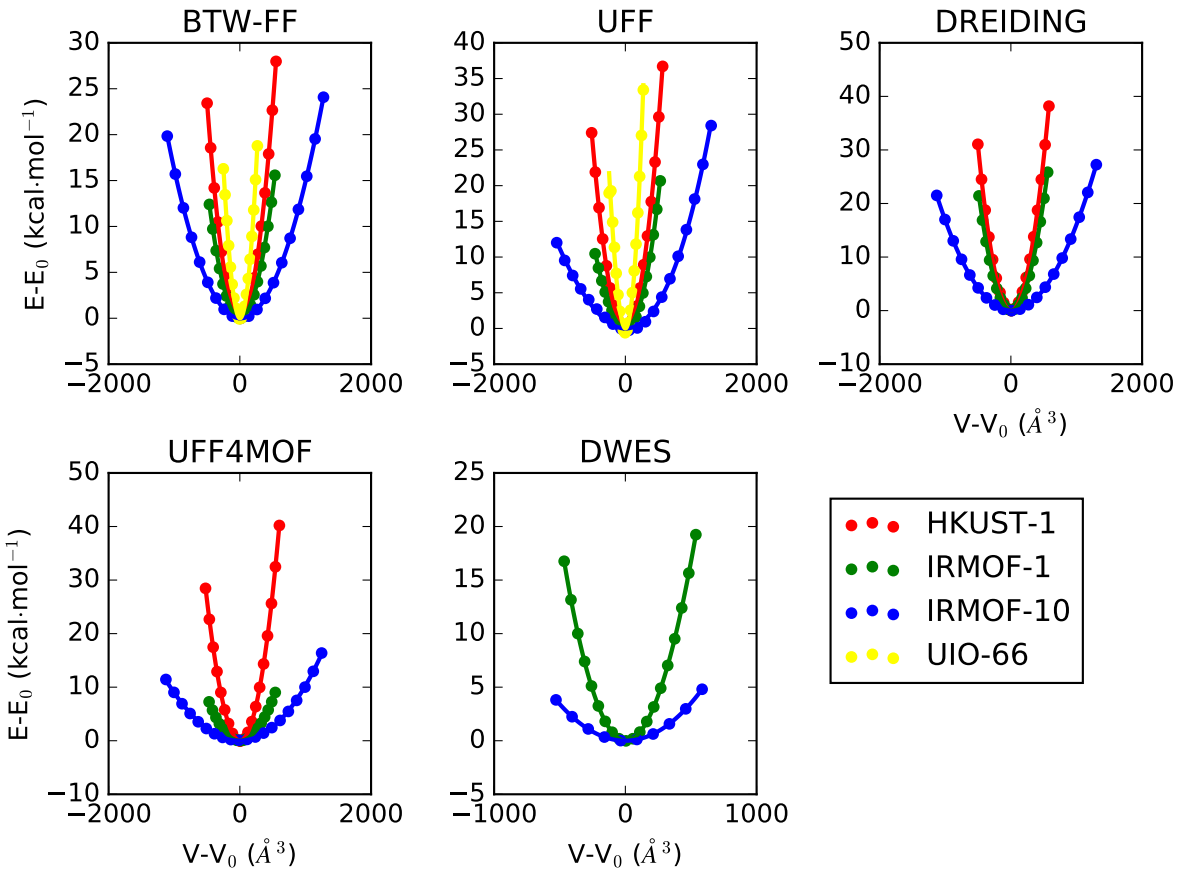

Figure S3: Energy vs. Volume plots for each MOF modelled with the force fields studied in this work. For the force fields UFF, DREIDING, and UFF4MOF, no charges were assigned to each atom. Whereas atoms parameterized with the BTW-FF and DWES force fields were assigned the partial charge ascribed to them from the original articles.

Table S28: Bulk Modulus Calculations with UFF. Units in GPa.

	No Charge	QEq	REPEAT	DDEC
IRMOF-1	14.47	28.56	27.71	10.68[c]
IRMOF-10	7.57	10.19[a]	11.56	0.18[c]
HKUST-1	28.70	28.53	28.44	28.40
UiO-66	42.44	38.91	37.07	14.70[c]

a - poor fit to Equation S33. See Figure S5.
c - poor fit to Equation S33. See Figure S6.

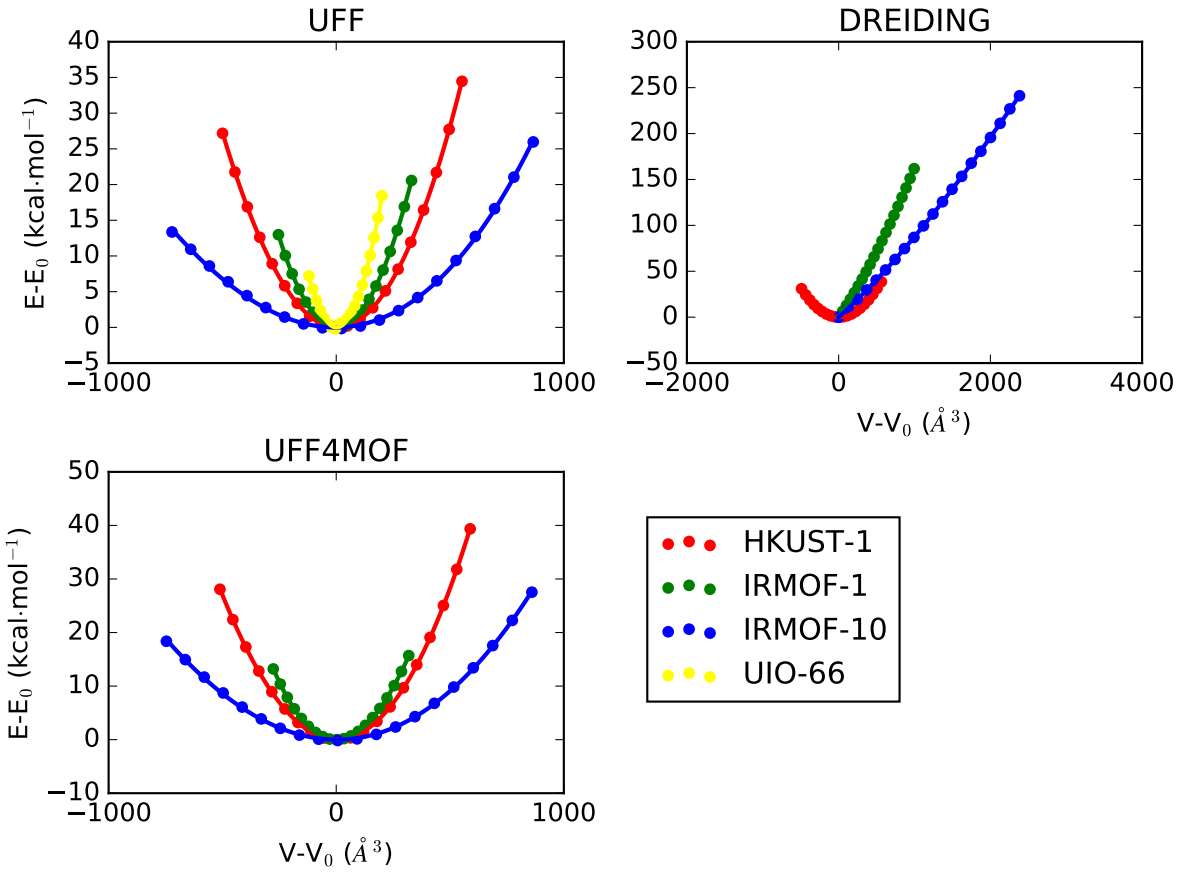

Figure S4: Energy vs. Volume plots for each MOF modelled with the UFF, DREIDING and UFF4MOF force fields. Each MOF was assigned partial charges from a REPEAT calculation.

Table S29: Bulk Modulus Calculations with DREIDING. Units in GPa.

	No Charge	QEq	REPEAT	DDEC
IRMOF-1	22.00	21.96	16.76	0.00[c]
IRMOF-10	9.75	35.36[a]	16.43[b]	0.42[c]
HKUST-1	31.13	31.65	32.19	31.85

a - poor fit to Equation S33. See Figure S5.
b - poor fit to Equation S33. See Figure S4.
c - poor fit to Equation S33. See Figure S6.

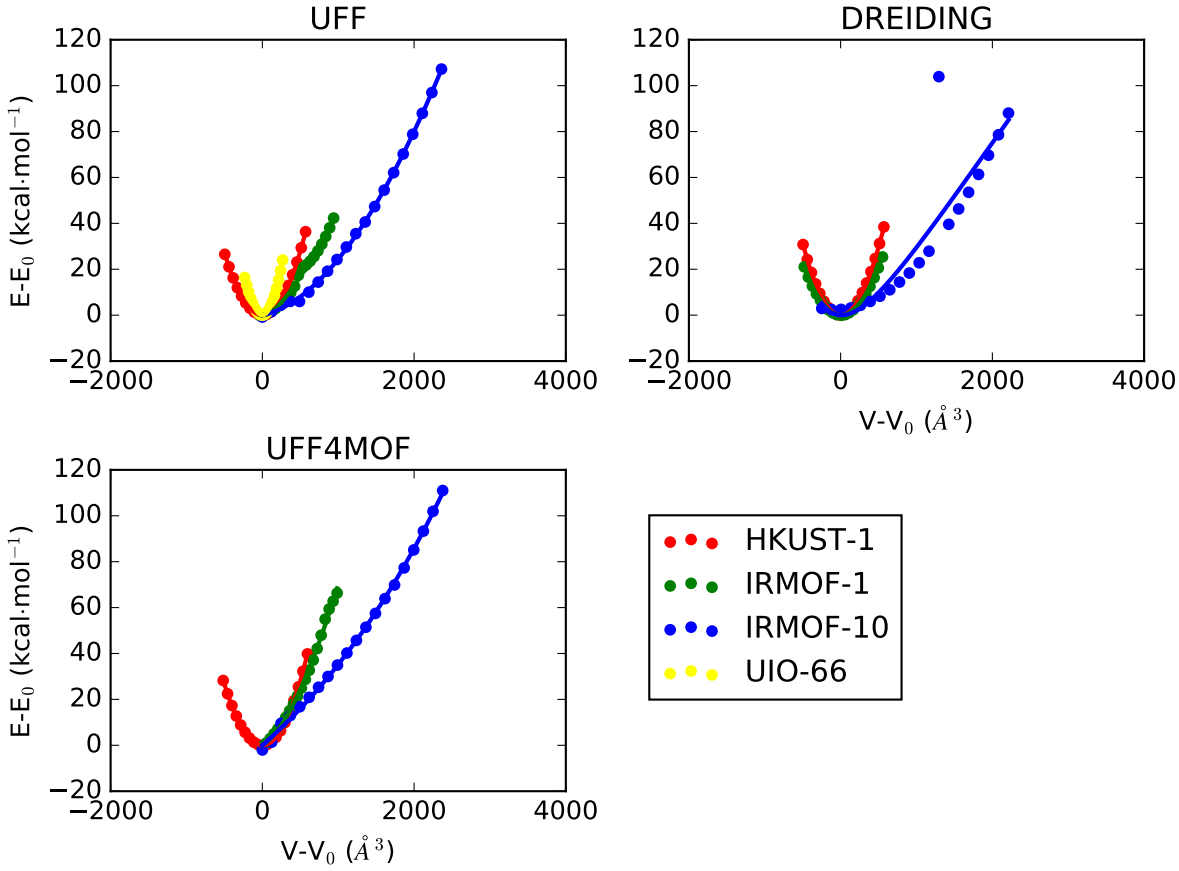

Figure S5: Energy vs. Volume plots for each MOF modelled with the UFF, DREIDING and UFF4MOF force fields. Each MOF was assigned partial charges from the charge equilibration method (QEq).

Table S30: Bulk Modulus Calculations with UFF4MOF. Units in GPa.

	No Charge	QEq	REPEAT	DDEC
IRMOF-1	16.84	17.90[a]	23.74	22.20
IRMOF-10	10.40	0.05[a]	13.81	9.78
HKUST-1	29.43	29.57	29.60	29.49

a - poor fit to Equation S33. See Figure S5.

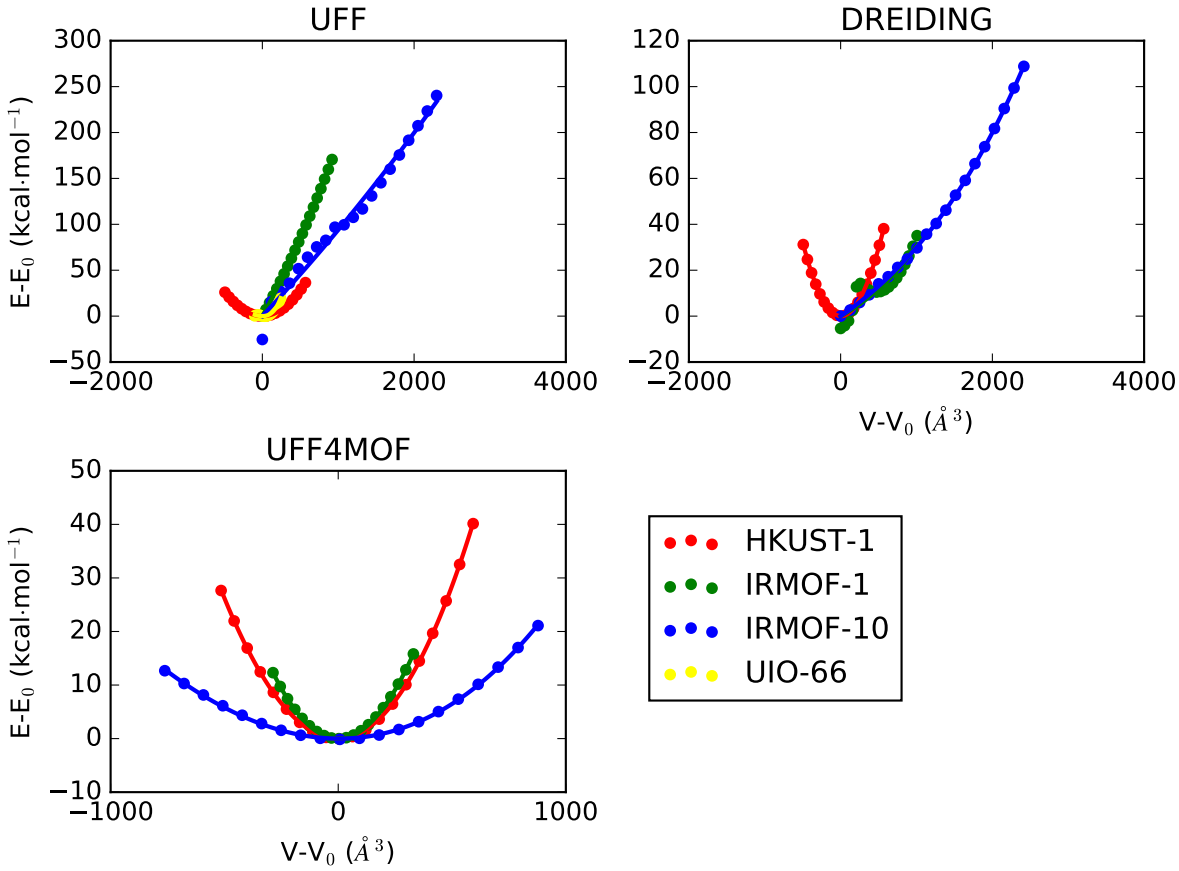

Figure S6: Energy vs. Volume plots for each MOF modelled with the UFF, DREIDING and UFF4MOF force fields. Each MOF was assigned partial charges from a DDEC calculation.

S.7 Elastic Constants of IRMOF-1

Each element C_{ij} of the coefficient matrix was computed by determining the curvature from a selection of deformations of the minimum energy structure. Energy minimisations were

carried out in a similar fashion to that described in Section S.6, with the exception that each curve is determined as a function of anisotropic deviations of the simulation cell. The bulk moduli in Table S31 are computed from the average of two separate methods, the Voigt bulk modulus computed as

$$K_V = \frac{(C_{11} + C_{22} + C_{33}) + 2\,(C_{12} + C_{23} + C_{31})}{9} \tag{S34}$$

and the Reuss bulk modulus;

$$K_R = \left[(s_{11} + s_{22} + s_{33}) + 2\,(s_{12} + s_{23} + s_{31})\right]^{-1} \tag{S35}$$

where $s_{ij} = C_{ij}^{-1}$.

The shear modulus was likewise computed as an average of the Voigt and Reuss shear moduli, where the Voigt shear modulus is calculated as

$$G_V = \frac{(C_{11} + C_{22} + C_{33}) - (C_{12} + C_{23} + C_{31}) + 3\,(C_{44} + C_{55} + C_{66})}{15} \tag{S36}$$

and the Reuss shear modulus is calculated using

$$G_R = \left[\frac{4\,(s_{11} + s_{22} + s_{33}) - 4\,(s_{12} + s_{23} + s_{31}) + 3\,(s_{44} + s_{55} + s_{66})}{15}\right]^{-1} \tag{S37}$$

The stiffness matrices for IRMOF-1 are reported in Equations S38 - S42 for each force field studied in the main text. All values are reported in GPa. It is clear that while the bulk moduli computed using these matrices are similar, there are notable differences in the stiffness coefficients between each force field. For example, both the stiffness matrix for the DWES force field (matrix S42) and UFF4MOF (matrix S40) do not satisfy the Born stability condition as they contain negative entries. This suggests that future force field developments can be better refined by capturing these tensor properties when comparing

with DFT calculations, for example.[S25]

Table S31: Materials properties for IRMOF-1 computed from the stiffness matrix. Units are GPa.

	BTW	UFF	UFF4MOF	DREIDING	DWES
bulk modulus	13.8	13.7	10.5	22.0	19.5
shear modulus	4.0	2.1	2.0	3.7	12.6

$$C_{BTW} = \begin{pmatrix} 37.6 & 1.9 & 1.9 & 0.0 & 0.0 & 0.0 \\ 1.9 & 37.5 & 1.9 & 0.0 & 0.0 & 0.0 \\ 1.9 & 1.9 & 37.4 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.4 \end{pmatrix} \tag{S38}$$

$$C_{UFF} = \begin{pmatrix} 25.8 & 7.6 & 7.6 & 0.0 & 0.0 & 0.0 \\ 7.6 & 25.9 & 7.4 & 0.0 & 0.0 & 0.0 \\ 7.6 & 7.4 & 26.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 \end{pmatrix} \tag{S39}$$

$$C_{UFF4MOF} = \begin{pmatrix} 21.3 & 5.0 & 4.9 & 0.0 & 0.1 & 0.0 \\ 5.0 & 23.0 & 3.9 & 0.0 & -0.2 & 0.0 \\ 4.9 & 3.9 & 23.2 & -0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.4 & 0.2 & 0.0 & 0.0 \\ 0.1 & -0.2 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \end{pmatrix} \tag{S40}$$

$$C_{DREIDING} = \begin{pmatrix} 43.1 & 11.4 & 11.4 & 0.0 & 0.0 & 0.0 \\ 11.4 & 43.1 & 11.4 & 0.0 & 0.0 & 0.0 \\ 11.4 & 11.4 & 43.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \end{pmatrix} \tag{S41}$$

$$C_{DWES} = \begin{pmatrix} 33.8 & 13.4 & 9.7 & -2.8 & 4.1 & 0.2 \\ 13.4 & 36.0 & 14.4 & -0.3 & 0.1 & 0.0 \\ 9.7 & 14.4 & 44.9 & -2.4 & -1.1 & -2.3 \\ -2.8 & -0.3 & -2.4 & 0.7 & -1.0 & -0.1 \\ 4.1 & 0.1 & -1.1 & -1.0 & 3.3 & -1.3 \\ 0.2 & 0.0 & -2.3 & -0.1 & -1.3 & 0.7 \end{pmatrix} \tag{S42}$$

S.8 Substructure Search

Common inorganic and organic SBUs were discovered using a 3-dimensional pattern recognition algorithm, to assign special flags to the bonds and atoms of the clusters for easy identification when applying a MOF related force field. The pattern recognition depends on a maximum clique detection algorithm described originally by Bron and Kerbosch.[S26] In graph theory, the maximum clique corresponds to the largest subset of nodes in a graph that are connected to each other by edges. However we will dispense with the mathematical terminology here and describe how cliques were used to find chemical structures in lay-person terms.

Maximum cliques are used in chemical pattern recognition by first constructing what is called a *correspondence graph*. Each node of a correspondence graph consists of two atoms, one from a representative template SBU and one from the MOF. Nodes are constructed only if the atoms are of the same element. Two nodes are connected by an edge if the

atoms of each node are spaced the same distance away from each other in their original structures (regardless of bonding information). Thus a clique in the correspondence graph will contain a set of atoms that are similarly spaced in two separate systems. A tolerance can be specified when determining edges in the correspondence graph, such that one can account for distortions in atomic positions in different materials.

The successful discovery of an inorganic SBU in a MOF structure is dependent on 1) the reference SBU's atomic positions used to construct the correpsondence graph, and 2) a user-defined tolerance for distance comparisons between reference SBUs and the MOF structure In this work the tolerance was set to 0.4 Å, which was sufficient to discover all of the inorganic SBUs in IRMOF-1, IRMOF-10,[S20] HKUST-1,[S27] and UiO-66[S28] the templates of which were extracted from the crystallographic coordinates of the SBUs within these MOFs.

S.8.1 Mining the CoRE database for metal SBUs

Using several SBUs that are well-studied and have been parameterized in at least one force field, the clique detection algorithm was used to determine how prevalent they were in the CoRE MOF database. To account for potential structural variability of these SBUs the tolerance was set to 0.7 Å when constructing the correspondence graphs. Setting a value much higher than this would result in some cases in 'false positive' recognition of SBUs that did not resemble the original template structure. Table S32 represents a detailed breakdown of the SBUs discovered in the CoRE database, using the SBU terminology presented in Figure 2 of the main text.

While fascinating how so many different elements can be found with the SBU configurations commonly known in MOFs, they ultimately represent a tiny fraction of the total known materials. It is therefore a worthy pursuit to assess if more generalized force fields are necessary, such that one can robustly model more elements and coordination environments in these remarkable materials.

Table S32: Metal SBU counts in the CoRE MOF database

Metal atom	M Paddlewheel	M_4O	M_3O	M pillar	M_6O_8
Ni	3	0	9	3	0
Mg	0	0	1	14	0
Co	22	0	5	27	0
Mn	8	0	1	7	0
Zn	125	54	5	13	0
Al	0	0	0	2	0
Fe	3	0	4	6	0
Ga	0	0	1	1	0
In	0	0	0	4	0
V	0	0	0	1	0
Cu	197	0	2	1	0
Zr	0	0	0	0	2
Sc	0	0	1	0	0
Cr	1	0	10	0	0
Ru	2	0	0	0	0
Cd	3	0	0	0	0
CoRE Coverage (%)	7.7	1.2	0.8	1.6	0.04

References

(S1) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, A Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(S2) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to MetalOrganic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.

(S3) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

(S4) Rappé, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(S5) Gale, J. D.; Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **2003**, *29*, 291–341.

(S6) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49*, 14251–14269.

(S7) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(S8) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(S9) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(S10) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(S11) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(S12) Campañá, C.; Mussard, B.; Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **2009**, *5*, 2866–2878.

(S13) Manz, T. A.; Sholl, D. S. 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.

(S14) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. Covalent Radii Revisited. *Dalton Trans.* **2008**, 2832–2838.

(S15) Garberoglio, G. OBGMX: A Web-Based Generator of GROMACS Topologies for

Molecular and Periodic Systems Using the Universal Force Field. *J. Comput. Chem.* **2012**, *33*, 2204–2208.

(S16) Bristow, J. K.; Tiana, D.; Walsh, A. Transferable Force Field for Metal-Organic Frameworks from First-Principles: BTW-FF. *J. Chem. Theory Comput.* **2014**, *10*, 4644–4652.

(S17) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(S18) Ponder, J. W. TINKER: Software tools for molecular design. https://dasher.wustl.edu/tinker/, Version 7.1.

(S19) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. MOF-FF - A Flexible First-Principles Derived Force Field for Metal-Organic Frameworks. *Phys. Status Solidi B* **2013**, *250*, 1128–1141.

(S20) Eddaoudi, M.; Kim, J.; Rosi, N. L.; Vodak, D. T.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage. *Science* **2002**, *295*, 469–472.

(S21) Dubbeldam, D.; Walton, K. S.; Ellis, D. E.; Snurr, R. Q. Exceptional Negative Thermal Expansion in Isoreticular Metal - Organic Frameworks. *Angew. Chem., Int. Ed. Engl.* **2007**, *119*, 4580–4583.

(S22) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(S23) Rogge, S.; Vanduyfhuys, L.; Ghysels, A.; Waroquier, M.; Verstraelen, T.; Maurin, G.; Van Speybroeck, V. A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks. *J. Chem. Theory Comput.* **2015**, *11*, 5583–5597.

(S24) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.

(S25) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F. X. Anisotropic Elastic Properties of Flexible Metal-Organic Frameworks: How Soft are Soft Porous Crystals? *Phys. Rev. Lett.* **2012**, *109*, 195502.

(S26) Bron, C.; Kerbosch, J. Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575–577.

(S27) S-Y Chui, S.; M-F Lo, S.; H Charmant, J. P.; Guy Orpen, A.; Williams, I. D. A Chemically Functionalizable Nanoporous Material [Cu 3 (TMA) 2 (H 2 O) 3] n. *Science* **1999**, *283*, 1148–1150.

(S28) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, *130*, 13850–13851.