Ab initio derived force fields for Zeolitic Imidazolate Frameworks: MOF-FF for ZIFs

Johannes P. Dürholt,† Guillaume Fraux,‡ François-Xavier Coudert,*‡ and Rochus Schmid*†

†Computational Materials Chemistry group, Lehrstuhl für Anorganische Chemie 2, Ruhr-Universität Bochum, Bochum, Germany
‡Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, Paris, France

E-mail: fx.coudert@chimieparistech.psl.eu; rochus.schmid@rub.de

Abstract

In this paper we parameterized in a consistent way a new force field for a range of different zeolitic imidazolate framework systems (ZIF-8, ZIF-8(H), ZIF-8(Br) and ZIF-8(Cl)), extending the MOF-FF parameterization methodology in two aspects. First, we implemented the possibility to use periodic reference data in order to prevent the difficulty of generating representative finite clusters. Second, a more efficient global optimizer based on the covariance matrix adaptation evolutionary strategy (CMA-ES) was employed during the parameterization process. We confirmed that CMA-ES, as a state-of-the-art black box optimizer for problems on continuous variables, is more suitable for force field optimization than the previous genetic algorithm. The obtained force field was then fully validated with respect to static and dynamic properties. Much effort was spent to ensure that the FF is able to describe the crucial linker swing effect in a large number of ZIF-8 derivatives. For this reason we compared our force field to
ab initio molecular dynamic simulations and found an accuracy comparable to those
obtained by different exchange–correlation functionals.

1 Introduction

Zeolitic imidazolate frameworks (ZIFs) are a subclass of metal-organic frameworks,1–4 con-
sisting of imidazolate linkers that bridge metal cations to form three-dimensional porous
crystalline solids, which are isomorphous to zeolitic frameworks.5 Like other MOFs, ZIFs
are neither static nor rigid, instead they exhibit different types of flexibility. Flexibility in
MOF's means that, upon external stimuli such as temperature, mechanical pressure or guest
molecule adsorption, the cell size and shape, and therefore the pore size and geometry can
change drastically — yet reversibly.6,7 Different manifestations of flexibility in MOFs include,
for example, negative thermal expansion,8 gate opening,9 breathing,10 negative gas adsorp-
tion11 or intraframework dynamics.5 Examples for flexibility in ZIFs are the reversible phase
transition upon cooling in ZIF-412,13 or the gate-opening effect in ZIF-7 when exposed to light
alkenes and alkanes.14,15 But the most prominent case is the so-called swing effect in ZIF-8
which can be classified as an example of intraframework dynamics. ZIF-8 is a low density
porous framework with the sodalite (sod) topology and chemical formula Zn(mim)$_2$ where
mim = 2-methylimidazolate. The sod topology features large spherical cages separated
by 6-ring windows of small aperture, and 4-rings that connect different cages (see Fig. S1).
By torsional motions of its imidazolate linkers, the frameworks can adsorb molecules with
a kinetic diameter larger than its geometric window size.16,17 It was recently shown by ab
initio molecular dynamic simulations that functionalization of the organic linker can have a
substantial influence on the swing effect.18

ZIFs present great potential for technical applications like gas capture (CO$_2$ in partic-
ular),19 sensing,20 encapsulation and controlled delivery,21 fluid separation22–24 and energy
storage25 For example, ZIF-8(Cl) exhibits a spring behavior with the highest amount of en-
nergy stored ever in high pressure intrusion-extrusion experiments.\textsuperscript{26} To exploit this potential, it is pivotal to gain an atomistic understanding of the underlying mechanisms, making molecular simulations a valuable tool, and the accurate description of the frameworks’ flexibility a crucial goal. Furthermore computational high-throughput studies based on atomistic simulations are nowadays an important approach to preselect promising candidates for further experimental studies.\textsuperscript{27} There is thus a strong need for molecular dynamics (MD) simulations of the flexibility of these MOFs, performed with high accuracy.

However the use of periodic Density Functional Theory (DFT) calculations, and in particular DFT-based MD (also called \textit{ab initio} MD) for such studies is limited to comparatively small length and time scales — because the computational cost increases very fast with both the size and number of atoms in the unit cell, and the necessary sampling time. As a consequence computationally more efficient parameterized force fields (FFs) have been applied where possible. As long as bond breaking is not involved, these more approximate molecular mechanics methods can be used to investigate larger systems for longer time scales. The difficulty is to define an energy expression that describes the relevant part of the potential energy surface with good accuracy and to determine the corresponding parameters. For computing the conformational flexibility of the porous MOF or ZIF matrix (as well as for the host–guest interactions, dominated by physisorption) conventional non-reactive force fields that employ a separation in bonded and non-bonded terms, are a sufficiently good approximation. However, the determination of parameters for the hybrid organic–inorganic part remains a challenging problem.

A frequently employed solution is to use so-called generic force fields, like UFF, where the parameters are generated by a rule based system from a much smaller set of atomic parameters.\textsuperscript{28} This allows the consistent treatment of a wide range of systems, including a number of MOFs, however, with a very limited and uncontrolled accuracy. This approach has recently been extended to MOFs, with the UFF4MOF extension of the UFF atomic parameter set.\textsuperscript{29–31} For most of the computational studies dealing with ZIFs, the same approach was
followed, by combining generic force field parameters from UFF and AMBER, modified in an 
*ad hoc* manner to obtain experimentally observed properties. The drawback of this approach 
is twofold: first, the force fields obtained may not describe physical properties that were not 
considered in the adjustment procedure. Second, the manual adjustment may lead to the 
correct macroscopic observables, at the price of an unphysical microscopic picture. A brief 
overview of the published force fields dealing with ZIFs is given in the following paragraph.

In 2012 Jiang *et al.* developed a flexible force field for ZIF-8. The equilibrium bond 
lengths and angles were set to the experimentally measured average values. Force constants 
for the organic linkers were adopted from the Amber force field. The parameters involving 
the Zn atoms were derived by fitting to experimental lattice constants. In 2013 the 
same group published a refined version of their original FF in order to be able to model 
the sorption-induced structural transition by hybrid MC/MD simulations. The crucial tor-
sions around the Zn atom were fitted to a single experimental N\textsubscript{2} isotherm. In a parallel 
effort, several FFs for ZIF-8 were developed by Demontis *et al.* In 2011 they published 
partial charges for several ZIF systems derived from cluster calculations and periodic DFT 
calculations. In 2012 the first FF based on these partial charges was published, which 
is again based on Amber. Missing parameters for the organic linker were obtained via the 
*parmcal* software, which calculates bond length and bond angle parameters based on em-
pirical rules. The parameters necessary to describe the tetrahedral ZnN\textsubscript{4} were taken from 
cluster based quantum-chemical calculations performed in order to be able to describe 
Zn containing biomolecules by Amber. In 2014 this group published a new FF based on a 
force-matching parameterization scheme. Structures and forces obtained from periodic *ab initio* MD simulations have been used as reference data, but it needs to be noted that the 
authors claim that already a trajectory of 500 fs contains sufficient reference data. Zheng *et al.* 
published an additional FF for ZIF-8 in 2013. They added an artificial long range bond 
between the carbon atoms in neighboring imidazolate linker in a 6-ring window in order to 
be able to describe the “swing effect” in the correct way. Wu *et al.* also developed a force
field for ZIF-8, again based on Amber, UFF and experimental data. The parameters for the
tetrahedral ZnN$_4$ were adopted from the already mentioned force field by Jiang et al. from
2013.$^{34}$ In 2018 Verploegh et al. published a study on the molecular diffusion in binary mixed
linker ZIFs using a flexible FF called intraZIF-FF, whose parameterization methodology has
not been published to date.$^{43}$

It is thus clear that the majority of these force fields have been parameterized in a
very ad hoc manner, often mixing different sources, ranging from different other force fields,
experimental and theoretical reference data. Clearly, the advantage of such system-specific
FFs is their partially higher accuracy, since they can be tuned in order to reproduce specific
properties or effects. As an example, Zheng et al. added an artificial bond to reproduce the
swing effect. However, this approach — and the resulting force field — can not easily be
transferred to other systems, making it difficult to use them for example in high-throughput
studies, or for novel systems where little data is available.

In order to overcome the mentioned shortcomings of generic FFs and ad hoc parameter-
ized FFs, we have developed a consistent parameterization strategy over the years with the
intention to trade transferability for accuracy. It is based on a machine learning approach
using evolutionary algorithms to derive all parameters at once, based on first principles
data.$^{44}$ This methodology is called MOF-FF and is based on a fit with respect to $ab\ initio$
calculated structure and Hessian $H$.$^{45}$ Recently we have extended our methodology in order
to parameterize also coarse-grained FFs for MOFs.$^{46-48}$

In the meantime, other groups have published related approaches to derive FFs for MOFs
in a consistent fashion,$^{49,50}$ including the Quick-FF methodology, which uses a different
approach to derive the parameters from the same reference data as in MOF-FF.$^{51}$ MOF-
FF and Quick-FF force fields are available for a variety of different MOF families. Yet,
force fields for the important family of ZIFs are still missing. A reason for this could be
that until recently both methodologies were not capable of treating periodic reference data,
which is almost necessary for deriving a ZIF force field from scratch since it is very difficult
to construct a representative cluster model for ZIFs. Recently Vanduyfhuys et al. extended the Quick-FF methodology in a way that also periodic reference data can be treated.\textsuperscript{52}

In this study we improved the MOF-FF methodology in several aspects in order to be able to parameterize force fields for ZIFs of varying topology and chemical composition (ZIF-8, ZIF-8(H), ZIF-8(Cl) and ZIF-8(Br)). Whereas several \textit{ad hoc} potentials for ZIF-8 exist, specific parameter sets for the halogenated systems are not available. We thus implemented the possibility of handling periodic reference data and introduced a more efficient optimizer. The obtained FFs were then validated in great detail against experimental and \textit{ab initio} calculated data. We focused especially on the question how well the ZIF flexibility (swing effect) is reproduced by our FFs, as it is crucial to their description and rather difficult to capture. Furthermore, we questioned the transferability of our FFs by applying them to polymorphs for which they were not parameterized.

\section*{2 Methods}

MOF-FF was developed to accurately simulate soft porous crystals by classical molecular dynamics simulations. It consists of two parts: first, a modified total energy expression based on the well known MM3 force field.\textsuperscript{53,54} This energy expression is implemented in our in-house developed \textsc{Pydlpoly} molecular mechanics suite\textsuperscript{45} and quite recently as user package in the well known \textsc{Lammps} molecular dynamics simulator.\textsuperscript{55} Second, a QM based parameterization methodology for the intramolecular interactions which is applicable to all kinds of classical force fields and molecules/materials like other QM based parameterization strategies.\textsuperscript{51,56} The methodology is implemented in our \texttt{FFgen} code. For this study we have revisited the parameterization methodology and let the force field expression untouched (see Ref. 45 or the supporting information for details on that part).
### 2.1 Fitting to Periodic Reference Data

The basic idea behind the parameterization procedure is visualized in Fig. 1. The reference information \( i.e. \) DFT optimized structure and the curvature information represented in the matrix of the second derivatives of the energy in respect to the coordinates \( \vec{\nabla}^2 E \) called Hessian \( H \) are calculated for a given reference system. The FFgen code is then used to find the best matching parameters in respect to the provided reference information for a given set of vdW potentials and charges. Whereas charges were calculated by a fit to the electrostatic potential of the optimized reference structure, vdW parameters were taken from the well known MM3 force field.\(^{53,54}\)

![Figure 1: General scheme of the MOF-FF parameterization methodology.](image)

The first innovation affects the input required by FFgen: For a lot of systems it is a sufficient approximation to represent the periodic structure by a zero dimensional cluster like for the IRMOF series or paddle-wheel based systems. But there are systems where this is not easily possible. Prominent representatives for these systems are, for example, MOF-74 or MIL-53. In these materials the inorganic building block is not an isolated metal oxide center, instead it is a one dimensional metal oxide chain. The case of ZIFs is somewhat different: while they are not composed by one dimensional metal oxide chains, it is impossible to cut a zero-dimensional cluster of charge zero. Especially for MOF-FF, where no 1–2 and 1–3 exclusions are used, this is a rather complicated problem, because a certain amount of
Coulombic interactions is therefore present in all bond and angle interactions. This difficulty is made irrelevant by the use of three dimensional periodic reference data. We have thus extended MOF-FF to use reference data that is periodic in one, two or three dimensions. Technically, the main implication is to include periodic boundary conditions in the calculation of Wilson’s $B$-matrix, which is explained in the next subsection.

### 2.2 Improvement of the Objective Function

The second improvement concerns the design of the objective function and its hyperparameters. For a better understanding of those the actual parameterization process is shortly reviewed.

In a classical FF the intramolecular energy is computed as the sum of parameterized many-body potentials. These many-body potentials are calculated in respect to internal coordinates like bonds $r$, angles $\phi$, dihedrals $\tau$ and improper dihedrals $\gamma$. These internal coordinates comprise more than the usual $3N - 6$ coordinates (where $N$ is the number of atoms) and are therefore labeled as redundant internal coordinates (RICs). It is a natural choice to use these RICs also during the parameterization process.

*À priori* to the evaluation of the objective function $Z_{MOF-FF}$, the geometry of the reference system is optimized with the current parameter set $P$ comprised of the individual parameters $p_i$, then the Hessian is calculated by a double sided finite difference approach with a stepsize of 0.001 Å. In the next step the RICs are computed and the Hessian $H$ is projected from Cartesian coordinates to RICs ($H_q$) by the help of Wilson’s $B$-matrix which is defined as:

$$B_{ij} = \sum_{ij} \frac{\partial q_i}{\partial x_j}, \quad (1)$$

where the $q_i$ are the internal coordinates and the $x_j$ are the Cartesian ones. The $B$-matrix is a rectangular matrix where the numbers of rows is equal to the number of internal coordinates $Q$ and the number of columns is equal to the number of Cartesian coordinates $3N$. The
main change necessary to be able to handle periodic reference data was to include periodic boundary conditions in the calculation of the $B$-matrix. Note that the RICs used for the calculation of the intramolecular energy are chosen here as well. A detailed description on how the projection of the Hessian $H$ into the RIC space is performed is shown in the supporting information.

After projecting both geometry and Hessian of the FF model (computed with trial parameter set $P$) and of the reference calculations, the objective function $Z_{MOF-FF}$ is determined by the difference between them in RIC space:

$$Z_{MOF-FF}(P) = \frac{w_{str}}{M_{str}} \sum_{i=1}^{N_{str}} \omega_i (r_i(P) - r_i^{ref})^2 + \frac{w_{ibe}}{M_{ibe}} \sum_{j=1}^{N_{ibe}} \omega_j (\theta_j(P) - \theta_j^{ref})^2 + \frac{w_{obe}}{M_{obe}} \sum_{k=1}^{N_{obe}} \omega_k (\gamma_k(P) - \gamma_k^{ref})^2 + \frac{w_{tor}}{M_{tor}} \sum_{l=1}^{N_{tor}} \omega_l (\tau_l(P) - \tau_l^{ref})^2 + \frac{w_{wdiag}}{M_{ric}} \sum_{m=1}^{N_{ric}} \omega_{mm} (H_{q,mm}(P) - H_{q,mm}^{ref})^2 + \frac{w_{wodiag}}{M_{ric}} \sum_{n=1}^{N_{ric}} \sum_{o \neq n} \omega_{no} (H_{q,no}(P) - H_{q,no}^{ref})^2$$

The individual weights $\omega$ per redundant internal coordinate $q$ are assigned based on the atomtypes in the system. If a RIC $q$, defined by its atomtypes, occurs $n_q$ times in the system, it gets a weight of $\omega_q = n_q^{-1}$. The condition, therefore, is such that the parameters belonging to the FF term describing the RIC of interest are in the set of variable parameters $P$. Furthermore, since $M_{ric} = \sum_{i=1}^{N_{ric}} \omega_i$, all single contributions to $Z_{MOF-FF}$ are weighted in the same manner. Of course this is somehow arbitrary, since different RIC types have different units. Bond lengths are compared in Å, angles in rad and force constants in mdyn/Å (bonds) or mdynÅ/rad² (angles). With this choice all contributions to $Z_{MOF-FF}$ are roughly in the same order of magnitude. The weights $w_{str}$, $w_{ibe}$, $w_{obe}$, $w_{wdiag}$ and $w_{wodiag}$ can be tuned by the user to increase the importance of a specific contribution. Note that the off-diagonal elements of the Hessian are only included in the objective function if also the corresponding cross terms are fitted.

The advantage of this reformulated objective function is a proper weighting, since in the
original formulation the RICs were weighted all equally with the consequence that RICs get a higher weight in the overall objective function when they occur more often in the system.

For the fitting of periodic systems we introduced an additional term to ensure that the force field also describes the lattice dimensions in the correct way. This additional term depends on the stress tensor $S$ and penalizes parameter sets $P$ which result in large stress tensor for the optimized geometry:

$$Z_{\text{lattice}} = \left( \frac{1}{9} \sum (C^{-1} \cdot (S \cdot V))^2 \right)^2$$

where $V$ is the cell volume and $C$ is the cell tensor. This additional term was then weighted by a factor of 0.1 and added to $Z_{\text{MOF-FF}}$ in this study. Another option to include the cell shape in the objective function would be to run, in addition to the geometry optimization, an additional lattice optimization; however we did not follow this approach due to its higher computational cost.

### 2.3 A new Optimizer: CMA-ES

Optimization of the objective function, with respect to the force field parameters, is complicated because the search landscape is relatively bumpy with a lot of local minima. For this purpose we rely on a stochastic zero-order optimizer which is able to escape from a local minimum. Such optimizers are also often referred to as black-box optimizers because they do not need any further information besides the actual value of the objective function at a given search point. In the original MOF-FF parameterization the PIKAIA optimizer was used, which is a genetic algorithm specifically designed for use on continuous variables. Generally, genetic algorithms are employed for optimization on discrete variables, thus in order to represent the continuous variables in a genome of discrete numbers, one has to predefined for every parameter $p_i$ a range consisting of an upper $p_{i}^{\text{max}}$ and a lower bound $p_{i}^{\text{min}}$ so that it holds $p_{i}^{\text{min}} \leq p_i \leq p_{i}^{\text{max}}$. One limitation of this approach is thus that it is necessary to define
these ranges *a priori*. If the solution is outside of the range it must be readjusted manually.

Furthermore, the periodic reference systems used in this study are substantially larger than the cluster models employed before. The primitive cell of ZIF-8 consists of 138 atoms and is more than a factor of two larger than e.g. a benzoate paddle-wheel unit. We thus needed to change the optimizer to reduce the computational cost of the parameterization process. We implemented in our FFgen code an algorithm called Covariance Matrix Adaption Evolution Strategy (CMA-ES), which was developed by Hansen for optimization problems on continuous variables.\textsuperscript{58,59} CMA-ES has previously been applied for force field development, namely for deriving coarse grained FFs for MOFs\textsuperscript{48} and recently for the parameterization of ReaxFF.\textsuperscript{60}

In a nutshell, CMA-ES iteratively improves a multivariate normal distribution in the parameter space to find a distribution whose random samples minimize the objective function. Thus, the individuals of a generation are sampled from a multivariate normal distribution where recombination is done by selecting a new mean for the distribution. During the optimization process, the covariance matrix is updated in order to optimize the shape of the multivariate normal distribution with respect to the search landscape defined by the objective function. This amounts to learning a second order model of the underlying objective function similar to the approximation of the inverse Hessian in quasi-Newton optimization methods, which are known for their quadratic convergence near the optimum. A full description of the optimizer can be found in the supporting information.

Using CMA-ES instead of PIKAIA leads to a substantially faster convergence. To validate this claim we fitted a force field for ZIF-8(Br), using the same setup as described in Sec. 4.1.3, with both optimizers. The convergence behavior is visualized in a logarithmic scale in Fig. 2. One evaluation of the fitness functions takes around 2 s on a single core of a current desktop GNU/LINUX workstation.

A further advantage of CMA-ES is that it is not necessary to define *a priori* any ranges for the parameters as needed in PIKAIA. But defining ranges can be very helpful for the following
Figure 2: Logarithmic convergence behavior of the originally employed PIKAIA genetic algorithm vs. CMA-ES for a parameterization run of ZIF-8(Br).

reasons: at the startup of the algorithm an isotropic normal distribution with a predefined stepsize $\sigma$ is initialized. Since the parameters differ by their units, they have different orders of magnitude. For this reason we define every parameter in reduced parameter units by the help of an upper and a lower bound:

$$\bar{p}_i = \frac{p_i}{p_{i}^{\text{max}} - p_{i}^{\text{min}}}$$ \hspace{1cm} (4)

However, in contrast to PIKAIA the ranges are in general not hard constraints, so the parameter is allowed to escape them. Hard ranges are only applied to prevent unphysical parameters like negative bond lengths, bond angles or force constants.

The optimizer is started from an educated guess for the parameters and their appropriate ranges. For bond lengths $r$ and bond angles $\theta$ it is started from the geometrically measured values of the reference system. Upper and lower bound were by default set to $\pm 10\%$ from the geometrically measured value. For bond and angle force constants the ranges are defined as $0 \text{mdyn}\text{Å}^{-1} \leq p_i \leq 8 \text{mdyn}\text{Å}^{-1}$ and $0 \text{mdynÅ}/\text{rad}^2 \leq p_i \leq 2 \text{mdynÅ}/\text{rad}^2$. For dihedral potentials $0 \text{kcal}/\text{mole} \leq p_i \leq 20 \text{kcal}/\text{mole}$ was used as range for the barrier and for out-of-
plane potentials $0 \text{mdynÅ}/\text{rad}^2 \leq p_i \leq 1 \text{mdynÅ}/\text{rad}^2$ was the default.

2.4 Technical Implementation

Originally, the FFgen code was a monolithic FORTRAN code using TINKER version 4.2\textsuperscript{61} as molecular mechanics back-end. We restructured the whole code in an object-oriented way using the Python programming language. Much effort was spent on isolating and modularizing the four main parts of the code to make it easy to implement new objective functions, molecular mechanics backends, optimizers or potentials.

Currently it is possible to use two different backends, namely our in-house developed PYDLPOLY code or LAMMPS package. The usage of LAMMPS for parameterization in respect to periodic reference data is recommended since LAMMPS allows the usage of smaller unit cells \textit{i.e.} primitive cells as reference systems. Furthermore we parallelized the CMA-ES optimizer to speed up the whole process by distributing the evaluation of the fitness function over parallel processes.

3 Computational Details

3.1 Obtaining the Reference Information

The reference information needed for FF optimization was obtained with periodic DFT calculations using the QUICKSTEP/CP2K package.\textsuperscript{62} QUICKSTEP/CP2K is based on a hybrid Gaussian plane-wave approach combining a Gaussian basis for the wave-functions with an auxiliary plane wave basis set for the representation of the density.

We found that it is crucial to use a high plane-wave cutoff $E_{\text{cut}}$ to obtain accurate Hessians, since in the QUICKSTEP module of CP2K, the computation of Coulomb and exchange-correlation energies is performed on a real space grid. This representation breaks the translational invariance of the system, which can lead to spurious forces on the atoms (egg box effect), which can have a large influence on the Hessian calculated by a finite difference
approach based on the atomic forces.\textsuperscript{63}

The gradient-corrected PBE functional\textsuperscript{64} was used with empirical correction for the dispersive interactions using the “D3” method by Grimme et al. \textsuperscript{65} Double-$\zeta$ valence polarized Gaussian basis sets were employed for all atoms. For C, H, and N basis sets optimized for usage with the PBE functional were employed, whereas on Zn, Cl and Br basis sets, optimized for molecules (MOLOPT) were employed. The interaction between ions and valence electrons was represented by Goedecker-Teter-Hutter (GTH) type pseudo-potentials.\textsuperscript{66–68} Given the systems’ size, the Brillouin zone was sampled at the $\Gamma$ point only.

In order to obtain accurate reference data, strict convergence criteria had to be chosen both for the SCF and for the geometry optimizations. For the SCF a convergence criterion $10^{-10}$ Hartree was applied, whereas for the optimizations the RMS force has to be lower than $10^{-7}$ Hartree/bohr. Wherever possible, primitive cells were used for all DFT calculations. The reference data generation can be divided into four steps:

1. Atomic coordinates and the lattice dimensions were optimized using a plane-wave cutoff for the density $E_{\text{cut}}$ of 600 Ry (systems without Cl or Br) \textit{i.e.} 700 Ry (systems with Cl or Br) together with a relative cutoff $E_{\text{cut}}^{\text{rel}}$ of 40 Ry.

2. Afterwards only the atomic coordinates were re-optimized using a cutoff $E_{\text{cut}}$ of 2500 Ry together with a relative cutoff of $E_{\text{cut}}^{\text{rel}}$ of 100 Ry. These cutoffs were used for all subsequent calculations.

3. The Hessian of the optimized structure was calculated by the help of a double sided finite difference scheme using a distortion of at least 0.001 bohr.

4. Charges were calculated by the REPEAT method\textsuperscript{69} using its implementation in \textsc{CP2K}. For this purpose two type of constraints were employed: The total charge of the system has to be zero and atoms with equal atom-types get same charges.
3.2 *Ab initio* Molecular Dynamics Simulations

We performed *ab initio* molecular dynamics simulations (AIMD) of the ZIF-8 isomorphs with Born-Oppenheimer dynamics using DFT for the calculation of the energy and the atomic forces using the QUICKSTEP/CP2K package. We used the same functional, dispersion correction and basis sets as for the static calculations described in Sec. 3.1. The cutoff for the density plane-wave basis set was set to 600 Ry.

We performed all AIMD simulations using periodic boundary conditions on a single crystallographic unit cell of the material ($a_{\text{ZIF-8}} = 16.991$ Å, $a_{\text{ZIF-8(Br)}} = 16.985$ Å, $a_{\text{ZIF-8(Cl)}} = 16.998$ Å). We used deuterated hydrogen atoms to allow for a larger time step (1 fs) in the integration of the equations of motion. Simulations were run in the canonical ensemble ($N, V, T$) using a CSVR\textsuperscript{70} thermostat with a time constant of 1 ps to control the temperature. The total simulation time was between 30 ps and 40 ps depending on the isomorph.

For comparison we used also data based on BLYP\textsuperscript{71,72} AIMD simulations published in a previous study. For further details on these simulations it is referred to the original publication.\textsuperscript{18}

3.3 Classical Molecular Dynamics Simulations

Classical molecular dynamics simulations were performed using our in-house developed PYDLPOLY code on $2 \times 2 \times 2$ crystallographic unit cells of the material, employing the same lattice constants as used for the AIMD simulations. Simulations were run in the canonical ensemble ($N, V, T$) using a Berendsen thermostat for equilibration runs of 0.1 ns and a Nosé-Hoover thermostat for sampling runs of 1 ns. Thermostat relaxation times were set to 0.2 ps and 2.0 ps.
3.4 Calculation of Elastic Constants

Elastic constants $C_{ij}$, which are the coefficients of the second-order elastic tensor $C$, were computed by using the numerical first derivative of the cell gradients corresponding to the each elastic coefficient $C_{ij}$. For this purpose the optimized structure was deformed in each possible direction, applying both negative and positive strain (corresponding to compression and tension), and for each deformation the atomic coordinates were energy-minimized. For deformations along the normal coordinates strains of $-1.0\%$ to $1.0\%$ in steps of $0.5\%$ were applied, whereas for shear deformations, strains from $-4.0\%$ to $4.0\%$ in steps of $2.0\%$. The PYMATGEN package was used to deform the structures and to perform the final analysis, using the crystallographic unit cell as a reference configuration. Elastic constants at the ab initio level of theory were computed using the QUICKSTEP/CP2K package employing the same setup as used for the production of the reference data. Elastic constants at the FF level were computed by using PYDLPOLY and a $2 \times 2 \times 2$ supercell.

4 Results and Discussion

4.1 FF Parameterization

4.1.1 Systems of Interest

We focused in this study on the development of four different FFs for ZIF-8, ZIF-8(H), ZIF-8(Br) and ZIF-8(Cl). All are composed of zinc and differently functionalized imidazolate linkers, and form 3D crystalline networks with the sod topology (chemical composition and atom types shown in Fig. 3).

Furthermore we calculated also the reference data for six other polymorphs with the ZIF-8(H) composition, which we used to probe into the transferability of the FFs between different topologies. Crystallographic and chemical information about the investigated systems are listed altogether in Table 1.
Figure 3: Chemical systems investigated in this study, showing the atom typing used throughout the text.

Table 1: Summary of the systems investigated in this study.

<table>
<thead>
<tr>
<th></th>
<th>Space group</th>
<th>Bravais lattice</th>
<th>Net topology</th>
<th>Functional group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIF-1</td>
<td>$P2_1/c$</td>
<td>mono</td>
<td>crb</td>
<td>H</td>
</tr>
<tr>
<td>ZIF-4</td>
<td>$Pbca$</td>
<td>ortho</td>
<td>cag</td>
<td>H</td>
</tr>
<tr>
<td>ZIF-6</td>
<td>$I4_1/amd$</td>
<td>tetra</td>
<td>gls</td>
<td>H</td>
</tr>
<tr>
<td>ZIF-8(H)</td>
<td>$I43m$</td>
<td>cubic</td>
<td>sod</td>
<td>H</td>
</tr>
<tr>
<td>coi</td>
<td>$I4_1$</td>
<td>tetra</td>
<td>coi</td>
<td>H</td>
</tr>
<tr>
<td>nog</td>
<td>$P2_1/c$</td>
<td>mono</td>
<td>nog</td>
<td>H</td>
</tr>
<tr>
<td>zni</td>
<td>$I4_1cd$</td>
<td>tetra</td>
<td>zni</td>
<td>H</td>
</tr>
<tr>
<td>ZIF-8(Br)</td>
<td>$I43m$</td>
<td>cubic</td>
<td>sod</td>
<td>Br</td>
</tr>
<tr>
<td>ZIF-8(Cl)</td>
<td>$I43m$</td>
<td>cubic</td>
<td>sod</td>
<td>Cl</td>
</tr>
<tr>
<td>ZIF-8</td>
<td>$I43m$</td>
<td>cubic</td>
<td>sod</td>
<td>Met</td>
</tr>
</tbody>
</table>
4.1.2 Partial Charges

The atomic partial charges obtained for ZIF-8, listed in Table 2, are in good agreement with those published by Rana et al.\textsuperscript{35} It should be noted that these authors computed the charges as average over several snapshots obtained from \textit{ab initio} MD simulations in the Born-Oppenheimer scheme and in the Car-Parrinello scheme. Obviously, charges obtained from a single energy-minimized structure are already sufficient.

For the seven hydrogen substituted ZIF polymorphs the charges are shown in Table S1. Depending on the topology the charges on the Zn range from 0.4645 e to 0.5971 e with a standard deviation of 0.04 e, indicating that the impact of the topology on the charges is relatively minor, compared to the natural methodology-related uncertainty. We thus chose as charges used for our FF a unique set of partial charges for this class of materials, obtained by averaging over the seven structures.

In addition the charges for the halogen-substituted structures are listed in Table S2. The charges in the Zn atoms for the four chemically distinct species in the sod topology differ almost by a factor of two (ZIF-8: 0.7290 e, ZIF-8(H): 0.5118 e, ZIF-8(Cl): 0.4986 e, ZIF-8(Br): 0.4714 e). This demonstrates the huge impact of the substituent on the whole framework and emphasizes the need for distinct parameter sets for every chemically distinct species, especially in the case of the non-polarizable MOF-FF where no Coulomb exclusions are applied.

Table 2: REPEAT charges (units in partial electron charges) for ZIF-8 computed by us in respect to the optimized structure in comparison to those published by Rana et al. based on snapshots extracted from BOMD and CPMD simulations.\textsuperscript{35}

<table>
<thead>
<tr>
<th></th>
<th>Zn</th>
<th>N</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>H1</th>
<th>H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>0.7290</td>
<td>−0.3417</td>
<td>−0.1923</td>
<td>0.4937</td>
<td>−0.5276</td>
<td>0.1627</td>
<td>0.1373</td>
</tr>
<tr>
<td>BOMD</td>
<td>0.7362</td>
<td>−0.3008</td>
<td>−0.1924</td>
<td>0.4339</td>
<td>−0.6042</td>
<td>0.1585</td>
<td>0.1572</td>
</tr>
<tr>
<td>CPMD</td>
<td>0.6894</td>
<td>−0.2800</td>
<td>−0.1910</td>
<td>0.4184</td>
<td>−0.5726</td>
<td>0.1536</td>
<td>0.1481</td>
</tr>
</tbody>
</table>
4.1.3 Intramolecular Parameterization

The force field energy expression was set up for the four systems in the way that only diagonal terms were applied and no cross terms like stretch-bend potentials were used. All dihedral potentials were set up by imposing a multiplicity of two, besides the C2–N–Zn–N and C1–N–Zn–N where a multiplicity of three was used. Due to the disorder of the methyl groups, no dihedral potential was applied for N–C2–C3–H3 dihedral in ZIF-8(H) and the involved parameters were predefined for the actual parameterization. The comparison of the actual dihedral angle values was not included in the fitness (except in the case of ZIF-8, where \( w_{\text{tor}} = 0.1 \) was employed), since the geometry is already imposed by the chosen multiplicity of the potential. Charges were chosen as described above and MM3 van der Waals parameters were employed. Equal weights of one were assigned to all different contributions (\( w_{\text{str}}, w_{\text{ibe}}, w_{\text{wdiag}} \)) of the objective function \( Z \).

When using heuristic optimizers like CMA-ES on bumpy search landscapes, it is recommended to run the optimizer several times with increasing population sizes.\(^7\) For this reason we started always several runs increasing the populations size \( \lambda \) from the default size (\( \lambda = 4 + (3 \ln N_{\text{par}}) \)) up to 32 individuals per generation. We then selected the FF with the lowest objective function value. A complete list of the optimized parameters is reported in the Supporting information.

4.2 Force Field Validation

4.2.1 Structural Properties

As a first validation step, we verified how well the parameterized FFs reproduce the structural properties of the materials, with respect to DFT-optimized structures. Fig. 4 shows a comparison between the final bond lengths, bending angles and dihedral angles obtained at the two levels of theory. Overall agreement is excellent, with the biggest deviations observed for ZIF-8, due to the disorder introduced by the methyl groups’ free rotation and the fact
that no dihedral potential is included for the N–C2–C3–H3 torsion. We also investigated
the lattice dimensions at zero Kelvin (full cell energy optimization) in comparison to the
DFT results; see Table 3. The lattice dimensions of the structures which served as reference
systems for the parameterization coincide almost perfectly. The biggest difference of 0.06 Å
arises for ZIF-8. But it should be noted that already the DFT calculations overestimates
the lattice dimension by 0.06 Å compared to the experimental lattice constants at 258 K.

Figure 4: Scatter plots visualizing the performance of our FFs with respect to the \textit{ab initio}
reference data. (a) Comparison of bond lengths (b) Comparison of bond angles (c) Comparison for dihedral angles (d) Comparison of vibrational normal modes.
Table 3: Lattice dimensions for the investigated ZIF systems computed by the FFs and different DFT methods. DFT results from the literature are listed only in extracts.

<table>
<thead>
<tr>
<th></th>
<th>a[Å]</th>
<th>b[Å]</th>
<th>c[Å]</th>
<th>α[°]</th>
<th>β[°]</th>
<th>γ[°]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIF-1 cp2k</td>
<td>9.94</td>
<td>14.93</td>
<td>16.45</td>
<td>90.00</td>
<td>118.51</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-1 FF</td>
<td>10.09</td>
<td>14.55</td>
<td>15.91</td>
<td>90.00</td>
<td>117.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-4 cp2k</td>
<td>14.73</td>
<td>18.30</td>
<td>15.26</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-4 FF</td>
<td>15.58</td>
<td>18.54</td>
<td>15.84</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-4 exp⁵</td>
<td>15.40</td>
<td>18.43</td>
<td>15.31</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-4 (H) cp2k</td>
<td>14.97</td>
<td>17.67</td>
<td>14.16</td>
<td>90.00</td>
<td>90.00</td>
<td>97.61</td>
</tr>
<tr>
<td>ZIF-6 cp2k</td>
<td>19.37</td>
<td>19.37</td>
<td>19.60</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-6 exp⁵</td>
<td>18.52</td>
<td>18.52</td>
<td>20.25</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-6 FF</td>
<td>19.23</td>
<td>19.23</td>
<td>19.80</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(H) cp2k</td>
<td>16.97</td>
<td>16.97</td>
<td>16.97</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(H) FF</td>
<td>16.97</td>
<td>16.97</td>
<td>16.97</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>coi cp2k</td>
<td>17.22</td>
<td>17.22</td>
<td>17.22</td>
<td>98.58</td>
<td>98.58</td>
<td>134.83</td>
</tr>
<tr>
<td>coi FF</td>
<td>17.23</td>
<td>17.23</td>
<td>17.22</td>
<td>98.63</td>
<td>98.63</td>
<td>134.34</td>
</tr>
<tr>
<td>nog cp2k</td>
<td>24.47</td>
<td>9.60</td>
<td>34.58</td>
<td>90.00</td>
<td>132.19</td>
<td>90.00</td>
</tr>
<tr>
<td>nog FF</td>
<td>24.68</td>
<td>9.64</td>
<td>35.03</td>
<td>90.00</td>
<td>134.67</td>
<td>90.00</td>
</tr>
<tr>
<td>zni cp2k</td>
<td>23.35</td>
<td>23.35</td>
<td>12.56</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>zni FF</td>
<td>23.76</td>
<td>23.76</td>
<td>12.50</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>zni exp⁷⁶</td>
<td>23.50</td>
<td>23.50</td>
<td>12.46</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>zni FF</td>
<td>23.23</td>
<td>23.23</td>
<td>12.79</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8 cp2k</td>
<td>17.03</td>
<td>17.03</td>
<td>17.03</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8 FF</td>
<td>17.09</td>
<td>17.09</td>
<td>17.09</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(Br) cp2k</td>
<td>17.25</td>
<td>17.25</td>
<td>17.25</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(Br) FF</td>
<td>17.25</td>
<td>17.25</td>
<td>17.25</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(Cl) cp2k</td>
<td>17.20</td>
<td>17.20</td>
<td>17.20</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>ZIF-8(Cl) FF</td>
<td>23.21</td>
<td>23.21</td>
<td>12.84</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
</tbody>
</table>
4.2.2 Deformations: Vibrations and Elasticity

Next we checked the accuracy of the FFs in representing deformations from the relaxed structures, comparing normal mode frequencies and elastic constants, which are directly linked to the second derivative of the energy in respect to the atomic coordinates, and unit cell parameters, respectively.

Fig. 4d shows the comparison of the FF normal modes against DFT data. The agreement is good, in particular for the low-frequency vibrations modes. Those modes are maximally delocalized and are mainly responsible for the lattice vibrations and flexibility of the framework, and consequently crucial to reproduce correctly. The modes between 500 cm$^{-1}$ and 2000 cm$^{-1}$ are more localized and involve especially distortions of the aromatic imidazolate rings. For a purely diagonal FF the accuracy obtained in this region has some limitations, which could only be lifted with the use of cross-terms. Though our FF derivation methodology is fully capable for the parameterization of cross terms, we deliberately dispensed them in this study in order to create a widely-usable FF, which can be input in a lot of molecular mechanics suites and does not depend on nonstandard terms.

Elastic constants calculated by different DFT methods — some published in the literature, and some computed as part of this work — are compared to those obtained by our FFs in Table 4. Zheng et al. predicted recently the elastic constants of differently functionalized ZIFs in the sod topology. They found that electron withdrawing groups improve the mechanical stability of the materials (ZIF-8(H) < ZIF-8(Cl) < ZIF-8(Br)).$^{77}$ Although the absolute numbers of our DFT calculations differ up to a few GPa this trend is also reproduced by our calculations. Furthermore this trend is even reproduced in our force field calculations. Only exception is $C_{44}$, but already the differences from the reference calculations are here very subtle in comparison to those published in the literature. Interestingly the elastic constants computed with the FF are closer to the numbers computed at the reference level of theory than to those computed by Zheng et al.
Table 4: Elastic constants of ZIF-8, ZIF-8(H), ZIF-8(Br) and ZIF-8(Cl) computed by the FFs and different DFT methods. DFT results from the literature are listed only in extracts.

<table>
<thead>
<tr>
<th></th>
<th>$C_{11}$ [GPa]</th>
<th>$C_{12}$ [GPa]</th>
<th>$C_{44}$ [GPa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIF-8 FF</td>
<td>10.03</td>
<td>6.21</td>
<td>1.31</td>
</tr>
<tr>
<td>ZIF-8(H) FF</td>
<td>6.65</td>
<td>4.95</td>
<td>1.12</td>
</tr>
<tr>
<td>ZIF-8(H) cp2k</td>
<td>5.39</td>
<td>4.51</td>
<td>0.32</td>
</tr>
<tr>
<td>ZIF-8(H)</td>
<td>8.95</td>
<td>7.59</td>
<td>2.36</td>
</tr>
<tr>
<td>ZIF-8(Cl) FF</td>
<td>9.92</td>
<td>7.84</td>
<td>0.46</td>
</tr>
<tr>
<td>ZIF-8(Cl) cp2k</td>
<td>9.23</td>
<td>7.35</td>
<td>0.86</td>
</tr>
<tr>
<td>ZIF-8(Cl)</td>
<td>12.30</td>
<td>9.98</td>
<td>3.58</td>
</tr>
<tr>
<td>ZIF-8(Br) FF</td>
<td>10.51</td>
<td>8.65</td>
<td>0.19</td>
</tr>
<tr>
<td>ZIF-8(Br) cp2k</td>
<td>10.33</td>
<td>8.31</td>
<td>0.88</td>
</tr>
<tr>
<td>ZIF-8(Br)</td>
<td>15.92</td>
<td>11.57</td>
<td>6.56</td>
</tr>
</tbody>
</table>

4.2.3 Transferability

In the next step we investigated the transferability of the FF optimized above. How does it perform when applied to topologies for which it has not been parameterized for? We thus optimized the structure of six ZIF-8(H) polymorphs and compared the energy to DFT results, obtained in this work as well as those of Baburin et al. Is it possible to reproduce the relative stabilities of the isomorphs and the correct energetic ordering? The results are shown in Fig. 5a. The errors of the force field with respect to our DFT results are in the same order of magnitude as the difference of our DFT results compared to those by Baburin et al., also computed at the DFT level but using a different methodology. DFT predicts the following energetic ordering: $\text{coi} < \text{zni} < \text{nog} < \text{ZIF-4} < \text{ZIF-1} < \text{ZIF-8(H)} < \text{ZIF-6}$, whereas the FF predicts $\text{coi} < \text{zni} < \text{ZIF-1} < \text{nog} < \text{ZIF-4} < \text{ZIF-8(H)} < \text{ZIF-6}$. Thus, the only difference is that the stability of ZIF-1 is overestimated. The reason for this can be found in the lower symmetry of the other polymorphs compared to ZIF-8(H) which was used as reference system. Fig. 5b shows the distribution of the $\text{N} - \text{Zn} - \text{N}$ angles for the other investigated polymorphs. They show substantially larger deviations from the ideal tetrahedron angle as compared to ZIF-8(H). So they are further away from the reference
structure which results in a worse performance. The same behavior can also be observed
for the lattice constants of the other polymorphs as shown in Table 3. In comparison to the
systems which served as reference information they show an inferior performance, however
they are still well within the range of lattice constants predicted by different DFT methods.
In the absence of experimental data on most of these systems, it is hard to go any further.
We note, however, that the predictive power in respect to the other polymorphs could be
increased by fitting a FF for each of them, but by this one would loose the possibility to
compare the relative stabilities between different polymorphs.

Figure 5: (a) Scatter plot of the total energies of the 7 ZIF-8(H) polymorphs computed by
the FF and two different DFT methods, namely by Baburin et al. and us as described in
Sec. 3.1. (b) Analysis of the N-Zn-N angles in the seven ZIF-8(H) polymorphs. Error bars
represent the standard deviation of all N-Zn-N angles in the unit cell.

4.2.4 Molecular Mechanism of the Swing Effect

The flexibility in ZIF-8 analogues is governed by intraframework dynamics and involves, as
detailed in the introduction, the so-called swing effect of the imidazolate linkers that allows
molecules larger than its window size to diffuse into the framework. Coudert has investigated
this effect in detail by ab initio MD simulations. The swinging motion of the imidazolate
linkers was characterized by the dihedral angle $\phi$ \text{Zn}_3 – Zn_2 – Zn_1 – CH$_3$ of the imidazolate
around the Zn$_1$ – Zn$_2$ axis, where the “reference” of 0° is the 6-ring of Zn (i.e. the window connection the cages) as shown in Fig. S1.

We used our force fields to run classical MD simulations in the (N, V, T) ensemble using the same lattice constants as in the corresponding AIMD for the different functionalizations to compare with the DFT results. Originally only ZIF-8 and ZIF-8(H) has been investigated using the BLYP functional.\textsuperscript{71,72} We have performed further AIMD simulations, in the course of the present work, for ZIF-8, ZIF-8(Br) and ZIF-8(Cl) using the PBE functional.

Histograms of the swinging angle $\phi$ for the four different systems are shown in Fig. 6. For ZIF-8, data from both PBE and BLYP is available (see Fig. 6a). From the difference between them one can estimate the errors made by different DFT exchange–correlation functionals which is an important measure for the accuracy of FFs. As one can see the differences are quite substantial, the histogram calculated with PBE has a larger spread, the “thermal” swing motion goes up to 20°, whereas in the case of BLYP it is only 15°. The histogram predicted by MOF-FF lies in between the two curves predicted by the two functionals.

For ZIF-8(H) only data computed with the BLYP functional is available (see Fig. 6c). As already predicted by Coudert ZIF-8(H) exhibits a swing motion which much larger amplitude than ZIF-8. This effect is nicely resembled by our force field. The thermal swing motion goes all the way up to 35° (instead of 20° for ZIF-8).

Additional AIMD simulations were performed for the halogenated species using the PBE functional. The agreement between FF and DFT calculated data is slightly less good than for ZIF-8 and ZIF-8(H). For ZIF-8(Br) the general shape of the curve is reproduced well but it is stretched along the x-axis towards higher swing angles by 5° (see Fig. 6b). In case of the ZIF-8(Cl) the FF predicted curve is compressed to lower swing angles by 5° degrees (see Fig. 6d). We emphasize, however, that these differences are of the same order of magnitude as the difference between the curves for ZIF-8 predicted at the DFT level of theory by the different exchange–correlation functionals.
Figure 6: Histograms of the swing angle of the imidazolate linkers, for ZIF-8 (a), ZIF-8(Br) (b), ZIF-8(H) (c) and ZIF-8(Cl) (d) computed at different levels of theory.
5 Conclusions

The aim of this study was to exploit the systematic and consistent MOF-FF force field parameterization methodology to arrive at an accurate and efficient potential for a range of ZIFs. By trading transferability for accuracy we parameterized explicitly for the chemically distinct systems, namely ZIF-8, ZIF-8(H), ZIF-8(Br) and ZIF-8(Cl), using the sod topology as a reference. For this purpose we needed to improve and extend the original methodology in several aspects. First, we added the possibility to employ also periodic reference systems, which enables us to use our approach also for those systems which can not be easily truncated to a cluster representations, as for example rod based MOFs like MIL-53 or MOF-74, or ZIFs in general. Second, we replaced the original genetic algorithm based global optimizer by another more efficient evolutionary strategy that is better suited for continuous variables. The CMA-ES converges substantially faster and without constraining to parameter ranges, which is extremely beneficial since the numerical effort for the evaluation of the target function is numerically much more involved in case of a fit to periodic reference systems like in the case of ZIFs. Furthermore, it also paves the way towards a completely automated black box algorithm for FF parameterization.

As expected, the force fields are able to well reproduce structure and lattice parameters as well as dynamic properties like vibrational normal modes and elastic constants in comparison to available experimental results as well as the computed reference data at DFT level of theory. Note, that the achieved accuracy is also due to the distinct parameter set for each chemically different system. In order to validate the transferability of the parameter set also the energetic ranking of other ZIF topologies was tested. Interestingly, we find that the energetic ordering of the polymorphs ZIF-1, ZIF-4, ZIF-6, coi, nog and zni computed with the force field fitted to ZIF-8(H) is nicely reproduced as compared to periodic DFT calculations. On the other hand, structural properties like N-Zn-N angles show a much wider scatter for the other polymorphs. In principle, a further improvement could be achieved by fitting explicitly to a reference in the corresponding polymorph, however, at the expense of
losing the possibility to compare and energetically rank the polymorphs.

A much more subtle property of ZIFs is their inherent flexibility, namely the so called swing effect, which allows molecules larger than the geometric window size to diffuse into the framework by a slight rotation of the imidazolate linkers. The ability of a force field to reproduce this behavior in an accurate way is crucial for its use in simulating guest molecule adsorption or for example heat conduction. We find that our force field is able to reproduce this flexibility within the same range of accuracy that is obtained in \textit{ab initio} MD simulations using different exchange–correlation functionals.

This demonstrates the potential of the here introduced extended MOF-FF parameterization methodology. Based on a single periodic reference structure (geometry and curvature) it is possible to derive an accurate force field in a consistent and systematic way which can be used to substantially extend length and time scales in MD simulations, within the constraint of no bond breaking, in an accuracy close to periodic DFT.

\textbf{Acknowledgement}

J.P.D. is grateful for the financial support by the Fonds der Chemischen Industrie (FCI) and by the Research School Plus from the Ruhr-Universität Bochum. Further financial support from the Deutsche Forschungsgemeinschaft (DFG) is acknowledged (grants SCHM1389/8-1 and SCHM1389/10-1). Access to HPC platforms was provided by a GENCI grant (A0030807069). We thank Hendrik Heenen for implementing MOF-FF in \textsc{LAMMPS}.

\textbf{References}


(78) Tan, J.-C.; Civalleri, B.; Lin, C.-C.; Valenzano, L.; Galvelis, R.; Chen, P.-F.; Bennett, T. D.; Mellot-Draznieks, C.; Zicovich-Wilson, C. M.; Cheetham, A. K. Exception-
ally Low Shear Modulus in a Prototypical Imidazole-Based Metal-Organic Framework.


Graphical TOC Entry