QuickFF: toward a generally applicable methodology to quickly derive accurate force fields for Metal-Organic Frameworks from ab initio input

L. Vanduyfhuys,[†] S. Vandenbrande,[†] T. Verstraelen,[†] R. Schmid,[‡] M. Waroquier,[†] and V. Van Speybroeck[*,†]

Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium, , and Lehrstul für Anorganische Chemie 2, Organometallics and Materials Chemistry, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44780 Bochum, Germany,

E-mail: Veronique.VanSpeybroeck@Ugent.be

Supporting Information

[*]To whom correspondence should be addressed
[†]Ghent University
[‡]Ruhr-Universität Bochum

Contents

1 Atom types

Atom types can be chosen using a built-in function varying from low to high level:

- Low: atom types are assigned only according to the element
- Medium: atom types are assigned according to the element and the number of neighbors
- High: atom types are assigned according to the element, the number of neighbors and the neighboring elements
- Highest: atom types are assigned according to the index of each atom, i.e. every atom has a unique atom type

When choosing atom types according to the level 'medium', atom type strings will be of the form 'EN' in which E is the element and N is the number of neighbors. These atom types are illustrated in figure S1 for 3 molecules. When choosing atom types according to the level 'high', atom type strings

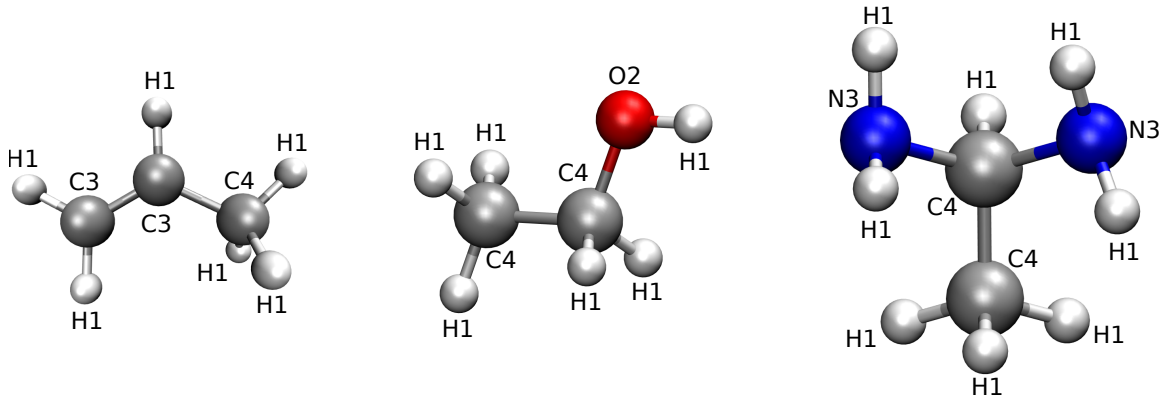

Figure S 1: Illustration of atom types assigned according to the level medium.

will be of the form 'ENs' in which E is the element, N is the number of neighbors and s is a string describing the neighbors. If the atom has only 1 neighbor, then s is equal '_e' with e the element of the neighbor. If the atom under consideration has 2 neighbors, then s is equal to '_ee' in which the first and second e represent the element of the first and second neighbor respectively. If the atom has more than 2 neighbors, then s will contain a substring '_en' for every neighbor element that is either carbon, nitrogen or oxygen. In this substring, e represents the neighbor element (and is either c, n or o) and n is the number of neighbors of that particular neighbor element. Multiple instances of this '_en' string are ordered according to atomic number. These atom types are illustrated in figure S2 for the same 3 molecules as before.

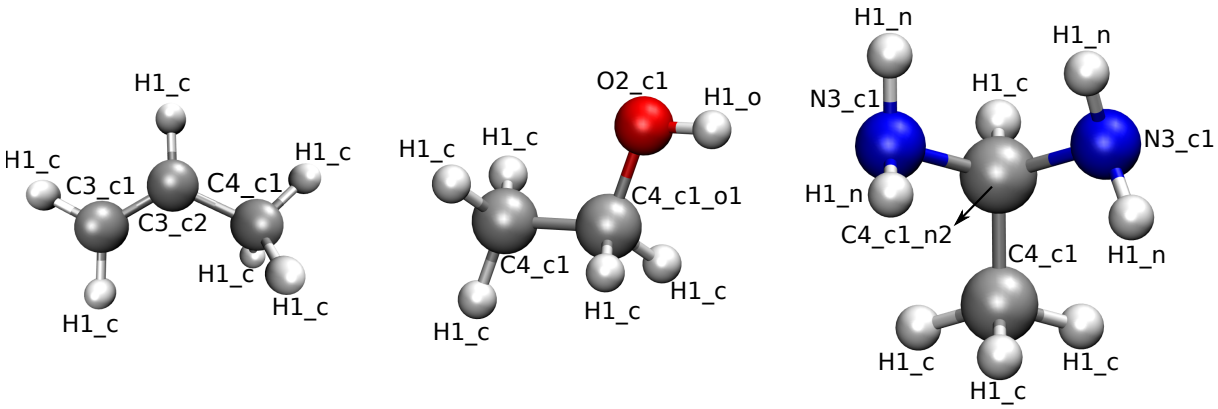

Figure S 2: Illustration of atom types assigned according to the level high.

2 Dihedral potential

We will start by explaining the general procedure to determine the dihedral potential and two examples will be given at the end. The potential energy of the dihedral terms is described using a single cosine expression:

$$V_{\text{torsion},k} = \frac{K_k^{\phi}}{2}\left[1 - \cos\left(m_k(\phi_k - \phi_{0,k})\right)\right] \tag{2.1}$$

For every dihedral pattern of atom types, such a term is included in the potential energy expression. Such a dihedral pattern can correspond to several dihedral angles (e.g. the dihedral pattern H-C-C-H in ethane corresponds to 9 dihedral angles). The potential is chosen to have local minima close to each of the ab initio values of the corresponding dihedral angles, if this turns out to be impossible, no dihedral term is included for that particular pattern. The general procedure to determine the multiplicity m_k and rest angle $\phi_{0,k}$ will be outlined in the next paragraph. This procedure will start by proposing a possible candidate of m_k and $\phi_{0,k}$ for every ab initio value of the corresponding dihedral angle. At the end, a unique potential can be defined if all candidates are equal. In some cases, the procedure does not allow for such a unique definition of m_k or $\phi_{k,0}$. In that case, we ignore the dihedral pattern and do not include a potential term for it. This is a simple and consistent way to assure that no dihedral terms are included with highly deficient parameters. However, it implies that some dihedral patterns will be ignored, which is mostly due to the fact that a single cosine function simply does not suffice to describe their contribution to the potential energy. We plan to address these with more complex energy terms, such as higher order Fourier series, in the future. On the other hand, dihedral patterns that can typically be accurately described by means of a single cosine, such as H-C-C-H in ethane, will be assigned a valid combination of m_k and $\phi_{k,0}$.

The rules for deriving the multiplicity and rest angle from a certain ab initio equilibrium angle ψ_i are as follows. First the multiplicity is determined based an the number of neighbors of the central 2 atoms of the dihedral (see Table 1). Once the multiplicity is known, the fundamental period of the cosine, i.e. $P = \frac{2\pi}{m}$, is divided into 3 intervals: $[0, \frac{P}{6}]$, $[\frac{2P}{6}, \frac{4P}{6}]$ and $[\frac{5P}{6}, P]$. Next, we consider the image of the absolute value of $|\psi_i|$ that lies in this fundamental period. If this image lies in the first or the third interval, the rest angle is chosen to be 0. If the image lies in the middle interval, the rest

angle is chosen to be $\frac{P}{2}$. If the image doesn't match any of the intervals, no dihedral potential will be assigned to the corresponding dihedral pattern.

Table S 1: Multiplicity as a function of the number of neighbors of the central atoms

Neighbours of the central atoms	Multiplicity
4, 4	3
4, 3	6
4, 2	3
3, 3	2
3, 2	2
2, 2	1

The procedure is illustrated for H-C-C-H in ethane and C-C-N-H in aminobenzene. Figure S3 shows the Newman projections of an ethane molecule and the the C-C-N-H dihedrals in aminobenzene. In the case of H-C-C-H in ethane, the central atoms, i.e. the two C atoms, both have 4 neighbors. Using Table 1, we can conclude that the multiplicity should be 3, which results in a fundamental period $P = 120°$. For every dihedral angle of type H-C-C-H in ethane, Table 2 contains the ab initio equilibrium value and its image in the fundamental period. From this table we can conclude that every instance has an image in the central interval, i.e. around 60°, and hence the rest angle should be 60°. The dihedral potential for H-C-C-H in ethane becomes:

$$V_{\mathrm{HCCH}} = \frac{K_k^{\psi}}{2} \left[1 - \cos\left(3(\phi_k - 60°)\right)\right] \tag{2.2}$$

The same procedure can be repeated for C-C-N-H in aminobenzene. The central C and N atoms both have 3 neihbors, which results in a multiplicity of 2. Hence the fundamental period is $P = 180°$. Table 2 again contains the ab initio equilibrium values and their image in the fundamental period of every instance of C-C-N-H in aminobenzene. The first instance has an image matching the interval around 0°, the third matches the interval around 180°, while the second and last match no interval. Hence, no conclusive decision could be made regarding the matching interval and hence no dihedral potential will be assigned to the dihedral pattern C-C-N-H in aminobenzene. Indeed, this type of dihedrals requires a more thorough investigation if one hopes to describe is accurately.

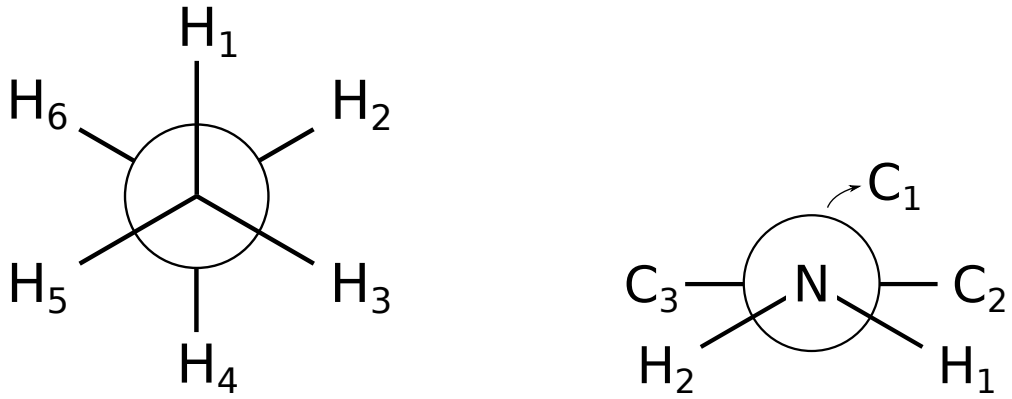

Figure S3: (left) Newman projection of an ethane molecule (right) Newman projection of the C-C-N-H dihedrals in aminobenzene.

Table S2: Images in the fundamental period of the ab initio dihedral angles in case of ethane(H-C-C-H) and aminobenzene(C-C-N-H).

Dihedral	ϕ_k^{ai}	Image
ethane($m = 3$) Intervals: $[0, 20], [40, 80], [100, 120]$		
H_1-C-C-H_2	65°	65°
H_1-C-C-H_4	178°	58°
H_1-C-C-H_6	-62°	62°
H_3-C-C-H_2	-52°	52°
H_3-C-C-H_4	63°	63°
H_3-C-C-H_6	-178°	58°
H_5-C-C-H_2	175°	55°
H_5-C-C-H_4	-65°	65°
H_5-C-C-H_6	57°	57°
aminobenzene($m = 2$) Intervals: $[0, 30], [60, 120], [150, 180]$		
C_2-C_1-N-H_1	-14°	14°
C_2-C_1-N-H_2	-148°	148°
C_3-C_1-N-H_1	166°	166°
C_3-C_1-N-H_2	32°	32°

3 Generation of perturbation trajectories

3.1 Perturbation trajectories

An innovative aspect of QuickFF is that trajectories are constructed along the multidimensional potential energy surface (PES) near the equilibrium structure. Changes to the potential energy along the perturbation trajectory can be modeled with the suggested FF expressions and fitted on the first principles predictions, achieved by a second order Taylor expansion of the ab initio PES. The fitting of the FF parameters completely relies on the Hessian matrix in Cartesian coordinates.

The transformation between the vector $\vec{q}$ containing the M selected internal coordinates (IC's) ($q_m, m = 1 \ldots M$), and the vector $\vec{R}$ containing the $3N_{\text{at}}$ Cartesian coordinates of the atomic positions in the molecular system, is non-linear. It is linearized by considering infinitesimal increments of the coordinates: $\delta\vec{q} = \mathbf{B}\,\delta\vec{R}$. $\mathbf{B}$ represents a $M \times 3N_{\text{at}}$ matrix, which has no uniquely defined inverse $\mathbf{B}^{-1}$ when the matrix $\mathbf{B}$ is overdetermined ($M > 3N_{\text{at}}$). Knowledge of the Hessian matrix in IC's leads unambiguously to the Cartesian Hessian matrix elements by the relation:

$$\left[\mathbf{H}_R\right]_{ij} = \left[\mathbf{B}^T\mathbf{H}_q\mathbf{B}\right]_{ij} + \sum_m \left(\frac{\partial E}{\partial q_m}\right) \frac{\partial^2 q_m}{\partial R_i \partial R_j} \tag{3.1}$$

Note that the second term is different from zero, because $0 = \frac{\partial E}{\partial \vec{R}} = \mathbf{B}^T \frac{\partial E}{\partial \vec{q}}$ does not imply $\frac{\partial E}{\partial \vec{q}} = 0$ due to the singular nature of $\mathbf{B}$. The reverse transformation is non trivial also due to the singular nature of $\mathbf{B}$. This implies that the conversion of Hessian matrix elements obtained from first principles to Hessian matrix elements expressed in the internal coordinates (coinciding with the force constants) is not uniquely determined and prone to large uncertainties. This is the main issue that various approximations for the parametrization of a FF in literature address. In a large number of protocols, one of the possible pseudo-inverses of $\mathbf{B}$ is chosen to transform the Hessian from Cartesian space to IC space and a cost function is constructed based on the Hessian in IC space. In this work we propose a new method. As a preliminary step before fitting the FF Hessian matrix to the ab-initio Hessian in Cartesian space, we explore the surroundings of the equilibrium structure on the PES by slightly changing the geometry of the system. The change of the potential energy can be evaluated using either the force field energy expression or the ab-initio expression. Equalization of both expressions results into constraints on the FF parameters.

We assume an arbitrary trajectory with small displacements of the atomic positions from the equilibrium structure: $\vec{R} = \vec{R}_0 + \delta\vec{R}$, where $\vec{R}_0$ represents the ab initio geometry of the equilibrium. This perturbation induces changes δq_m $(m = 1 \ldots M)$ of the internal coordinates, which are well defined. The ab initio total energy V^{ai} along the perturbation trajectory can be approximated by means of a Taylor expansion up to second order:

$$V^{\mathrm{ai}}\left(\vec{R}\right) = V_0^{\mathrm{ai}} + \left[\nabla V^{\mathrm{ai}}\right]^T \cdot \left(\vec{R} - \vec{R}_0\right) + \frac{1}{2}\left(\vec{R} - \vec{R}_0\right)^T \cdot \mathbf{H}^{\mathrm{ai}} \cdot \left(\vec{R} - \vec{R}_0\right) \tag{3.2}$$

Note that the ab initio forces should be zero in equilibrium, but there can be a deviation due to an imcomplete geometry optimization in the ab initio reference data.

Similarly, the force field energy along this arbitrary trajectory can also be computed:

$$V^{\mathrm{ff}}(\vec{R}) = V_{\mathrm{nb}}^{\mathrm{ff}}(\vec{R}) + V_{\mathrm{cov}}^{\mathrm{ff}}(\vec{R}) \tag{3.3}$$

with the non bonding contribution

$$V_{\mathrm{nb}}^{\mathrm{ff}}(\vec{R}) = V_{\mathrm{el}}^{\mathrm{ff}}(\vec{R}) + V_{\mathrm{vdW}}^{\mathrm{ff}}(\vec{R}) \tag{3.4}$$

Equalization of the expressions results into a relation on the FF parameters. This procedure can be repeated for several frames, systematically inducing perturbations of the atomic positions and hence generating a trajectory of perturbed configurations. For each frame characterized by $\vec{R}^{(p)}$ and $\vec{q}^{(p)}$ we generate an equation that must be fulfilled by the FF parameters. In principle, one can simulate any number of random perturbation trajectories. However, doing so without any attention for the number of equations and their linear independency may result in an ill-conditioned system of non-linear equations. To remedy this inconvenience, we carefully constuct very specific perturbation trajectories. Instead of allowing each internal coordinate to contribute to the change of the potential energy along a trajectory, we search for a trajectory where the main contribution in the covalent energy arises from a single internal coordinate. In case of a molecule with solely uncoupled internal coordinates each IC can vary without disturbing the others. Along this path only one term, corresponding to the perturbed IC, contributes to the change of the covalent potential energy. On the contrary, a coupled IC directly affects others, when it is submitted to a small perturbation. A ring structure is a nice example of a

system with coupled internal coordinates. Consider a benzene ring and induce a change δq_n of the bond length $q_n = AB$ (see Figure 4). It is obvious that this perturbation disturbs the other IC's of the ring: $q_m \rightarrow q_m + \delta q_m$ with $m \neq n$. The IC's are coupled by geometrical constraints, but many solutions (trajectories) exist. One can opt for a constrained structure optimization keeping the bond length AB fixed and relaxing the other bond lengths, bending angles and out-of-plane distances. This trajectory corresponds with a minimum of total energy (trajectory 1 in Figure 5). However, along this trajectory no attempt is made to restrict the changes of the IC's as response of a perturbation. Another option

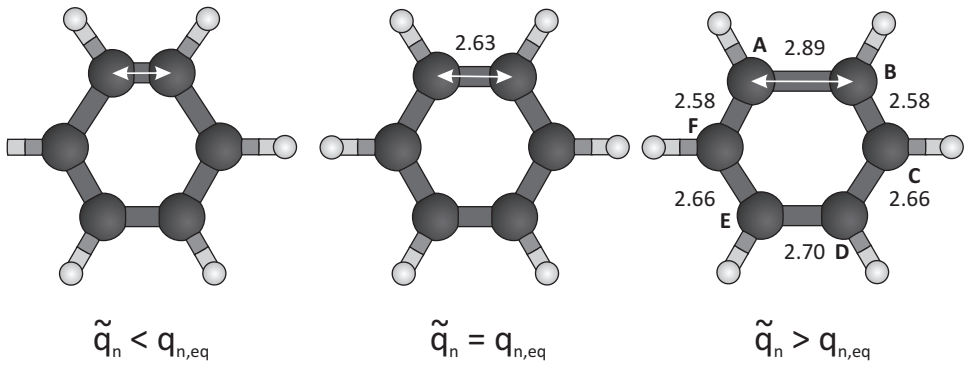

Figure S 4: Three frames from the perturbation trajectory $\vec{R}(\widetilde{q_n})$ in case q_n represents a C-C bond in benzene. (left) frame with a shrunken C-C bond (middle) frame with benzene in equilibrium (right) frame with an elongated C-C bond. The perturbations are highly exaggerated for the sake of clarity. The numbers indicate bond length in atomic units.

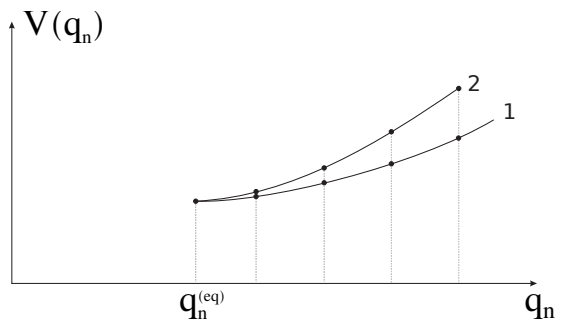

Figure S 5: Schematical behavior of the potential energy along perturbation trajectories

is the construction of a trajectory where the changes of the relaxed IC's are restricted to a minimum. Along this trajectory (labeled 2 in Figure 5) one may assume that the IC's are globally varying as decoupled as possible. To illustrate with the benzene ring of Figure 4, we assume a stretching of the C-C bond AB (denoted as q_i) of 10 % relative to the equilibrium value of $q_{i,eq}$ =1.39 Å (or 2.63 au). The concept of a minimal strain

$$\min \chi^S\left(\vec{R}^N\right) \quad \text{with} \quad \chi^S\left(\vec{R}^N\right) = \frac{1}{2}\sum_{j\neq i}\left(q_j - q_{j,eq}\right)^2 \tag{3.5}$$

leads to a solution $\vec{R}^N$, i.e. the molecular geometry, as displayed in Figure 4. The cost function amounts to 0.006 au^2. We neglected the deviations of the bending angles from the equilibrium value of 120 ° as they are too small to contribute to the cost function. To get an idea about the magnitude of the strain in other solutions, we consider the trajectory keeping all IC's in their original (equilibrium) values apart from the opposite C-C bond DE, which is submitted to the same elongation of AB. This solution leads to a single contribution in the cost function, but its magnitude of 0.034 au^2 exceeds the strain of previous case with a factor of 6.

A general formulation of the concept of perturbation trajectories belonging to a minimum of strain is now presented. Consider a particular IC q_n and assume some small perturbation bringing it to $\widetilde{q}_n$. The new structure, obtained by minimizing the strain releasing all IC's other than q_n, is represented by the vector matrix $\vec{R}(\widetilde{q}_n)$. The following cost function is used for the minimization of the internal strain:

$$\chi_n^S\left(\vec{R}\right) = \frac{1}{2}\sum_{m\neq n}\left[q_m(\vec{R}) - q_m(\vec{R}_0)\right]^2 \tag{3.6}$$

under the constraint $q_n(\vec{R}) = \widetilde{q}_n$. The summation runs over all internal coordinates q_m, different from the perturbed IC q_n. Each IC in the summation is expressed in atomic units as a means of preconditioning. We approximate $q_m(\vec{R})$ by its first order Taylor expansion:

$$q_m(\vec{R}(\widetilde{q}_n)) = q_m(\vec{R}_0) + \left[\nabla q_m(\vec{R}_0)\right]^T \cdot \left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right) \tag{3.7}$$

Note that the gradient matrix $\nabla q_m(\vec{R}_0)$ represents an $3N_{\text{at}} \times 1$ column matrix, and that it can be

evaluated from the first principles input on the equilibrium structure:

$$\frac{\partial q_m}{\partial R_i}\left(\vec{R}_0\right) = B_{mi} \tag{3.8}$$

Introducing this Taylor expansion in the cost function results in:

$$\chi_n^S\left(\vec{R}(\widetilde{q}_n)\right) = \frac{1}{2}\sum_{m\neq n}\left[\left[\nabla q_m(\vec{R}_0)\right]^T \cdot \left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right)\right]^2 \tag{3.9}$$

$$= \frac{1}{2}\sum_{m\neq n}\left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right)^T \cdot \left(\left[\nabla q_m(\vec{R}_0)\right]\left[\nabla q_m(\vec{R}_0)\right]^T\right) \cdot \left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right) \tag{3.10}$$

$$= \frac{1}{2}\left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right)^T \cdot \mathbf{S}_n \cdot \left(\vec{R}(\widetilde{q}_n) - \vec{R}_0\right) \tag{3.11}$$

with $\mathbf{S}_n$ the strain matrix corresponding to strain of all IC's except q_n:

$$(\mathbf{S}_n)_{ij} = \sum_{m\neq n}\left(\left[\nabla q_m(\vec{R}_0)\right]\left[\nabla q_m(\vec{R}_0)\right]^T\right)_{ij} \tag{3.12}$$

$$= (\mathbf{B}^T\mathbf{B})_{ij} - B_{ni}B_{nj} \tag{3.13}$$

Some manipulations have been applied to the strain matrix or $\mathbf{B}^T\mathbf{B}$ to reduce the condition number (see further).

This procedure of minimal internal strain yields trajectories along an internal coordinate, that are as much as possible decoupled from the other IC's. Hence, all contributions to the covalent force field energy along the trajectory will be small, except for the harmonic term related to q_n. Therefore, we approximate the covalent force field energy $V_{\text{cov}}^{\text{ff}}(\widetilde{q}_n)$ along the perturbation trajectory, by a single harmonic potential[1] in $\widetilde{q}_n$.

$$V_{\text{cov}}^{\text{ff}}(\widetilde{q}_n) = \frac{K_n}{2}\left(\widetilde{q}_n - q_{n,0}\right)^2 \tag{3.14}$$

instead of the complete covalent energy, where the variations of all bond lengths, bending angles, out-of-plane distances and dihedral angles are taken into consideration. This is a serious simplification for the

[1] As we will explain further on, we do not apply these perturbation trajectories for dihedrals. All remaining ICs have a harmonic contribution to the potential energy. Hence, it makes sense to always approximate the covalent energy by means of a harmonic term, independant of the nature of the perturbed IC.

fitting of the relevant force constants K_n and $q_{n,0}$. Equalization of the Eqs. (3.2) and (3.3), taking into account Eq. (3.14), gives a relation that the FF parameters should satisfy. We can construct multiple frames along the trajectory in which the IC q_n is constrained to various perturbed values $\widetilde{q}_n$, while all other degrees of freedom are relaxed. The above procedure can be repeated for each of these frames[2] yielding multiple relations for the force constant K_n and rest value $q_{n,0}$. These relations all express that the force field energy should be equal to the ab initio energy along the perturbation trajectory for each chosen $\widetilde{q}_n$ apart from a constant shift c:

$$V^{\mathrm{ai}}\left(\vec{R}(\widetilde{q}_n)\right) = V^{\mathrm{ff}}_{\mathrm{nb}}\left(\vec{R}(\widetilde{q}_n)\right) + \frac{K_n}{2}\left(\widetilde{q}_n - q_{n,0}\right)^2 + c \tag{3.15}$$

Hence, one can fit a parabola (as a funtion of $\widetilde{q}_n$) to the difference between V^{ai} and $V^{\mathrm{ff}}_{\mathrm{nb}}$ yielding directly an estimate of the force constant K_n and rest value $q_{n,0}$. The procedure can be repeated for each internal coordinate q_n, however, we only apply it to bond lengths, bending angles and out-of-plane distances. No dihedral angles are considered in this step of the procedure. The main reason is that the goal of this step is to get accurate estimates of the rest values, while the rest values of the dihedral angles were already determined in the previous step. Finally, an averaging procedure is applied to all IC's belonging to the same atom types. The standard deviation can be regarded as a measure to assess the quality of the atom types which compose the IC's.

Note that $q_{n,0}$ does not always coincide with the equilibrium $q_n(\vec{R}_0)$ value. We illustrate this with two examples. Figure 6 shows the different energy contributions for the perturbation trajectory of a carbon-carbon bond in benzene as well as the result of the fit. In this first example the rest value $q_{n,0}$ is almost equal to the equilibrium value, due to the weak electrostatic interaction (and excluded vdW interactions) in the C-C bond. The second example illustrates the case of a metal-oxygen bond in MIL-53(Al), where the elongation of the bond induces significant changes of the electrostatic interaction, which are of the same magnitude as the covalent bond interaction (Figure 7). The rest value of 2.06 Å differs significantly from the equilibrium value of 1.91 Å. It is one of the interesting features of this perturbation trajectory concept in step 2, that rest values can be accurately determined, which in hybrid systems, such as MOFs, can be substantially different from the equilibrium values predicted from first

[2]By default the procedure is repeated for 11 frames, with perturbations in a range of 0.05 Å for distances and 5 ° for angles.

principles.

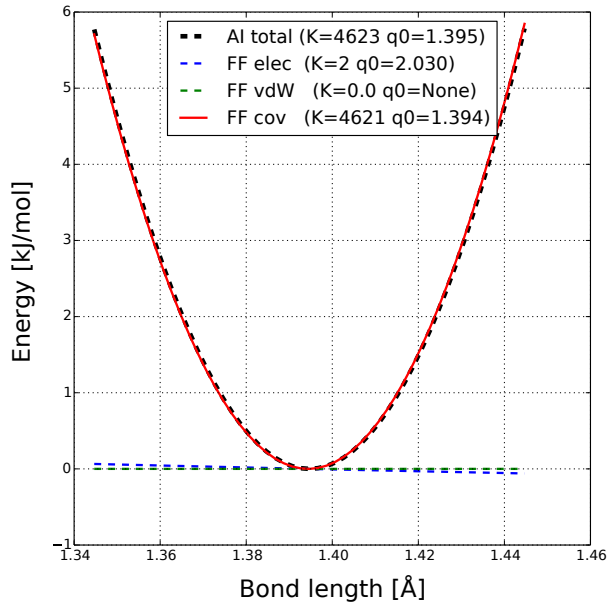

Figure S 6: Energy contributions for the perturbation trajectory $\vec{R}(\tilde{q}_n)$ in case q_n represents a C-C bond in benzene. Dashed lines represent properties that serve as input for the fit, while the solid line represents the result of the fit.

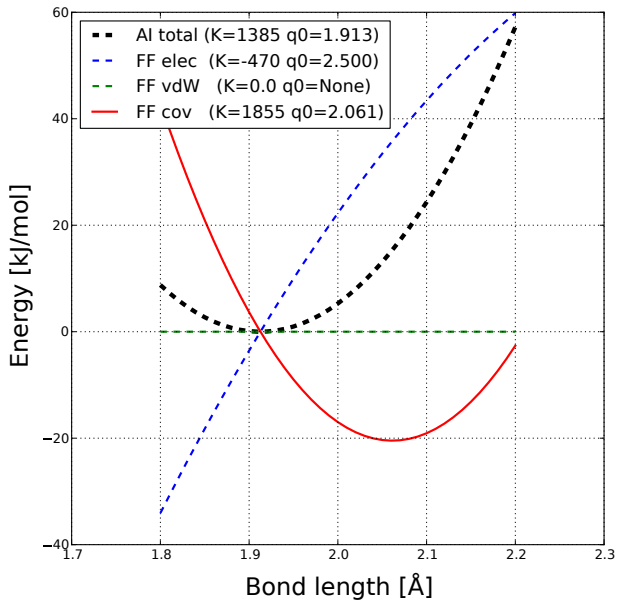

Figure S 7: Energy contributions for the perturbation trajectory $\vec{R}(\tilde{q}_n)$ in case q_n represents a $Al - O_{ca}$ bond in MIL-53(Al). Dashed lines represent properties that serve as input for the fit, while the solid line represents the result of the fit. This plot also nicely illustrates the difference between rest value and equilibrium value: the equilibrium value is the minimum of the ab initio curve, while the rest value is the minimum of the FF covalent curve.

3.2 Strain matrix

For the construction of these perturbation trajectories, a strain cost χ_n^S was introduced. This cost is minimized to relax the molecular geometry under the constraint that IC q_n is fixed at the perturbation value $q_n = \widetilde{q}_n$:

$$\chi_n^S\left(\vec{x}(\widetilde{q}_n)\right) = \frac{1}{2}\vec{x}(\widetilde{q}_n)^T \cdot \mathbf{S}^{(n)} \cdot \vec{x}(\widetilde{q}_n) \tag{3.16}$$

$$\vec{x}(\widetilde{q}_n) = \vec{R}(\widetilde{q}_n) - \vec{R}_0 \tag{3.17}$$

The strain cost is expressed by means of a strain matrix $\mathbf{S}^{(n)}$:

$$\mathbf{S}^{(n)} = \sum_{m \neq n} \left[\nabla q_m(\vec{R}_0)\right] \left[\nabla q_m(\vec{R}_0)\right]^T \tag{3.18}$$

which may be expressed in terms of the transformation matrix $\mathbf{B}$ between the IC's and the Cartesian coordinates of the atomic positions at equilibrium:

$$\left[\mathbf{S}^{(n)}\right]_{ij} = \left[\mathbf{B}^T\mathbf{B}\right]_{ij} - B_{ni}B_{nj} \tag{3.19}$$

$$= \left[\left(\mathbf{B^{(n)}}\right)^T \mathbf{B^{(n)}}\right]_{ij} \tag{3.20}$$

$$B_{mi} = \frac{\partial q_m}{\partial R_i}\left(\vec{R}_0\right) \tag{3.21}$$

Remember that $\mathbf{B}$ represents a $M \times 3N_{\text{at}}$ matrix, which has no uniquely defined inverse $\mathbf{B}^{-1}$ when the matrix $\mathbf{B}$ is overdetermined ($M > 3N_{\text{at}}$) with M the total number of internal coordinates. Matrix $\mathbf{B^{(n)}}$ is of dimension $(M-1) \times 3N_{\text{at}}$ and is identical to $\mathbf{B}$, but it does not contain the column corresponding to q_n. To make the notation not to cumbersome, from now on, we will work with matrices $\mathbf{S}^{(n)}$ and $\mathbf{B^{(n)}}$, but we will drop the superscript. Before we use this strain matrix to estimate force field parameters, we manipulate the expressions to enhance the condition number of the matrix. The IC's can be classified in classes: each IC q_n belongs to a specific class of IC's characterized by a index p ($p = 1 \ldots I$). In each class there are N_p elements, so that the index n corresponds to (p, k) with p the class of q_n and k an index distinguishing q_n from the other members of the same IC class. Assume I as the number of total IC classes, then $M = \sum_{p=1}^{I} N_p$. The strain matrix can be decomposed in strain submatrices $\mathbf{S_p}$ belonging

to each class p:

$$\mathbf{S} = \sum_{p=1}^{I} \mathbf{S_p} \tag{3.22}$$

The inner summation over the IC's in the matrix product $\mathbf{B}^T\mathbf{B}$ can now be decomposed in the different classes p:

$$[\mathbf{S}]_{ij} = \left[\mathbf{B}^T\mathbf{B}\right]_{ij} = \sum_{p=1}^{I}\sum_{k=1}^{N_p} \left[\mathbf{B}^T\right]_{i;p,k} [\mathbf{B}]_{p,k;j} = \sum_{p=1}^{I}\sum_{k=1}^{N_p} \left[\mathbf{B_p}^T\right]_{ik} [\mathbf{B_p}]_{kj} \tag{3.23}$$

This leads to:

$$[\mathbf{S_p}]_{ij} = \left[\mathbf{B_p}^T\mathbf{B_p}\right]_{ij} \tag{3.24}$$

By this decomposition the submatrices $\mathbf{B_p}$ are less overconstrained and are reduced in dimension: $N_p \times 3N_{\text{at}}$. We proceed by applying the Singular Value Decomposition to $\mathbf{B_p}$:

$$\mathbf{B_p} = \mathbf{U_p} \cdot \mathbf{\Sigma_p} \cdot \mathbf{V_p}^T \tag{3.25}$$

$\mathbf{\Sigma_p}$ is a diagonal matrix with dimension $3N_{\text{at}} \times 3N_{\text{at}}$ and contains the singular values $\{\sigma_{p,k}\}_{k=1}^{N_p}$, in order of decreasing value . $\mathbf{U_p}$ is an $N_p \times 3N_{\text{at}}$ orthogonal matrix whose columns are the eigenvectors of $\mathbf{B_p}\mathbf{B_p}^T$, $\mathbf{V_p}$ is an $3N_{\text{at}} \times 3N_{\text{at}}$ orthogonal matrix whose columns are the eigenvectors of $\mathbf{B_p}^T\mathbf{B_p}$. The contribution $\mathbf{B_p}^T\mathbf{B_p}$ to the strain matrix becomes:

$$\mathbf{B_p}^T\mathbf{B_p} = \mathbf{V_p} \cdot \mathbf{\Sigma_p}^2 \cdot \mathbf{V_p}^T \tag{3.26}$$

To enhance the performance of the strain relaxation, we manipulate the singular values:

1. Any singular value $\sigma_{p,k}$ smaller than 10^{-6} a.u, is set to $\frac{1}{3N_{at}}$
2. Any singular value $\sigma_{p,k}$ larger than 10^{-6} a.u, is set to 1

All columns of $\mathbf{V_p}$ correspond to deformations associated to ICs of class p. The columns of $\mathbf{V_p}$ corresponding to large singular values (set to 1) represent linear combinations of (changes in) internal

coordinates that are easily transformed in Cartesian coordinates, because of their linear independence. These columns are denoted as $\mathbf{V_{p,o}}$. They form the deformations in the molecular geometry yielding the main contributions to the strain $\mathbf{S_p}$. All other columns are denoted as $\mathbf{V_{p,\perp}}$, and are hence orthogonal to $\mathbf{V_{p,o}}$. The approximation imposed on the eigenvalues $\sigma_{k,p}$ to be 1 or 0, has a lot of advantages: (i) the condition number of this block is perfect (i.e. one) and (ii) any ambiguity due to the units of the internal coordinates disappears. The singular value decomposition (SVD) of $\mathbf{S_p}$ (Eqs. 3.24 and 3.26) can now be rewritten as:

$$\mathbf{S_p} = \mathbf{V_{p,o}} \cdot (\mathbf{V_{p,o}})^T + \frac{\mathbf{V_{p,\perp}} \cdot (\mathbf{V_{p,\perp}})^T}{3N_{\text{at}}} \tag{3.27}$$

The last contribution can be omitted as no strain should in principle be in associated with these spurious degrees of freedom. However, assigning a small singular value is an effective regularization technique that lowers the condition number of the total strain matrix.

4 Construction of a large set of organic molecules

The procedure to construct a set of 1100 organic molecules (including technical details) is as follows:

1. All molecules from the PubChem Compound database were downloaded that (i) contain at most 10 elements other than hydrogen, (ii) contain at least one hydrogen and one carbon atom and (iii) consist of a single covalent unit. Furthermore, we considered only molecules for which an initial guess of the 3D structure was available. This brings us to an initial set of 408316 molecules. For each molecule, the InChIKey[1,2] and the SMILES[3] code were computed with OpenBabel.[4,5]

2. All molecular structures were converted to an efficient HDF5 file for further processing.[6] At this stage, duplicates were eliminated based on the InChIKey. It was also found that some of the molecules had apparently an invalid initial geometry, i.e. with all atomic Cartesian coordinates equal to zero. These cases were also eliminated. After this check, the set was reduced to 376407 molecules.

3. The following formal criteria were introduced to select a suitable subset: (i) at most three hetero elements are allowed per molecule, (ii) molecules with elements heavier than fluorine are excluded, (iii) the formal charge of each molecule (based on the SMILES code) must be zero, (iv) the number of electrons must be even and (v) zwitterions are not allowed. A zwitterion is detected by the presence of any of the following strings in the SMILES code: [CH2-], [NH3+] or [NH2+]. After imposing these selection criteria, only 150875 molecules remained, in which the elements helium, lithium, beryllium an boron do not occur.

4. From the set of 150875 molecules, a random selection of 1100 molecules was made. For each of these molecules, the geometry was optimized and the Hessian in the equilibrium was calculated using Gaussian.[7] For 36 molecules, one or more normal mode frequencies were negative. This is due to an inefficient geometry optimization and several methods exists to solve this issue. However, these methods involve a more advanced treatment of the ab initio calculations, which is not a mandatory skill of the intended QuickFF user. Therefore, we also deleted these molecules from the set, leaving us with a set of 1064 molecules.

5 Application: Mil-53(Al)

5.1 System

The atom types and definition of the unit cell parameters are shown in Figure 8.

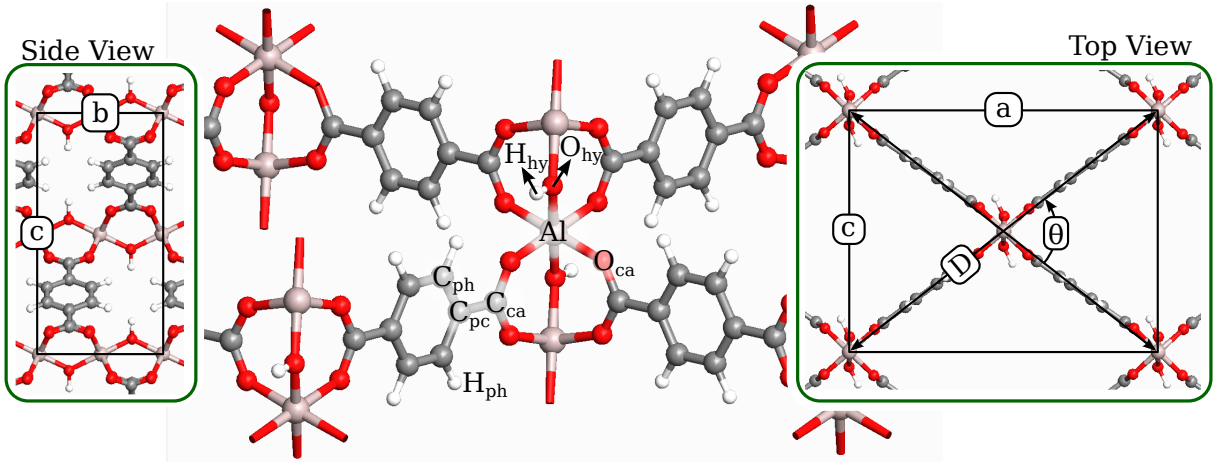

Figure S8: Illustration of the definition of atom types and unit cell dimensions in Mil-53(Al). Reprinted (adapted) with permission from ref. 8. Copyright (2014) American Chemical Society.

5.2 Force field

The parameters of the QuickFF-derived force field are shown and compared with our previous force field in the table below.

Table S3: Force field parameters from QuickFF compared with our previous force field.

Atypes	**QuickFF**				**Previous**	
	oop		**nooop**			
BONDS	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$
AL-O_{CA}	1192	2.077	1196	2.077	1079	2.151
AL-O_{HY}	2066	2.056	2070	2.056	1738	2.137
C_{CA}-C_{PC}	2169	1.496	2169	1.496	2580	1.508
C_{CA}-O_{CA}	5499	1.333	5498	1.333	5466	1.337
C_{PC}-C_{PH}	3120	1.395	3118	1.395	2897	1.398
C_{PH}-C_{PH}	3338	1.380	3338	1.380	3413	1.386
C_{PH}-H_{PH}	3394	1.086	3394	1.086	3420	1.086
H_{HY}-O_{HY}	7148	1.075	7147	1.075	7030	1.046
BENDS	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$
AL-O_{CA}-C_{CA}	114	132	123	132	58	88
AL-O_{HY}-AL					173	73
AL-O_{HY}-H_{HY}	60	115	60	115	104	85
C_{CA}-C_{PC}-C_{PH}	602	120	610	120	678	121
C_{PC}-C_{CA}-O_{CA}	159	115	143	115	136	124
C_{PC}-C_{PH}-C_{PH}	446	120	444	120	448	122
C_{PC}-C_{PH}-H_{PH}	279	120	277	120	297	121
C_{PH}-C_{PC}-C_{PH}	476	120	473	120	345	118
C_{PH}-C_{PH}-H_{PH}	348	121	349	121	351	122
O_{CA}-AL-O_{CA}			2	104	–	–
O_{CA}-AL-O_{HY}			28	74	121	91
O_{CA}-C_{CA}-O_{CA}	761	124	757	124	672	123
O_{HY}-AL-O_{HY}			0	181	–	–
DIHEDRALS	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$	**K**	$\mathbf{q_0}$
AL-O_{CA}-C_{CA}-C_{PC}	11	0	0	0	14	0
AL-O_{CA}-C_{CA}-O_{CA}			23	0	11	0
AL-O_{HY}-AL-O_{CA}			0	60	–	–
AL-O_{HY}-AL-O_{HT}			0	0	–	–
C_{CA}-C_{PC}-C_{PH}-C_{PH}	51	0	70	0	58	0
C_{CA}-C_{PC}-C_{PH}-H_{PH}	22	0	21	0	29	0
C_{PC}-C_{PH}-C_{PH}-C_{PC}	36	0	57	0	41	0
C_{PC}-C_{PH}-C_{PH}-H_{PH}	33	0	43	0	36	0
C_{PH}-C_{PC}-C_{CA}-O_{CA}	10	0	66	0	23	0
C_{PH}-C_{PC}-C_{PH}-C_{PH}	29	0	6	0	32	0
C_{PH}-C_{PC}-C_{PH}-H_{PH}	29	0	5	0	27	0
H_{PH}-C_{PH}-C_{PH}-H_{PH}	17	0	32	0	23	0
OOPDIST	**K**	$\mathbf{d_0}$	**K**	$\mathbf{d_0}$	**K**	$\mathbf{d_0}$
C_{CA}-C_{PH}-C_{PH}-C_{PC}	135	0				
C_{PC}-C_{PH}-H_{PH}-C_{PH}	149	0				
C_{PC}-O_{CA}-O_{CA}-C_{CA}	1474	−0.02				

The tables below show the ff parameters and their standard deviation of the linker and oxide clusters as obtained from the perturbation trajectories. These tables are included to be able to assess the quality of atom types based on the standard deviation. As we can see from the tables, the standard deviations are very small for the rest values. The force constants of some ICs, e.g. AL-O_{CA}-C_{CA}, can be higher, this is mainly due to the fact that those ICs are strongly correlated with other ICs and are therefore more liable to error when approximating the FF energy along the perturbation trajectory with only one harmonic contribution. However, the force constant refinement step will further tune the force constants, eliminating these errors.

Table S 4: Force field parameters and their standard deviation of the linker cluster as obtained from the perturbation trajectories.

BONDS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
C_{CA}-C_{PC}	2927.455 ± 0.658	1.493 ± 0.000
C_{PC}-C_{PH}	4199.753 ± 23.473	1.395 ± 0.000
C_{PH}-C_{PH}	4569.008 ± 8.228	1.380 ± 0.000
C_{PH}-H_{PH}	3416.378 ± 5.472	1.086 ± 0.000
BENDS	K [kJ mol^{-1} rad^{-2}]	q_0 [deg]
C_{CA}-C_{PC}-C_{PH}	1003.816 ± 74.090	119.570 ± 0.109
C_{PC}-C_{PH}-C_{PH}	1973.564 ± 65.309	119.795 ± 0.033
C_{PC}-C_{PH}-H_{PH}	631.722 ± 7.330	119.609 ± 0.056
C_{PH}-C_{PC}-C_{PH}	2110.444 ± 1.264	120.465 ± 0.023
C_{PH}-C_{PH}-H_{PH}	598.441 ± 7.078	120.719 ± 0.053
OOPDISTS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
C_{CA}-C_{PH}-C_{PH}-C_{PC}	408.771 ± 17.945	0.007 ± 0.045
C_{PC}-C_{PH}-H_{PH}-C_{PH}	941.199 ± 22.443	0.001 ± 0.005

Table S 5: Force field parameters and their standard deviation of the oxide cluster as obtained from the perturbation trajectories.

BONDS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
AL-O_{CA}	1764.410 ± 99.099	2.077 ± 0.020
AL-O_{HY}	2501.821 ± 14.122	2.056 ± 0.000
C_{CA}-C_{PC}	3074.634 ± 9.421	1.498 ± 0.001
C_{CA}-O_{CA}	6283.640 ± 55.024	1.333 ± 0.001
H_{HY}-O_{HY}	7068.934 ± 0.000	1.075 ± 0.000
BENDS	K [kJ mol^{-1} rad^{-2}]	q_0 [deg]
AL-O_{CA}-C_{CA}	1398.267 ± 210.624	131.916 ± 1.918
AL-O_{HY}-AL	1676.984 ± 0.000	125.646 ± 0.000
AL-O_{HY}-H_{HY}	1083.551 ± 6.039	114.847 ± 1.193
C_{PC}-C_{CA}-O_{CA}	1238.970 ± 13.474	118.012 ± 0.174
O_{CA}-AL-O_{CA}	4256.978 ± 133.800	91.764 ± 0.272
O_{CA}-AL-O_{HY}	2722.300 ± 185.929	94.201 ± 1.028
O_{CA}-C_{CA}-O_{CA}	1749.090 ± 3.008	124.108 ± 0.035
OOPDISTS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
AL-AL-H_{HY}-O_{HY}	2097.810 ± 0.000	0.481 ± 0.000
C_{PC}-O_{CA}-O_{CA}-C_{CA}	1086.319 ± 56.316	0.003 ± 0.070

5.3 Geometry

Table 6 below contains the values of bond lengths, bending angles and some dihedral angles in equilibrium. These values are reported for the ab initio calculations on both linker and oxide clusters as well as for the periodic force field calculations using QuickFF/oop, QuickFF/nooop and our previous force field. Table 7 contains the unit cell dimensions obtained from experiment and predicted by our previous force field as well as the two QuickFF force fields (oop and noopp), both for large pore and narrow pore phases.

Table S 6: Ab initio optimized bond lengths of the linker and oxide cluster compared with the periodic predictions made by our previous[8] force field and the present QuickFF force field. Two options are investigated in QuickFF: inclusion of out-of-plane distance terms (oop) or not (nooop).

ATYPES	**Linker**	**Oxide**	**Previous**	**QuickFF**	
				nooop	**oop**
BOND LENGTHS [Å]					
AL-O_{CA}	1.93 ± 0.01	1.92 ± 0.01	1.93 ± 0.01	1.87 ± 0.00	1.86 ± 0.00
AL-O_{HY}	–	1.86 ± 0.00	1.88 ± 0.01	1.83 ± 0.00	1.82 ± 0.00
C_{CA}-C_{PC}	1.50 ± 0.00	1.49 ± 0.00	1.50 ± 0.00	1.49 ± 0.00	1.49 ± 0.00
C_{CA}-O_{CA}	1.27 ± 0.00	1.27 ± 0.00	1.27 ± 0.00	1.26 ± 0.00	1.26 ± 0.00
C_{PC}-C_{PH}	1.40 ± 0.00	1.41 ± 0.00	1.41 ± 0.00	1.40 ± 0.00	1.41 ± 0.00
C_{PH}-C_{PH}	1.39 ± 0.00	–	1.39 ± 0.00	1.39 ± 0.00	1.39 ± 0.00
C_{PH}-H_{PH}	1.08 ± 0.00	1.08 ± 0.00	1.08 ± 0.00	1.08 ± 0.00	1.08 ± 0.00
H_{HY}-O_{HY}	–	0.96 ± 0.00	0.91 ± 0.00	0.96 ± 0.00	0.96 ± 0.00
BENDING ANGLES [deg]					
AL-O_{CA}-C_{CA}	–	133 ± 2	132 ± 1	136 ± 0	135 ± 1
AL-O_{HY}-AL	–	126 ± 0	123 ± 0	134 ± 0	134 ± 0
AL-O_{HY}-H_{HY}	–	114 ± 0	118 ± 0	113 ± 0	113 ± 0
C_{CA}-C_{PC}-C_{PH}	120 ± 0	–	121 ± 0	120 ± 0	120 ± 0
C_{PC}-C_{CA}-O_{CA}	–	118 ± 0	118 ± 0	118 ± 0	118 ± 0
C_{PC}-C_{PH}-C_{PH}	120 ± 0	–	121 ± 0	120 ± 0	120 ± 0
C_{PC}-C_{PH}-H_{PH}	119 ± 0	–	120 ± 0	120 ± 0	120 ± 0
C_{PH}-C_{PC}-C_{PH}	120 ± 0	–	118 ± 0	120 ± 0	120 ± 0
C_{PH}-C_{PH}-H_{PH}	121 ± 0	–	120 ± 0	120 ± 0	120 ± 0
O_{CA}-AL-O_{CA}	–	91 ± 0	90 ± 3	90 ± 2	90 ± 1
O_{CA}-AL-O_{HY}	–	92 ± 1	90 ± 3	90 ± 1	90 ± 2
O_{CA}-C_{CA}-O_{CA}	–	125 ± 0	124 ± 0	124 ± 0	124 ± 0
O_{HY}-AL-O_{HY}	–	–	180 ± 0	180 ± 0	178 ± 0
DIHEDRAL ANGLES [deg]					
AL-O_{CA}-C_{CA}-C_{PC}	–	171 ± 4	165 ± 9	174 ± 0	170 ± 2
AL-O_{CA}-C_{CA}-O_{CA}	–	9 ± 5	15 ± 10	5 ± 0	11 ± 3

Table S 7: Comparison of the unit cell predicted by the force field with the experimental cell parameters (a) Ref. 9 (b) Ref. 8

	Narrow pore				**Large pore**			
	Exp[a]	**QuickFF**		**Prev**[b]	**Exp**[a]	**QuickFF**		**Prev**[b]
		oop	**nooop**			**oop**	**nooop**	
a [Å]	20.82	19.32	19.22	19.57	16.91	16.66	16.05	17.05
b [Å]	6.61	6.75	6.73	6.53	6.62	6.71	6.73	6.59
c [Å]	6.87	6.10	6.10	6.25	12.67	12.77	13.59	12.91
α [deg]	90.00	89.98	95.26	87.89	90.00	89.98	90.00	90.00
β [deg]	90.00	90.20	90.41	89.43	90.00	90.01	90.00	90.00
γ [deg]	113.95	92.60	92.58	97.15	90.00	90.01	90.00	90.63
D [Å]	21.93	20.24	20.14	20.54	21.13	20.99	21.03	21.38
θ [deg]	36.52	35.04	35.15	35.41	73.68	74.95	80.49	74.25

6 Application: MOF-5

6.1 System

Figure 9 illustrates the clusters on which QuickFF was applied to. The atom types are also depicted on the figure.

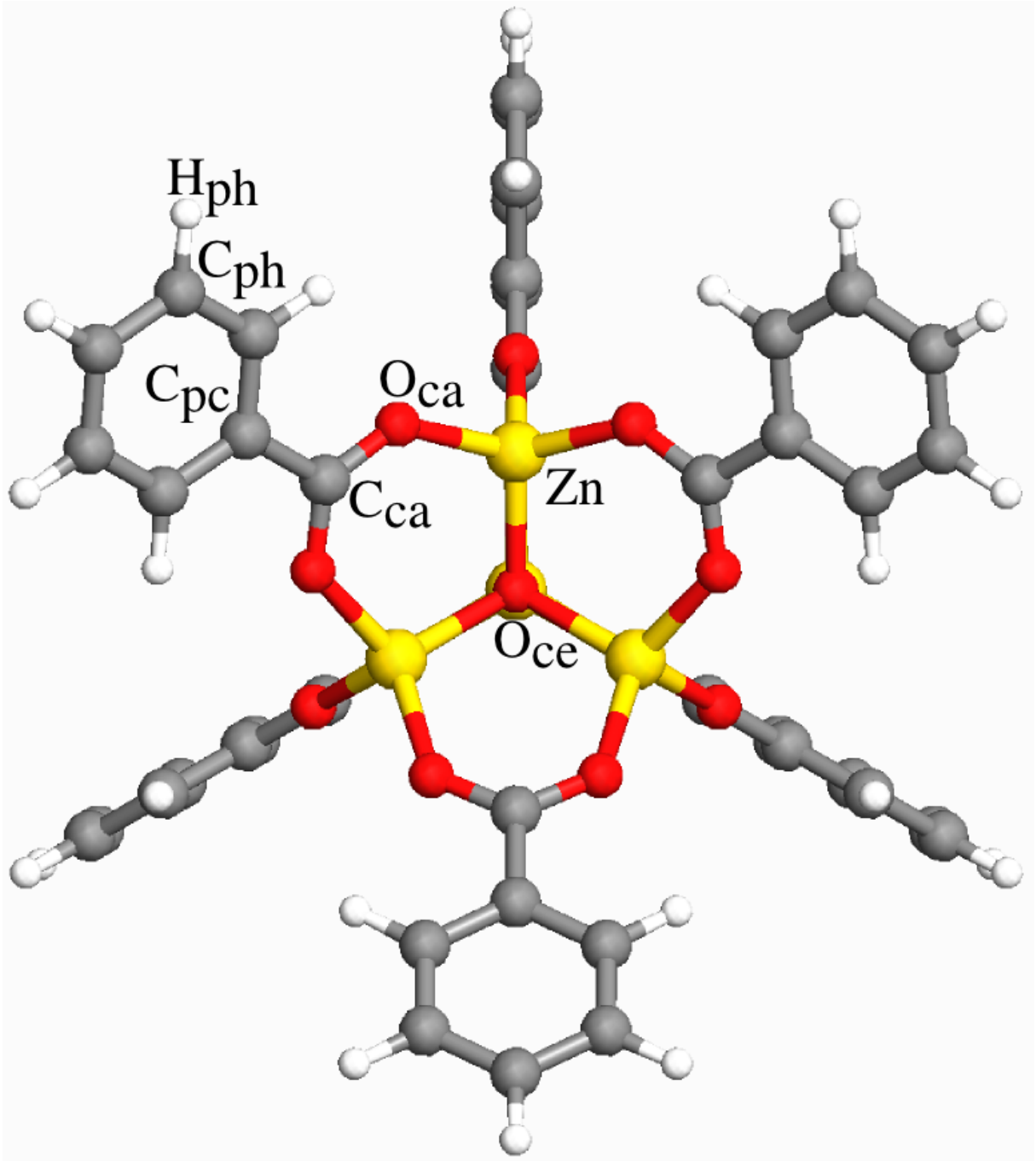

Figure S 9: Illustration of the definition of atom types in MOF-5.

6.2 Force field

QuickFF was used on the benzoate cluster as depicted in Figure 9. For the parameterization of the QuickFF force field, identical reference data was used as was for MOF-FF: the same ab initio equilibrium geometry and Hessian, identical atom types (see Figure 9), the same set of atomic charges q_i, gaussian charge radii r_i and MM3-Buckingham van der Waals parameters σ_i and ϵ_i. The electrostatic and van der Waals parameters are given in Table 8. The covalent parameters of the resulting force field is shown and compared with MOF-FF in Table 9.

Table S8: Electrostatic and van der Waals parameters needed as input for both QuickFF and MOF-FF

Atom type	**Electrostatics**		**van der Waals**	
	q_i [e]	r_i [A]	ϵ_i [kcal/mol]	σ_i [A]
Zn	1.42	2.073	0.276	2.29
O_{ca}	−0.72	1.118	0.059	1.82
O_{ce}	−1.78	1.118	0.059	1.82
C_{ca}	0.61	1.163	0.056	1.94
C_{pc}	0.18	1.163	0.056	1.94
C_{ph}	−0.12	1.163	0.056	1.94
H_{ph}	0.12	0.724	0.020	1.50

Table S 9: Covalent force-field parameters for MOF-5 of QuickFF and MOF-FF

	MOF-FF		QuickFF	
BOND	K_r	r_0	K_r	r_0
C_{ca}-C_{pc}	2973	1.488	2087	1.502
C_{ca}-O_{ca}	5198	1.275	4551	1.278
C_{pc}-C_{ph}	4264	1.394	3201	1.405
C_{ph}-C_{ph}	4264	1.394	3274	1.398
C_{ph}-H_{ph}	3270	1.094	3353	1.091
O_{ca}-Zn	1003	1.917	779	1.917
O_{ce}-Zn	897	1.987	655	1.983
BEND	K_θ	θ_0	K_θ	θ_0
C_{ca}-C_{pc}-C_{ph}	485	117	588	120
C_{ca}-O_{ca}-Zn	57	136	151	133
C_{pc}-C_{ca}-O_{ca}	646	116	205	118
C_{pc}-C_{ph}-C_{ph}	446	127	464	120
C_{pc}-C_{ph}-H_{ph}	303	120	308	119
C_{ph}-C_{pc}-C_{ph}	446	127	438	119
C_{ph}-C_{ph}-C_{ph}	446	127	480	120
C_{ph}-C_{ph}-H_{ph}	303	120	322	120
O_{ca}-C_{ca}-O_{ca}	936	123	764	123
O_{ca}-Zn-O_{ca}	48	123	25	109
O_{ca}-Zn-O_{ce}	0	114	0	111
Zn-O_{ce}-Zn	420	104	49	110

	MOF-FF		QuickFF	
TORSION	K_ϕ	ϕ_0	K_ϕ	ϕ_0
C_{ca}-C_{pc}-C_{ph}-C_{ph}	0	0	50	0
C_{ca}-C_{pc}-C_{ph}-H_{ph}	0	0	19	0
C_{ca}-O_{ca}-Zn-O_{ca}	0	0	3	0
C_{ca}-O_{ca}-Zn-O_{ce}	0	0	−0	0
C_{pc}-C_{ca}-O_{ca}-Zn	8	0	−0	0
C_{pc}-C_{ph}-C_{ph}-C_{pc}	18	0	33	0
C_{pc}-C_{ph}-C_{ph}-H_{ph}	26	0	28	0
C_{ph}-C_{pc}-C_{ca}-O_{ca}	8	0	8	0
C_{ph}-C_{pc}-C_{ph}-C_{ph}	18	0	32	0
C_{ph}-C_{pc}-C_{ph}-H_{ph}	26	0	32	0
H_{ph}-C_{ph}-C_{ph}-H_{ph}	0	0	15	0
O_{ca}-C_{ca}-O_{ca}-Zn	0	0	−0	0
O_{ca}-Zn-O_{ce}-Zn	0	0	−0	0
OOPDIST	K_d	d_0	K_d	d_0
C_{ca}-C_{ph}-C_{ph}-C_{pc}			346	0.00
C_{pc}-C_{ph}-H_{ph}-C_{ph}			278	0.00
C_{pc}-O_{ca}-O_{ca}-C_{ca}			1706	0.00
C_{ph}-C_{ph}-H_{ph}-C_{ph}			244	0.00
OOPANGLE	K_ψ	ψ_0	K_ψ	ψ_0
C_{ca}-C_{pc}-O_{ca}-O_{ca}	113	0.00		
C_{pc}-C_{ca}-C_{ph}-C_{ph}	53	0.00		

The tables below again show the ff parameters and their standard deviation as obtained from the perturbation trajectories. The same conclusions can be made as in the case of Mil-53(Al).

Table S 10: Force field parameters and their standard deviation as obtained from the perturbation trajectories.

BONDS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
C_{ca}-C_{pc}	2928.932 ± 0.657	1.502 ± 0.000
C_{ca}-O_{ca}	5396.727 ± 0.661	1.278 ± 0.000
C_{pc}-C_{ph}	4436.413 ± 0.386	1.405 ± 0.000
C_{ph}-C_{ph}	4428.605 ± 97.003	1.398 ± 0.002
C_{ph}-H_{ph}	3367.249 ± 27.697	1.091 ± 0.002
O_{ca}-Zn	1015.739 ± 0.471	1.917 ± 0.000
O_{ce}-Zn	1773.410 ± 0.359	1.983 ± 0.000
BENDS	K [kJ mol^{-1} rad^{-2}]	q_0 [deg]
C_{ca}-C_{pc}-C_{ph}	1134.645 ± 0.089	120.342 ± 0.000
C_{ca}-O_{ca}-Zn	873.290 ± 0.262	133.033 ± 0.000
C_{pc}-C_{ca}-O_{ca}	1145.518 ± 0.359	118.479 ± 0.000
C_{pc}-C_{ph}-C_{ph}	2206.740 ± 0.156	120.183 ± 0.000
C_{pc}-C_{ph}-H_{ph}	954.425 ± 0.168	119.169 ± 0.000
C_{ph}-C_{pc}-C_{ph}	1918.204 ± 0.080	119.468 ± 0.000
C_{ph}-C_{ph}-C_{ph}	1908.858 ± 77.611	120.081 ± 0.003
C_{ph}-C_{ph}-H_{ph}	724.275 ± 66.068	119.961 ± 0.231
O_{ca}-C_{ca}-O_{ca}	1488.539 ± 0.022	123.399 ± 0.000
O_{ca}-Zn-O_{ca}	911.791 ± 0.199	108.808 ± 0.000
O_{ca}-Zn-O_{ce}	1675.501 ± 0.102	110.949 ± 0.000
Zn-O_{ce}-Zn	1477.403 ± 0.100	109.598 ± 0.000
OOPDISTS	K [kJ mol^{-1} Å^{-2}]	q_0 [Å]
C_{ca}-C_{ph}-C_{ph}-C_{pc}	1197.081 ± 0.169	0.000 ± 0.000
C_{pc}-C_{ph}-H_{ph}-C_{ph}	1300.538 ± 0.390	0.000 ± 0.000
C_{pc}-O_{ca}-O_{ca}-C_{ca}	1787.092 ± 0.146	0.000 ± 0.000
C_{ph}-C_{ph}-H_{ph}-C_{ph}	1211.564 ± 28.286	0.000 ± 0.000

6.3 Geometry and unit cell

As a first validation, both the cluster and periodic structure were optimized and their geometries were compared with the ab initio geometry of the clusters. Table 11 compares the equilibrium geometry of the benzoate cluster according to DFT (i.e. the reference data), QuickFF and MOF-FF. Table 12 compares the equilibrium geometry of the periodic MOF-5 as predicted by QuickFF and MOF-FF. Finally, the dimensions of the periodic unit cell are compared in Table 13.

Table S 11: Comparison of the equilibrium geometry of the benzoate cluster.

ATYPES	**Clus DFT**	**Clus QuickFF**	**Clus MOF-FF**
C_{ca}-C_{pc}	1.497 ± 0.000	1.510 ± 0.000	1.497 ± 0.000
C_{ca}-O_{ca}	1.271 ± 0.000	1.274 ± 0.000	1.271 ± 0.000
C_{pc}-C_{ph}	1.404 ± 0.000	1.413 ± 0.000	1.402 ± 0.000
C_{ph}-C_{ph}	1.398 ± 0.002	1.402 ± 0.000	1.399 ± 0.001
C_{ph}-H_{ph}	1.090 ± 0.001	1.090 ± 0.001	1.093 ± 0.001
O_{ca}-Zn	1.960 ± 0.000	1.968 ± 0.000	1.962 ± 0.000
O_{ce}-Zn	1.963 ± 0.000	1.921 ± 0.000	1.967 ± 0.000
C_{ca}-C_{pc}-C_{ph}	120 ± 0	121 ± 0	120 ± 0
C_{ca}-O_{ca}-Zn	131 ± 0	130 ± 0	131 ± 0
C_{pc}-C_{ca}-O_{ca}	117 ± 0	117 ± 0	117 ± 0
C_{pc}-C_{ph}-C_{ph}	120 ± 0	120 ± 0	120 ± 0
C_{pc}-C_{ph}-H_{ph}	119 ± 0	120 ± 0	121 ± 0
C_{ph}-C_{pc}-C_{ph}	120 ± 0	119 ± 0	120 ± 0
C_{ph}-C_{ph}-C_{ph}	120 ± 0	120 ± 0	120 ± 0
C_{ph}-C_{ph}-H_{ph}	120 ± 0	120 ± 0	120 ± 0
O_{ca}-C_{ca}-O_{ca}	126 ± 0	126 ± 0	126 ± 0
O_{ca}-Zn-O_{ca}	108 ± 0	106 ± 0	108 ± 0
O_{ca}-Zn-O_{ce}	111 ± 0	113 ± 0	111 ± 0
Zn-O_{ce}-Zn	109 ± 0	109 ± 0	109 ± 0
C_{ca}-C_{pc}-C_{ph}-C_{ph}	180 ± 0	180 ± 0	180 ± 0
C_{ca}-C_{pc}-C_{ph}-H_{ph}	0 ± 0	0 ± 0	0 ± 0
C_{ca}-O_{ca}-Zn-O_{ca}	122 ± 0	124 ± 0	122 ± 0
C_{ca}-O_{ca}-Zn-O_{ce}	0 ± 0	0 ± 0	0 ± 0
C_{pc}-C_{ca}-O_{ca}-Zn	180 ± 0	180 ± 0	180 ± 0
C_{pc}-C_{ph}-C_{ph}-C_{ph}	0 ± 0	0 ± 0	0 ± 0
C_{pc}-C_{ph}-C_{ph}-H_{ph}	180 ± 0	180 ± 0	180 ± 0
C_{ph}-C_{pc}-C_{ca}-O_{ca}	90 ± 90	90 ± 90	90 ± 90
C_{ph}-C_{pc}-C_{ph}-C_{ph}	0 ± 0	0 ± 0	0 ± 0
C_{ph}-C_{ph}-C_{pc}-C_{ph}	0 ± 0	0 ± 0	0 ± 0
C_{ph}-C_{ph}-C_{ph}-C_{ph}	0 ± 0	0 ± 0	0 ± 0
H_{ph}-C_{ph}-C_{pc}-C_{ph}	180 ± 0	180 ± 0	180 ± 0
H_{ph}-C_{ph}-C_{ph}-C_{ph}	180 ± 0	180 ± 0	180 ± 0
H_{ph}-C_{ph}-C_{ph}-H_{ph}	0 ± 0	0 ± 0	0 ± 0
O_{ca}-C_{ca}-O_{ca}-Zn	0 ± 0	0 ± 0	0 ± 0
O_{ca}-Zn-O_{ce}-Zn	80 ± 57	80 ± 57	80 ± 57
Zn-O_{ca}-C_{ca}-O_{ca}	0 ± 0	0 ± 0	0 ± 0
C_{ca}-C_{ph}-C_{ph}-C_{pc}	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
C_{pc}-C_{ph}-H_{ph}-C_{ph}	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
C_{pc}-O_{ca}-O_{ca}-C_{ca}	0.000 ± 0.000	-0.000 ± 0.000	0.000 ± 0.000
C_{ph}-C_{ph}-H_{ph}-C_{ph}	0.000 ± 0.000	-0.000 ± 0.000	0.000 ± 0.000

Table S 12: Comparison of the equilibrium geometry of the periodic MOF-5.

ATYPES	**Per QuickFF**	**Per MOF-FF**
C_{ca}-C_{pc}	1.511 ± 0.000	1.498 ± 0.000
C_{ca}-O_{ca}	1.273 ± 0.000	1.271 ± 0.000
C_{pc}-C_{ph}	1.413 ± 0.000	1.402 ± 0.000
C_{ph}-C_{ph}	1.402 ± 0.000	1.398 ± 0.000
C_{ph}-H_{ph}	1.092 ± 0.000	1.094 ± 0.000
O_{ca}-Zn	1.968 ± 0.000	1.964 ± 0.000
O_{ce}-Zn	1.927 ± 0.000	1.970 ± 0.000
C_{ca}-C_{pc}-C_{ph}	121 ± 0	120 ± 0
C_{ca}-O_{ca}-Zn	130 ± 0	131 ± 0
C_{pc}-C_{ca}-O_{ca}	117 ± 0	117 ± 0
C_{pc}-C_{ph}-C_{ph}	121 ± 0	120 ± 0
C_{pc}-C_{ph}-H_{ph}	120 ± 0	121 ± 0
C_{ph}-C_{pc}-C_{ph}	119 ± 0	120 ± 0
C_{ph}-C_{ph}-H_{ph}	120 ± 0	119 ± 0
O_{ca}-C_{ca}-O_{ca}	126 ± 0	126 ± 0
O_{ca}-Zn-O_{ca}	106 ± 0	108 ± 0
O_{ca}-Zn-O_{ce}	112 ± 0	111 ± 0
Zn-O_{ce}-Zn	109 ± 0	109 ± 0
C_{ca}-C_{pc}-C_{ph}-C_{ph}	180 ± 0	180 ± 0
C_{ca}-C_{pc}-C_{ph}-H_{ph}	0 ± 0	0 ± 0
C_{ca}-O_{ca}-Zn-O_{ca}	123 ± 0	122 ± 0
C_{ca}-O_{ca}-Zn-O_{ce}	0 ± 0	0 ± 0
C_{pc}-C_{ca}-O_{ca}-Zn	180 ± 0	180 ± 0
C_{pc}-C_{ph}-C_{ph}-C_{pc}	0 ± 0	0 ± 0
C_{pc}-C_{ph}-C_{ph}-H_{ph}	180 ± 0	180 ± 0
C_{ph}-C_{pc}-C_{ph}-C_{ph}	0 ± 0	0 ± 0
C_{ph}-C_{pc}-C_{ph}-H_{ph}	180 ± 0	–
C_{ph}-C_{ph}-C_{pc}-C_{ph}	0 ± 0	0 ± 0
H_{ph}-C_{ph}-C_{ph}-H_{ph}	0 ± 0	0 ± 0
O_{ca}-C_{ca}-C_{pc}-C_{ph}	90 ± 90	–
O_{ca}-Zn-O_{ce}-Zn	80 ± 57	80 ± 57
Zn-O_{ca}-C_{ca}-O_{ca}	0 ± 0	–
C_{ca}-C_{ph}-C_{ph}-C_{pc}	-0.000 ± 0.000	-0.000 ± 0.000
C_{pc}-C_{ph}-H_{ph}-C_{ph}	0.000 ± 0.000	0.000 ± 0.000
C_{pc}-O_{ca}-O_{ca}-C_{ca}	-0.000 ± 0.000	-0.000 ± 0.000

Table S 13: Unit cell of MOF-5 according to different force fields

	QuickFF	**MOF-FF**
a [A]	26.173	26.080
b [A]	26.173	26.080
c [A]	26.173	26.080
α [deg]	90.000	90.000
β [deg]	90.000	90.000
γ [deg]	90.000	90.000

6.4 NMA frequencies

As a second validation, we investigate the NMA frequencies of both the benzoate cluster and the periodic MOF-5. For the case of QuickFF, the force field Hessian of both the cluster as the periodic structure was calculated by means of analytical first order derivatives combined with finite difference second order derivatives. For the case of DFT and MOF-FF, the Hessian was taken from Ref. 10. For all three Hessians, the external degrees of freedom were projected out followed by a regular NMA analasys (using TAMkin). Table 14 compares the DFT, QuickFF and MOF-FF frequencies of the benzoate cluster. Table 15 compares the QuickFF and MOF-FF frequencies of the benzoate cluster. No DFT frequencies are at hand for the periodic structure because no periodic DFT calculations were performed.

Table S 14: Comparison between DFT, QuickFF and MOF-FF of the first 40 NMA frequencies of the benzoate cluster.

Mode	**Frequency [1/cm] (degeneracy)**			**Mode**	**Frequency [1/cm] (degeneracy)**		
	Clus DFT	Clus QuickFF	Clus MOF-FF		Clus DFT	Clus QuickFF	Clus MOF-FF
1	8 (3)	2 (3)	7 (3)	21	235 (2)	343 (3)	247 (2)
2	13 (3)	4 (3)	12 (3)	22	263 (1)	435 (6)	322 (3)
3	18 (1)	32 (3)	26 (1)	23	320 (3)	454 (4)	327 (3)
4	30 (3)	34 (2)	31 (3)	24	322 (3)	457 (2)	391 (6)
5	34 (2)	36 (4)	32 (2)	25	417 (6)	481 (3)	472 (3)
6	42 (3)	50 (3)	41 (3)	26	456 (3)	482 (3)	474 (3)
7	53 (3)	55 (2)	55 (3)	27	457 (3)	498 (3)	479 (3)
8	63 (1)	77 (3)	75 (2)	28	488 (3)	502 (3)	484 (2)
9	65 (2)	79 (3)	83 (6)	29	495 (2)	503 (3)	486 (1)
10	73 (3)	126 (4)	92 (1)	30	498 (1)	572 (6)	550 (3)
11	96 (3)	131 (1)	119 (3)	31	518 (3)	630 (2)	573 (3)
12	115 (3)	139 (3)	120 (3)	32	606 (3)	632 (3)	587 (3)
13	120 (3)	147 (3)	125 (1)	33	613 (3)	638 (1)	637 (3)
14	134 (2)	152 (2)	144 (3)	34	626 (6)	690 (3)	640 (3)
15	145 (3)	181 (3)	149 (3)	35	685 (2)	691 (3)	644 (3)
16	149 (3)	204 (3)	157 (2)	36	686 (3)	735 (2)	645 (3)
17	155 (1)	212 (3)	179 (3)	37	691 (1)	738 (3)	686 (2)
18	186 (3)	218 (3)	196 (3)	38	703 (3)	743 (1)	688 (6)
19	195 (3)	227 (3)	235 (3)	39	704 (3)	765 (3)	692 (3)
20	226 (3)	337 (3)	246 (1)	40	721 (3)	768 (3)	694 (1)

Table S15: Comparison between QuickFF and MOF-FF of the first 40 NMA frequencies of the periodic MOF-5.

Mode	**Frequency [1/cm] (degeneracy)**	
	Per QuickFF	Per MOF-FF
1	5 (3)	7 (3)
2	16 (3)	21 (9)
3	20 (3)	27 (6)
4	21 (3)	28 (3)
5	32 (12)	29 (6)
6	37 (3)	35 (3)
7	41 (6)	37 (3)
8	42 (6)	38 (1)
9	48 (6)	41 (6)
10	49 (3)	42 (9)
11	50 (6)	44 (3)
12	51 (7)	48 (3)
13	52 (2)	49 (2)
14	54 (12)	52 (3)
15	56 (3)	53 (6)
16	60 (2)	59 (6)
17	64 (3)	60 (3)
18	72 (3)	64 (3)
19	73 (6)	73 (3)
20	74 (9)	76 (6)

Mode	**Frequency [1/cm] (degeneracy)**	
	Per QuickFF	Per MOF-FF
21	79 (3)	78 (1)
22	81 (6)	79 (5)
23	82 (3)	82 (2)
24	83 (6)	83 (6)
25	85 (3)	89 (3)
26	97 (3)	90 (7)
27	101 (3)	91 (3)
28	103 (3)	93 (3)
29	115 (6)	95 (3)
30	116 (6)	96 (3)
31	118 (3)	98 (6)
32	119 (3)	99 (1)
33	123 (3)	100 (3)
34	126 (3)	104 (3)
35	127 (6)	106 (3)
36	128 (1)	112 (3)
37	129 (5)	120 (3)
38	131 (3)	121 (3)
39	132 (1)	123 (9)
40	134 (6)	127 (6)

6.5 Infrared spectrum

To calculate the infrared spectrum, a molecular dynamics simulation of the unit cell was performed in the NVT ensemble at 300 K. The Nose-Hoover thermostat was used with a chainlength of 1 and a timeconstant of 100 fs. The equations of motion were integrated with a timestep of 1.0 fs. First 100 ps of equilibration were performed, followed by 1 ns sampling time in which a sample was taken every 20 fs. The infrared absorption spectrum is shown in Figure 10 and compared with the MOF-FF simulated spectrum and the experimental[11] one.

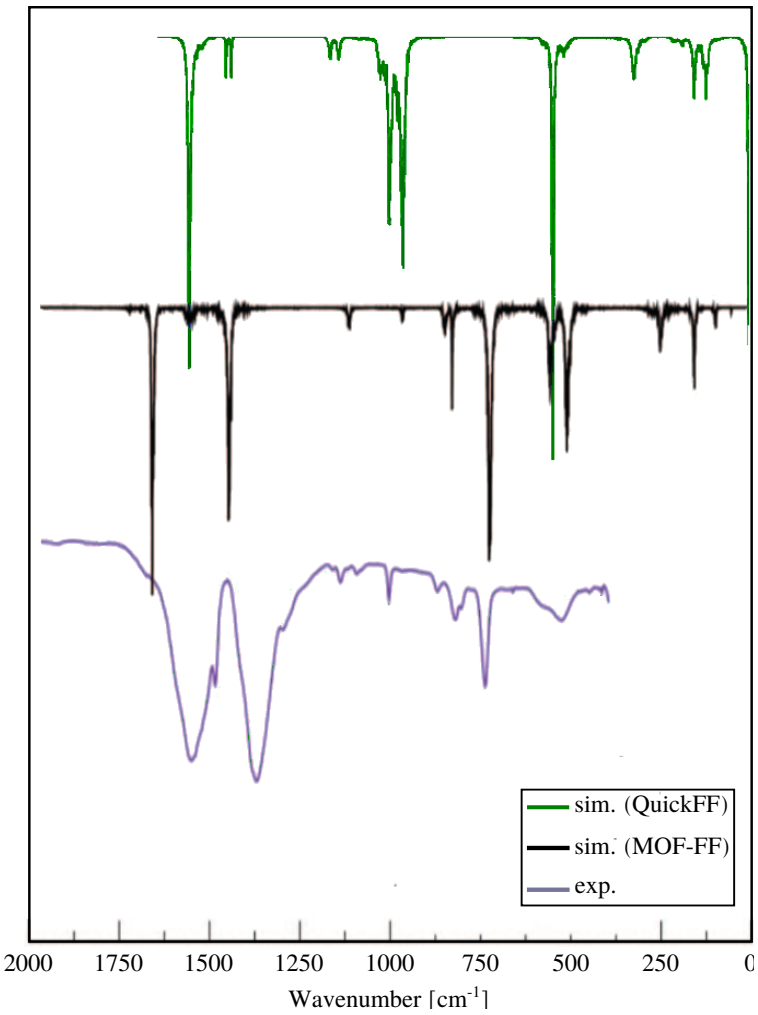

Figure S 10: Infrared absorption spectrum at 300 K predicted by QuickFF using NVT-MD.

References

(1) Stein, S.; Heller, S.; Tchekhovski, D. An Open Standard for Chemical Structure Representation - The IUPAC Chemical Identifier. http://www.inchi-trust.org/home/.

(2) http://www.jcheminf.com/content/4/1/39.

(3) http://dx.doi.org/10.1021/ci00057a005.

(4) http://openbabel.org/.

(5) http://www.jcheminf.com/content/3/1/33.

(6) The HDF Group. Hierarchical data format version 5, 2000-2013, http://www.hdfgroup.org/HDF5.

(7) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas, O.; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. Gaussian 09 Revision A.02, Gaussian Inc. Wallingford CT 2009.

(8) Vanduyfhuys, L.; Verstraelen, T.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. *J. Chem. Theory Comput.* **2012**, *8*, 3217–3231.

(9) Liu, Y.; Her, J.; Dailly, A.; Ramirez-Cuesta, A.; Neumann, D.; Brown, C. *J. Am. Chem. Soc.* **2008**, *130*, 11813–11818.

(10) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. *Phys. Status Solidi B* **2013**, *250*, 1128–1141.

(11) Hermes, S.; Schroeder, F.; Amirjalayer, S.; Schmid, R.; Fischer, R. *J. Mater. Chem.* **2006**, *16*, 2464.