QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from *Ab Initio* Input

Louis Vanduyfhuys,[a] Steven Vandenbrande,[a] Toon Verstraelen,[a] Rochus Schmid,[b] Michel Waroquier,[a] and Veronique Van Speybroeck*[a]

QuickFF is a software package to derive accurate force fields for isolated and complex molecular systems in a quick and easy manner. Apart from its general applicability, the program has been designed to generate force fields for metal-organic frameworks in an automated fashion. The force field parameters for the covalent interaction are derived from *ab initio* data. The mathematical expression of the covalent energy is kept simple to ensure robustness and to avoid fitting deficiencies as much as possible. The user needs to produce an equilibrium structure and a Hessian matrix for one or more building units. Afterward, a force field is generated for the system using a three-step method implemented in QuickFF. The first two steps of the methodology are designed to minimize correlations among the force field parameters. In the last step, the parameters are refined by imposing the force field parameters to reproduce the *ab initio* Hessian matrix in Cartesian coordinate space as accurate as possible. The method is applied on a set of 1000 organic molecules to show the easiness of the software protocol. To illustrate its application to metal-organic frameworks (MOFs), QuickFF is used to determine force fields for MIL-53(Al) and MOF-5. For both materials, accurate force fields were already generated in literature but they requested a lot of manual interventions. QuickFF is a tool that can easily be used by anyone with a basic knowledge of performing *ab initio* calculations. As a result, accurate force fields are generated with minimal effort.

DOI: 10.1002/jcc.23877

Introduction

Force fields have become a very powerful tool in molecular simulation nowadays. They are used in a very broad range of research fields to describe the intermolecular and intramolecular interactions of molecular systems. This is mainly because they allow to perform molecular simulations on a length and time scale unaccessible to *ab initio* calculations. The following examples show the versatility of their usage: the investigation of the protein–ligand structure for rational drug design,[1–4] the design of new materials for methane storage,[5] the validation of continuum models describing the van der Waals interface in nanopumps,[6] an atomistic study of the temperature influence on the reactivity of plasma species.[7] Specifically in the field of nanoporous materials and in particular metal-organic frameworks (MOFs), they are used intensively to study adsorption,[8–10] diffusion,[8,11–13] separation,[14] and breathing[15,16] processes.

The literature on force field development is very rich and it is not possible to give a complete overview here. There are polarizable versus nonpolarizable force fields, all-atom versus coarse-grained force fields, diagonal force fields versus force fields with cross terms. Another possible scheme to classify them can be the transferability, the range of systems to which they are applicable. On one hand, one has universal or general force fields, these are force fields that are applicable to a very wide range of systems, for example, organic molecules. Examples of such force fields are UFF,[17] GAFF,[18] and DREIDING.[19] Force fields-like OPLS,[20,21] AMBER,[22–24] CHARMM,[25–27] MM3,[28] and MM4[29] can also be labeled as transferable, but between a more specific set of systems such as proteins or organic molecules. Recently, various independent groups proposed automated procedures to derive force fields. The group of Barone developed JOYCE,[30,31] a program to derive all-atom and united-atom force fields for small to medium-sized molecules. The force fields were parameterized by minimizing a cost function that measures the error between force field energy, gradient and Hessian on the one hand, and *ab initio* energy, gradient and Hessian on the other hand. However, the user needs to determine weight factors for the different contributions in the cost function, which is a nontrivial task that has to be repeated for every molecule separately. Another approach was used by the group of Ayers, they proposed an

[a] *L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, M. Waroquier, V. Van Speybroeck*
Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium
E-mail: Veronique.VanSpeybroeck@Ugent.be

[b] *R. Schmid*
Lehrstul für Anorganische Chemie 2, Organometallics and Materials Chemistry, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44780 Bochum, Germany

Contract grant sponsor: Scientific Research Flanders (FWO); Contract grant sponsor: Research Board of Ghent University (BOF); Contract grant sponsor: BELSPO (in the frame of IAP/7/05); Contract grant sponsor: European Research Council under European Community's Seventh Framework Programme; Contract grant numbers: FP7(2007–2013) and 240483

automated procedure for parameterizing Amber-compatible force fields.[32] This procedure requires input from AMBER force fields and the force constants of the harmonic terms are derived by means of a term-by-term projection of the *ab initio* Hessian. Very recently, Huang and Roux proposed GAAMP,[33] General Automated Atomic Model Parameterization, a program to derive AMBER or CHARMM compatible force fields. The harmonic terms were mainly taken from GAFF or CGenFF[34] and they focused their attention to the derivation of AMBER or CHARMM compatible charges and reliable potentials for soft dihedrals. Mayne et al. developed the Force Field Toolkit as a VMD plug-in to automatically parametrize CHARMM compatible force fields for small molecules.[35] Also very recently, Stefan Grimme developed Quantum Mechanically derived Force Field,[36] a procedure to automatically derive force fields from quantum mechanical input. The methodology also includes a parameterization of the van der Waals interactions inspired by earlier work of Grimme regarding density functional theory (DFT)-D3.[37] The force field (FF) was shown to be accurate for organic molecules and transition metal complexes. However, the rest values of the covalent terms are systematically set to the *ab initio* equilibrium values. It remains to be tested whether this remains valid for MOFs, especially if a force field is used with no exclusions of 1–2 and 1–3 electrostatic interactions. Force fields without such exclusion rules can, however, proof necessary for describing highly polarized systems, such as the metal-oxide unit in MOFs.

The development of a new program package QuickFF for deriving force fields is inspired by the general quest from the MOF-community to develop in a transparent way accurate force fields for these systems. New MOFs are being synthesized at a considerable rate. Furthermore, large databases of hypothetical MOFs have been proposed[38,39] recently. A tool that is able to generate force fields with a minimum of manual interventions would be very valuable to screen a large number of materials in a fast and easy way. Once force fields are generated, they may be used in molecular dynamics (MD) and Monte-Carlo simulations to describe phenomena, with a fixed connectivity of the material, not reachable by quantum mechanical approaches due to the high computational cost. MOFs differ from other nanoporous materials, such as zeolites, in various aspects, but the specific framework flexibility is by far the feature that attracts the most attention. This lattice flexibility has a strong influence on physical properties such as elastic constants, thermal conductivity, diffusion,[11,40] and adsorption properties of guest molecules in the pores.[41–45] The latter phenomenon is mostly referred to as the breathing effect,[46–48] by which the host framework can shrink or expand. The force fields generated by QuickFF need to be able to describe these processes, which is a very challenging task.

When generating force fields for MOFs, special attention needs to be payed to describe the metal—ligand bond. Indeed these interactions are not easily described by suitable coordination bond terms in force fields, as they may also have large ionic contributions. For the organic linker a wealth of reliable FFs exists in literature but the combination of organic entities with inorganic building blocks poses extra complexities. Recently, various groups have proposed theoretical schemes to construct force fields for MOFs. First of all, MOF-FF was developed by the group of Schmid and coworkers.[49] The force field energy expression is very similar to the MM3 expression.[28] The electrostatic interactions are described using Gaussian charges. The covalent parameters are fitted to *ab initio* cluster data using a genetic algorithm. MOF-FF has been able to produce accurate force fields for a series of well-known MOFs such as MOF-5, HKUST-1, UIO-66, and others. During the fitting procedure of the force fields, additional user interference could be necessary to adjust the allowed range of the parameters or to run several parallel genetic algorithm runs and to combine them afterward. Very recently, the group of Walsh and coworkers developed the BTW-FF procedure.[50] Periodic *ab initio* calculations are used to fit the force field. The energy expression is identical to that of the MM3 force field, with some refinements for some parameters. The resulting force fields have been used to calculate bulk moduli, thermal expansion coefficients, and heat capacities. Presently, BTW-FF has been parameterized for a set of well-known MOFs such as MOF-5, HKUST-1, UiO-66, and others. Finally, UFF4-MOF[51] is an extension for the well-known Universal Force Field to describe MOFs. New parameters are provided for Al and all row four transition metal elements. Furthermore, additional O parameters are proposed that provide reliable structures of the metal oxide clusters of the connectors. These extra parameters were extracted from fits to *ab initio* cluster calculations. This extension can be used to construct a UFF force field for any MOF containing row 4 transition metals or Al. Although UFF4MOF has been shown to be very accurate in reproducing the unit cell dimensions of several MOFs, it needs to be tested in how far the parameter set will be accurate enough to simulate more exotic MOFs or physical phenomena that are more sensitive to the specific shape of the potential energy surface (PES), such as breathing.

QuickFF is an easy to use program that uses as input the equilibrium geometry and the Hessian matrix elements in cartesian coordinates. The intention is to derive force fields based on a maximal transfer of knowledge from the quantum mechanical system and that are transferable to larger systems. We implemented the procedure in a user-friendly Python code.* Currently, the Python code is written to read formatted checkpoint files of a frequency job performed in Gaussian,[52] but it can easily be extended to read input from other programs.

The article is structured as follows. In section 2, we briefly discuss the methodology as embedded in QuickFF for the derivation of force fields from *ab initio* calculations and elaborate on the practical usage. In section 3.1, the method is applied on a set of 1000 organic molecules to illustrate the ease with which one can derive accurate force fields for a large test set. As a proof of principle for the construction of force fields on MOFs, QuickFF is used to construct a force field for Mil-53(Al) and for MOF-5 (section 3.2).

*The Python code can be downloaded from the web-interface to the revision control system Git: https://github.com/molmod/QuickFF.

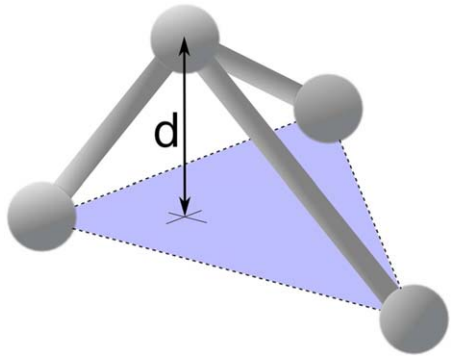

Figure 1. Illustration of the oop distance *d*. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Implemented Methods for Generation of Force Fields with QuickFF

Force field potential energy expression

In a force field, the various contributions to the potential energy are expressed in terms of atom types, that is, atoms with a similar chemical identity. In QuickFF, the atom type of each atom can be determined automatically based on local environment, or defined manually by the user. More information about the automatic assignation of atom types can be found in the Supporting Information. The force fields generated with QuickFF are composed of three contributions: a valence contribution describing covalent interactions between chemically bonded atoms ($V_{\mathrm{cov}}^{\mathrm{ff}}$), a van der Waals part describing the Pauli repulsion and dispersion interactions (using a Lennard-Jones or an MM3-Buckingham potential) and an electrostatic contribution governed by the Coulomb interaction between point charges or Gaussian distributed charge densities ($V_{\mathrm{el}}^{\mathrm{ff}}$):

$$V^{\mathrm{ff}} = V_{\mathrm{cov}}^{\mathrm{ff}} + V_{\mathrm{vdW}}^{\mathrm{ff}} + V_{\mathrm{el}}^{\mathrm{ff}} \tag{2.1}$$

$$V_{\mathrm{vdW}}^{\mathrm{ff}} = \begin{cases} \frac{1}{2}\sum_{\substack{ij=1\\(i\neq j)}}^{N_{\mathrm{at}}} \epsilon_{ij}\left[1.84\cdot 10^5 \cdot e^{-12\frac{r_{ij}}{\sigma_{ij}}} - 2.25\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] (\mathrm{MM3}) \\ \frac{1}{2}\sum_{\substack{ij=1\\(i\neq j)}}^{N_{\mathrm{at}}} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] (\mathrm{LJ}) \end{cases} \tag{2.2}$$

$$V_{\mathrm{el}}^{\mathrm{ff}} = \frac{1}{2}\sum_{\substack{ij=1\\(i\neq j)}}^{N_{\mathrm{at}}} \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} f(r_{ij}) \tag{2.3}$$

with

$$f(r_{ij}) = \begin{cases} 1 & \text{point charges} \\ \mathrm{erf}\left(\frac{r_{ij}}{d_{ij}}\right) & \text{Gaussian charges} \end{cases} \tag{2.4}$$

N_{at} is the total number of atoms in the system, Q_i is the charge of atom i and r_{ij} is the distance between atoms i and j, erf is the error function and $d_{ij} = \sqrt{d_i^2 + d_j^2}$ is the mixed radius of the Gaussian charges (d_i is the radius of the Gaussian charge distribution of atom i).

The valence contribution to the potential energy has the following mathematical form:

$$V_{\mathrm{cov}}^{\mathrm{ff}} = V_{\mathrm{bond}} + V_{\mathrm{bend}} + V_{\mathrm{torsion}} + V_{\mathrm{oopdist}} \tag{2.5}$$

$$V_{\mathrm{bond}} = \sum_{i=1}^{N_r} \frac{1}{2} K_{r,i}\left(r_i - r_{0,i}\right)^2 \tag{2.6}$$

$$V_{\mathrm{bend}} = \sum_{j=1}^{N_\theta} \frac{1}{2} K_{\theta,j}\left(\theta_j - \theta_{0,j}\right)^2 \tag{2.7}$$

$$V_{\mathrm{torsion}} = \sum_{k=1}^{N_\phi} \frac{1}{2} K_{\phi,k}\left[1 - \cos\left(m_k\left(\phi_k - \phi_{0,k}\right)\right)\right] \tag{2.8}$$

$$V_{\mathrm{oopdist}} = \sum_{l=1}^{N_d} \frac{1}{2} K_{d,l}\left(d_l - d_{0,l}\right)^2 \tag{2.9}$$

The force field includes harmonic bonds V_{bond}, harmonic bends V_{bend}, cosine dihedrals V_{torsion}, and harmonic out-of-plane (oop) distances V_{oopdist}. N_r is the total number of stretch bonds, N_θ the number of bending angles, N_ϕ the number of dihedral angles, and N_d the number of oop distances. An oop distance represents the distance from a plane determined by three atoms to a fourth atom that is only bonded to each of these three atoms (see Fig. 1). These oop distances are related to oop bends,[53] also called inversion terms. We prefer to work with oop distances because for every oop pattern, there is a unique oop distance. The introduction of oop terms enables us to accurately describe both the planarity of conjugated π-systems and the nonplanarity of some sp^3-units such as amines. The valence FF contains several unknown parameters: force constants ($K_{r,i}, K_{\theta,j}, K_{\phi,k}$, and $K_{d,l}$), rest values ($r_{0,i}, \theta_{0,j}, \phi_{0,k}$, and $d_{0,l}$) and multiplicity factors (m_k). All these parameters will be estimated in such a way that the force field reproduces the *ab initio* equilibrium geometry and *ab initio* Hessian in equilibrium as well as possible.

To account for the van der Waals (vdW) interaction, a repulsive short-range term—modeling the Pauli exclusion principle—and attractive long-range dispersion terms should be added to the force field potential. In the current version of QuickFF two vdW potentials are implemented, the Buckingham potential, as used in MM3,[28] and the Lennard-Jones (LJ) 6–12 type potential, as used in UFF.[17] The repulsive interaction is described better by the exponential part of the Buckingham potential than by the steep r^{-12} part of the LJ potential. The user has the choice to use one of the implemented potentials. More details on the practical implementation of these van der Waals interactions can be found in section 2.3. If necessary, QuickFF can easily be extended with other van der Waals schemes such as the ones developed by Grimme et al.

The last contribution to the FF potential is the electrostatic interaction. Atomic point charges can be derived from the quantum mechanical wave function belonging to the equilibrium geometry using one of the various partitioning schemes available in literature, for example, Hirshfeld-I,[54,55] Hirshfeld-E,[56] RESP[57] charges,... Hirshfeld-based schemes apply the

atom-in-molecule principle to partition the *ab initio* molecular electron density into overlapping atomic electron densities from which the charges can be derived. The RESP method estimates point charges by fitting them to the *ab initio* electrostatic potential. In QuickFF, the user is free to apply one of the available charge population schemes. The derivation of atomic charges is not part of the QuickFF procedure. Only the exclusion rule for the nonbonding interactions should still be chosen. An option is built in to exclude some interatomic electrostatic force terms (e.g., 1–3, 1–4 bonded pairs). More details on the practical implementation of these electrostatic interactions can be found in section 2.3.

Parameterization protocol implemented in QuickFF

The parameterization implemented in QuickFF aims at determining values for the force constants and rest values figuring in the covalent part of the force field expression. A small note regarding the nomenclature is in order. In this work, the term rest value is used for the parameters figuring in the harmonic energy term expressions while the term equilibrium value is used for the value of an internal coordinate in the equilibrium structure. These two values are not necessarily identical (which is also illustrated in the Supporting Information), as opposed to many force fields. The procedure consists of three steps as explained below.

Step 1: Determining the dihedral multiplicities and rest angles. In the first step, the dihedral multiplicities m_k and rest angles $\phi_{0,k}$ are determined directly from the equilibrium geometry based on local symmetry. Dihedral patterns belonging to the same atom types may have widely varying equilibrium angles. For example, the H—C—C—H dihedrals in ethane have values of 60°, 180°, and 300°. Hence, the dihedral potential should have local minima at each of these values. Therefore, we need to choose the multiplicity and rest angle accordingly, which is $m_k = 3$ and $\phi_{0,k}=60^\circ$ in the case of ethane. The general procedure for determining m_k and $\phi_{0,k}$ is explained in detail in the Supporting Information. In some cases, this procedure will not result in a unique dihedral potential of the type as given in expression (2.8) because of its simple mathematical form; we then choose to ignore that dihedral all together, in accordance with the QuickFF philosophy. If necessary one could construct more complicated forms of the potential afterward to represent the dihedral angles in particular cases, such as internal rotors.[33,36,58]

Step 2: Perturbation trajectories. In a second step, a new methodology is implemented to determine values of the force constants and rest values and addresses the correlation between the force field parameters. To this end trajectories are constructed along the multidimensional PES near the equilibrium structure. Changes to the potential energy along such trajectories can be modeled with the suggested FF expressions and compared with the first principles predictions, achieved by a second order Taylor expansion of the *ab initio* PES along the perturbation trajectory. Force field parameters can be extracted directly from the *ab initio* Hessian for each internal coordinate (IC) consequently by carefully choosing the trajectories. This is an important feature, as it does not require an ambiguous Cartesian-to-IC transformation of the Hessian, nor does it require complicated cost functions. As a result, a unique set of force field parameters will be obtained. In this section, we will briefly outline this procedure, more details can be found in the Supporting Information. To extract the force field parameters for a certain IC, we construct so-called perturbation trajectories. Each frame in the trajectory corresponds to a certain perturbation imposed on the IC under consideration. Consider a particular IC q_n and assume some small perturbation bringing it to the value $\tilde{q}_n$. The geometry corresponding to this perturbation is determined by relaxing all other degrees of freedom (apart from q_n) by minimizing the strain corresponding to all these other degrees of freedom. The strain is quantified by means of the following function:

$$\chi_n^S\left(\vec{R}\right)=\frac{1}{2}\sum_{m\neq n}\left[q_m(\vec{R})-q_m(\vec{R}_0)\right]^2 \tag{2.10}$$

The summation runs over all internal coordinates q_m, different from the perturbed IC q_n, and the minimization was converged with a threshold on the change of χ_n^S of 10^{-9}. Each IC in the summation is expressed in atomic units as a means of preconditioning. Repeating this procedure of minimal internal strain for several perturbation values $\tilde{q}_n$[†] yields trajectories $\vec{R}(\tilde{q}_n)$ along an internal coordinate q_n, that are as much as possible decoupled from the other IC's and in which all other IC's are relaxed as much as possible. 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 $V_{\rm cov}^{\rm ff}(\tilde{q}_n)$ along the perturbation trajectory, by a single harmonic potential in $\tilde{q}_n$.

$$V_{\rm cov}^{\rm ff}(\tilde{q}_n)=\frac{K_n}{2}\left(\tilde{q}_n-q_{n,0}\right)^2 \tag{2.11}$$

instead of the full expression of the covalent energy, where the variations of all bond lengths, bending angles, oop distances, and dihedral angles are taken into account. The unknown force field parameters figuring in expression 2.11, that is K_n and $q_{n,0}$, can now be estimated by expressing that the force field energy should be equal to the *ab initio* energy along the perturbation trajectory, apart from a constant shift c, for each chosen $\tilde{q}_n$:

$$V^{\rm ai}\left(\vec{R}(\tilde{q}_n)\right)=V_{\rm el}^{\rm ff}\left(\vec{R}(\tilde{q}_n)\right)+V_{\rm vdW}^{\rm ff}\left(\vec{R}(\tilde{q}_n)\right)+\frac{K_n}{2}\left(\tilde{q}_n-q_{n,0}\right)^2+c \tag{2.12}$$

Hence, one can fit a parabola to the difference between $V^{\rm ai}$ and $V_{\rm el}^{\rm ff}+V_{\rm vdW}^{\rm ff}$ 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 oop distances. No dihedral angles are considered in this step of the procedure. The main

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

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 based on local symmetry. 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.

The force constants derived in this step are overestimated because even along the perturbation trajectories with minimal strain, coupled IC's cannot be completely decoupled and by consequence contributions from other IC's along the trajectory $\vec{R}(\tilde{q}_n)$ should be considered instead of the single harmonic potential of eq. (2.11). The third step consists of a fine tuning to address this error and match the force field Hessian with the *ab initio* Hessian.

Step 3: Refinement of the force constants. In the third step, the harmonic force constants from the previous step are refined, and the still missing dihedral force constants are fitted according to a simple least-square cost function that measures the error between the various *ab initio* Hessian and the force field Hessian matrix elements expressed in Cartesian coordinates:

$$\chi^H(\vec{K}) = \sum_{i \le j} \left(\left[\mathbf{H}^{\mathrm{ai}}\right]_{ij} - \frac{\partial^2 V^{\mathrm{ff}}_{\mathrm{el}}}{\partial R_i \partial R_j} - \frac{\partial^2 V^{\mathrm{ff}}_{\mathrm{vdW}}}{\partial R_i \partial R_j} - \frac{\partial^2 V^{\mathrm{ff}}_{\mathrm{cov}}}{\partial R_i \partial R_j}\left(\vec{K}\right) \right)^2 \quad (2.13)$$

$\vec{K}$ represents a vector containing all force constants. All force constants are constrained to be positive. In addition, we impose that all dihedral force constants are smaller than 200 kJ/mol to prevent possible compensation effects. The rest values, as extracted from step 2, are kept fixed. In this way, troublesome fitting deficiencies due to the correlations between force constants and rest values are avoided. The constrained optimization is performed using the Sequential Least SQuares Programming[59] algorithm, with a convergence threshold on the change of χ^H of 10^{-9}.

Practical usage

QuickFF can be used in two ways: from a command-based terminal by means of a single command using the script `qff-est.py` or by writing an external script and importing QuickFF as a Python library. The former is very straightforward to use and will be outlined in this section, the latter is less straightforward but allows more control. Both usages are also documented on line at `http://molmod.github.io/QuickFF/`.

The most straightforward way to construct a force field from the command line is by means of the following command:

```
qff-est.py [options] fns
```

`fns` is a space-separated list of input file names, while options is a list of space-separated optional keyword arguments with the format `--key=value`. Several input file formats are supported for reading input data. For example, a Gaussian formatted checkpoint file can be specified to read the *ab initio* geometry, forces and Hessian in equilibrium and a HDF5 file can be used to specify the electrostatic and/or van der Waals parameters. All supported file formats are discussed in detail in the online documentation. The optional keyword arguments can be used to control the force field model, all of these options are documented on line as well as by means of the `--help` option. The list below enumerates the most important options:

- `--atypes-level`: The level of automatic atom type assignation (`low`, `medium`, `high`, or `highest`) as explained in the Supporting Information. By default, the atom types are taken from the input files.
- `--ei-model`: Defines the potential used for the electrostatic interactions. Can be `CoulPoint`, `CoulGauss`, `HarmPoint`, `HarmGauss`, or `Zero`. The Harm variants use a second-order Taylor expression to speed up the calculation considerably (and remain accurate). The default is `HarmPoint`.
- `--ei-scales`: Defines the scaling rule for the electrostatic interactions. Three comma-separated floats are required. The first one sets the scale for atoms separated by one bond, the second for atoms separated by two bonds, and the third for atoms separated by three bonds. The default is 0.0,0.0,1.0.
- `--ei-path`: Defines the path in the HDF5 file, given as an input file, from which the charges (and optionally the radii in case gaussian distributed charges are required) will be extracted.
- `--vdw-model`: Defines the potential used for the van der Waals interactions. Can be `LJ`, `MM3`, `HarmLJ`, `HarmMM3`, or `Zero`. The default is `Zero`.
- `--vdw-scales`: Defines the scaling rule for the van der Waals interactions. Three comma-separated floats are required. The default is 0.0,0.0,1.0.
- `--vdw-path`: Defines the path in the HDF5 file, given as an input file, from which the van der Waals parameters, epsilons and sigmas, will be extracted.

To illustrate the usage of the `qff-est.py` script, suppose one disposes of a Gaussian formatted checkpoint file (`gaussian.fchk`) that is the result of a frequency job performed on the equilibrium geometry and a HDF5 file (`gaussian_wpart.h5`) that contains Iterative Hirshfeld charges in the path `/wpart/hi`. One can now easily construct a force field using automatically assigned atom types according to the level high, containing a covalent term for all IC's, an electrostatic part in which all atom pairs are allowed to interact (including bonded atoms) using the Iterative Hirshfeld point charges and no van der Waals terms. This can be achieved by means of the following command:

```
qff-est.py --atypes-level=high --ei-path=/wpart/hi --ei-scales=1,1,1
gaussian.fchk gaussian_wpart.h5
```

The option `--ei-model` is not needed here because its default value is the desired one.

Applications

QuickFF on a large set of organic molecules

The ease of generating accurate force fields in a fast and efficient way for a large set of molecules is illustrated by applying

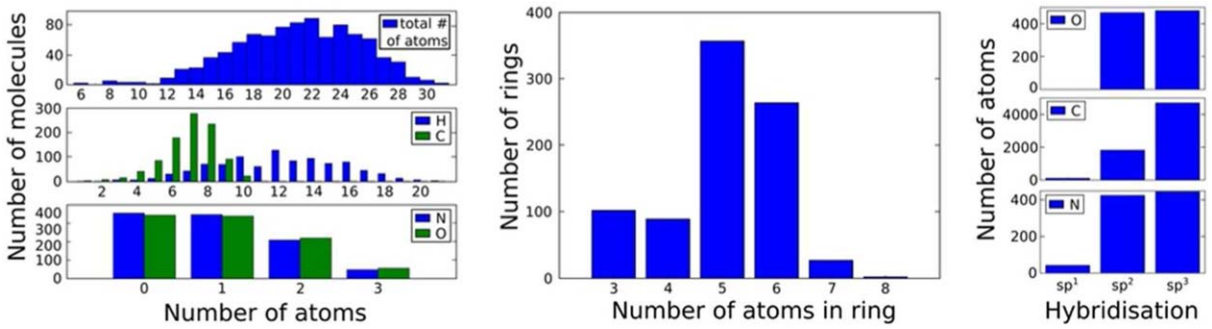

Figure 2. Histograms to illustrate the diversity of the test set. Left: distribution of the total number of atoms in a molecule and number of C, H, O, and N atoms in a molecule. Middle: distribution of the size of cyclic patterns. Right: distribution of the hybridization of C, N, and O atoms.

the program on a set of 1100 organic molecules. To this end, 1100 molecules were randomly selected from a subset of the PubChem Compound database.[60] Several selection criteria were introduced to define a subset that can be accepted as sufficient for a serious assessment of the QuickFF procedure for deriving force fields. The detailed procedure to select randomly a 1000 molecules from an initial set of 4,083,66 molecules from the PubChem Compound database is outlined in the Supporting Information. To illustrate the chemical diversity in this final set of molecules, we constructed various histograms, shown in Figure 2. The figure illustrates the distribution of several properties: the total number of atoms in a molecule, the number of C, H, N, and O atoms in a molecule, the size of cyclic patterns and the hybridization of C, N, and O atoms. OpenBabel[61] was used to determine the hybridization of each atom.

For every molecule in this set, *ab initio* calculations have been performed to generate the reference data, serving as input for the derivation of the force field. First, the geometry of the molecule was optimized and the *ab initio* Hessian was constructed with corresponding frequencies using DFT. The B3LYP[62–64] density functional was chosen for its well-known accuracy in describing small molecules consisting of first and second row atoms,[65,66] together with the 6–311+G(d,p) basis set.[67–69] These calculations were performed with the Gaussian 09[52] program. The optimization was done using the VeryTight convergence criterion and the scf(tight, xqc) and int(grid = ultrafine) options were selected for both the optimization and frequency tasks. Next, the atomic charges were computed using the Hirshfeld-E[56] partitioning scheme with Horton,[70] a development platform for electronic structure methods. With the Gaussian output and the atomic charges, we derived a force field using QuickFF without an electrostatic exclusion rule, and atom types derived according to the atom type level "high" (see Supporting Information). Next, the force fields were used to perform a geometry optimization and to calculate the Hessian with YAFF,[71] an in-house developed force-field code for molecular simulations of both periodic and nonperiodic systems. Finally, the frequencies were calculated using TAMkin,[72] a postprocessing toolkit for normal mode analysis, thermochemistry, and reaction kinetics. One remark concerns the *ab initio* method used to generate the reference data. Due to computational efficiency, DFT methods are mostly preferred. For the training set used here B3LYP was selected, which includes no dispersion interaction. In this case, the van der Waals interactions can be added to full strength afterward to the force field potential without the risk of double counting, as the reference *ab initio* data did not contain any dispersion forces. For some molecules, such as large organic molecules with interacting phenyl groups, accurate geometries may only be obtained using functionals including long-range dispersion (such as M06-2X[73,74]) or with B3LYP+D3 correction terms of Grimme[37] (at each optimization cycle). However, such molecules are absent in the test set presented here, and it is hence justified to optimize the geometry at the *ab initio* level without dispersion interactions. Fitting the parameters of the bonded FF terms indirectly incorporates the influence of dispersion. Including additional vdW terms to the total FF potential may lead to an overestimation of the dispersion effects. One can account for this double counting by introducing a scale factor to the nonbonded interactions, but there is no conclusive protocol to fix it. Another possibility is to restrict the vdW terms to 1,4 interactions and higher, that is excluding 1,2 and 1,3 interactions, to restrict the van der Waals terms from influencing the bond lengths and bending angles.

The quality of the force fields generated by QuickFF for the large dataset is first validated by comparing equilibrium bond lengths, bending angles, dihedral angles, oop distances, and frequencies with regard to the reference data and by comparing its performance to the results obtained with two well-known general force fields, UFF[17] and GAFF.[18] These two force fields are universal force fields adjusted to reproduce mostly experimental data. The validity of QuickFF on small organic units is also a necessary condition for its application to hybrid materials such as MOFs.

The results are shown in Figure 3 for four force fields (QuickFF with van der Waals terms taken from UFF, QuickFF without van der Waals terms, GAFF, and UFF). Each row in Figure 3 displays the scatter plots of a certain observable (bond lengths, bending angles, etc.). Each dot represents the value of an internal coordinate or a vibrational mode of a molecule in the dataset. The *x*-value represents the *ab initio* value, while the value on the *y*-axis reports the force field prediction. The diagonal line corresponds to the situation where the force

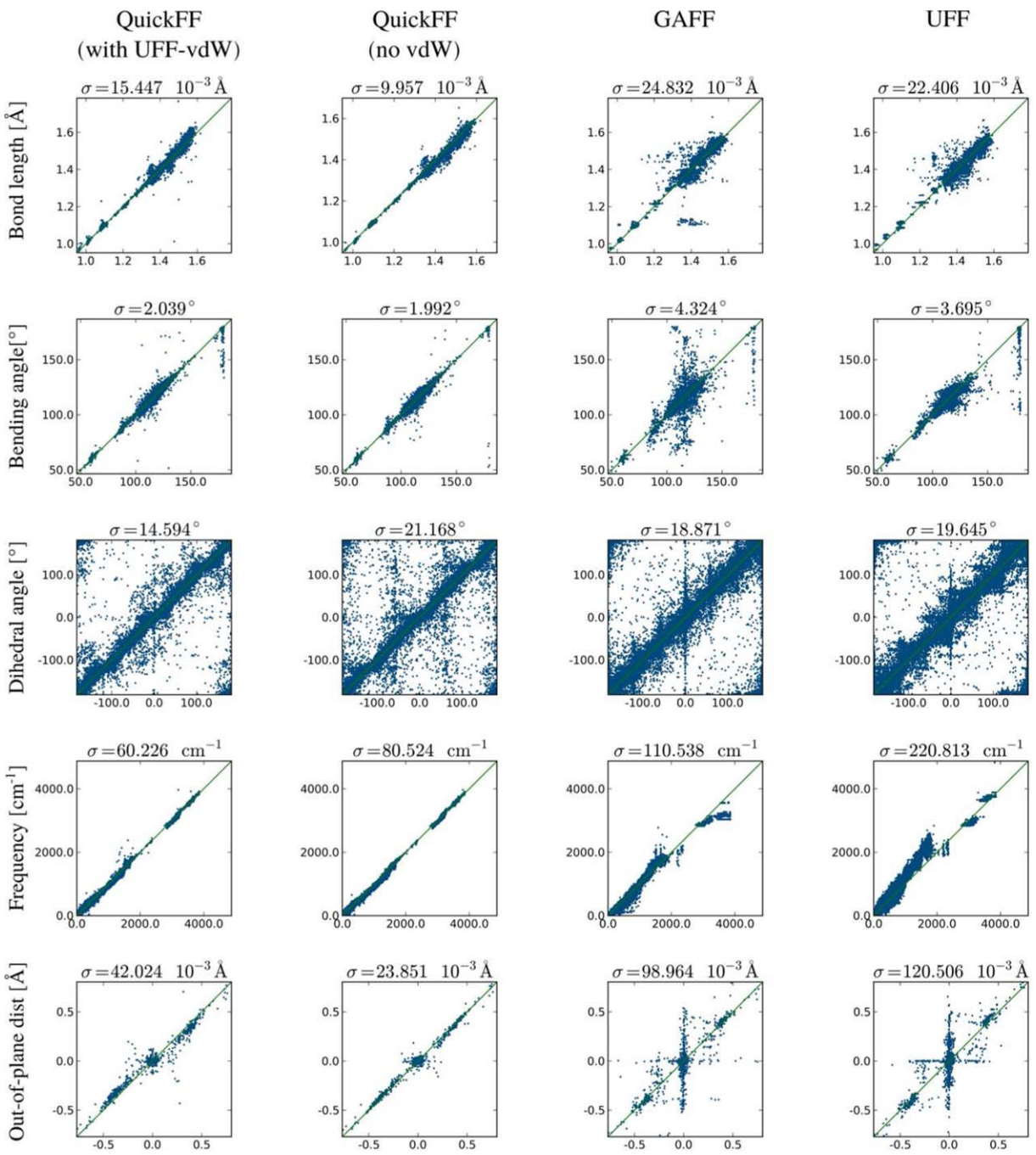

Figure 3. Scatter plots visualizing the performance of the force fields [QuickFF (with vdW from UFF), QuickFF (without vdW), GAFF, and UFF] in reproducing geometrical quantities and frequencies with respect to the *ab initio* reference data. The *x*-value reports the *ab initio* value of the various properties for each molecule of the dataset, while the FF prediction is given on the *y*-axis. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

field exactly reproduces the *ab initio* reference value. Above each scatter plot, the standard deviation is included, which gives an indication of the error between the force field predictions and the *ab initio* reference data.

The scatter plots of Figure 3 show that QuickFF performs very well compared to the other general force fields. The standard deviation of any observable is lower for QuickFF than for both GAFF and UFF (except for the dihedrals, see later). This is not really surprising as the QuickFF parameters have been fitted for each molecule separately and sufficient flexibility was taken into account in the assignment of atom types. By construction, the two universal FF's GAFF and UFF allow less flexibility in the assignment of atom types. Bond lengths are relatively well reproduced but a significantly larger scattering is noticed for the bending angles. The lack of an oop term in the covalent force field expression in GAFF and UFF gives rise to a cross pattern centered at the origin. It implies that *ab initio* and force field predictions for the oop structure belonging to the equilibrium

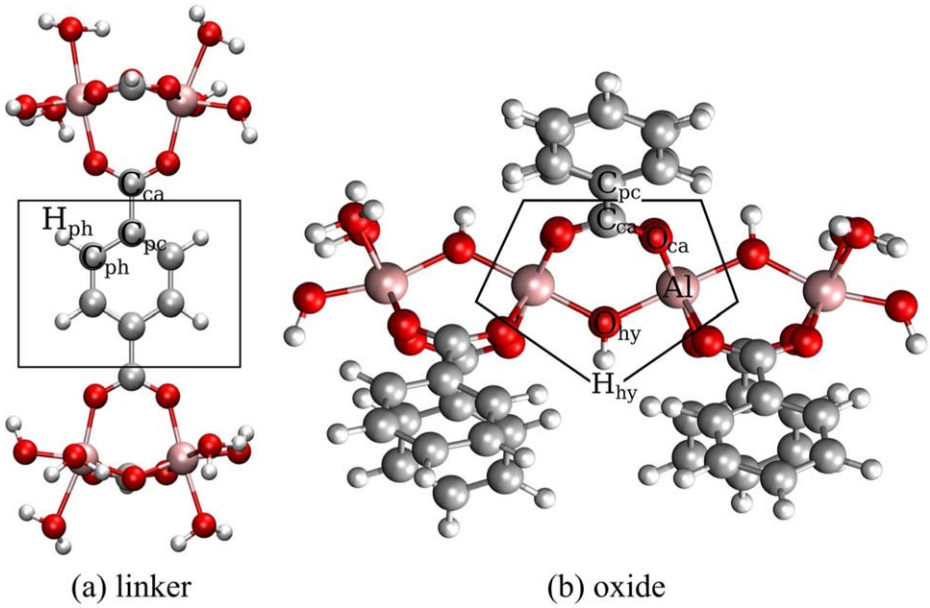

Figure 4. (left) Linker cluster and (right) oxide cluster on which QuickFF was performed. The solid line defines the core region used to define atom types relevant for the periodic system. Reprinted (adapted) with permission from Ref. [15]. Copyright (2014) American Chemical Society. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

geometry are not consistent with each other. Planar configurations may result from *ab initio* calculations, while force fields rather predict a nonplanar equilibrium, or *vice versa*. In the scatter plots, these cases give rise to the presence of horizontal and vertical branches. No such cross pattern is present in the scatter plot of QuickFF, which clearly illustrates the added value of the valence force field terms containing the oop distances.

The series of scatter plots related to the dihedral angles show that these internal degrees of freedom are more difficult to reproduce with all investigated force fields. This could be anticipated because soft dihedral angles, for example, dihedrals related to internal rotors, are typically the internal degrees of freedom that are the most flexible. Moreover, these soft dihedrals correspond to anharmonic motions while the Hessian used in the reference data for QuickFF only involves harmonic motions. Hence, insufficient information is incorporated in the model to accurately fit the force field terms of soft dihedrals. As QuickFF is designed to accurately generate first-generation force fields for a large set of molecules, this does not pose a serious problem. In cases where such soft dihedral angles come into play prominently, a second generation force field can be built including information on the torsional potential of the internal rotor. Such refinements come at a higher computational cost as additional *ab initio* data need to be performed, as such this procedure is not taken up in the standard workflow of QuickFF which is designed to generate force fields in a fast and easy way.

The scatter plots of Figure 3 further reveal that van der Waals interactions only have a small impact on the geometry and the frequencies of these small organic molecules. However, one must be careful not to extrapolate these conclusions to larger molecules without further investigation. As our final aim is the construction of a transparent methodology for hybrid materials, these extensions are beyond the scope of the present article.

Quick and easy force field generation for MIL-53(Al) and MOF-5

The main reason to develop QuickFF is to have a transparent and easy protocol for the construction of viable force fields for MOFs. Easy generation of force fields for these materials could be beneficial to screen these materials quickly for optimal properties. Hereafter, QuickFF is used to generate force fields for two materials MIL-53(Al)[75] and MOF-5.[76] MIL-53(Al) was chosen as it is one of the prototype MOFs that shows an intriguing flexible behavior, having the capacity to undergo large structural deformations.[75,77] For this material, some of the present authors already constructed a force field earlier, but the procedure required various manual interventions and cannot be regarded as quickly applicable to a larger set of materials.[15] However, the original constructed force field was very accurate, as it was able to reproduce the breathing behavior of the material properly and predict the transition pressure of the mechanically induced transition from large pore to narrow pore.[78] QuickFF can only be approved for further applications if it equally well reproduces all these properties. The other benchmark system is MOF-5, which is the first MOF for which a specific force field was developed by the group of Schmid and coworkers.[79] This force field was able to reproduce geometry, negative thermal expansion, and benzene diffusion.

Validation for MIL-53(Al). The generation of the force field for MIL-53(Al) relies on *ab initio* calculations on two isolated clusters, which are representative for the metal-oxide unit at one

hand and the terephthalate linker on the other hand (see Fig. 4). The computational details can be found in Ref. [15]. In a next step, QuickFF is applied to each individual clusters. To be consistent with our previous work, some assumptions are made, which facilitates the comparison of the newly derived QuickFF force field directly with our previous force field. First, no van der Waals interactions were included *a priori*, instead they will be added a posteriori and taken from the MM3 force field. Second, we include all electrostatic interactions without any exclusions and we use the same charges. Third, a covalent term was associated with all internal coordinates. As oop vibrations have not been included in the covalent terms of the previous FF, we constructed two force fields with QuickFF with and without $V_{oopdist}$. As a result, the performance of the two force fields can be compared on equal basis, and at the same time the added value of the oop term in the covalent interaction can be examined.

In a next step, the data of the individual clusters need to be merged to produce a force field for the periodic crystal. The same procedure was followed as in our earlier work, in which a core region is defined in both systems. These core regions are illustrated in Figure 4 and consist of atom types that are considered relevant for the periodic crystal. From each cluster force field, energy terms are only retained when their internal coordinates have at least one atom in the core region. Whenever a term is present in the force fields of both clusters, their parameters are averaged over both clusters. The resulting force field parameters can be found and compared with our previous force field in the Supporting Information. This comparison reveals that the force field parameters are not differing much from each other, the largest deviations are noticed in the dihedrals. The concept applied here nicely illustrates how to build force fields for periodic systems from *ab initio* data generated on smaller building units. Such a procedure is beneficial in terms of computational time, as the computation of accurate Hessians for periodic systems is nontrivial and comes at a large computational cost.[80] The validation of the FF should take place at the level of periodic calculations, for which the flexibility, the geometry, unit cell dimensions, energy profile of the breathing mode should be described appropriately.

Starting from an initial structure of the material, the atomic coordinates and unit cell parameters are fully relaxed during the geometry optimization of the periodic structure with the force field. The resulting equilibrium values of several bond lengths are compared with the *ab initio* predictions extracted from the extended cluster calculations as they are used as reference data in the parametrization of the FF (see Table 1). Both force fields (with and without oop terms) succeed in accurately reproducing the bond lengths (a more extended comparison including bending angles and dihedral angles is given in the Supporting Information). Furthermore, the table also reveals that the new QuickFF force field and the previous force field perform equally.

The ultimate validation of the force field generated with QuickFF, is the prediction of a correct flexibility behavior for the unit cell. By starting from initial geometries close to the experimental structure of the narrow pore (np) and large pore (lp) phase, we can determine the unit cell dimensions for both phases with the force field. The definition of the unit cell dimensions is illustrated in the Supporting Information. In Table 2, we compare the various force-field predictions with the experimental results of Liu et al.[77] We can conclude that all FF variants perform fairly well, however, the presence of an oop term in the valence potential energy improves the prediction of the large pore unit cell.

Table 1. *Ab initio* optimized bond lengths (in Å) of the linker and oxide cluster compared with the periodic predictions made by our previous[15] force field and the present QuickFF force field.

				QuickFF	
ATYPES	Linker	Oxide	Previous	nooop	oop
$AL—O_{CA}$	1.93	1.92	1.93	1.87	1.86
$AL—O_{HY}$	–	1.86	1.88	1.83	1.82
$C_{CA}—C_{PC}$	1.50	1.49	1.50	1.49	1.49
$C_{CA}—O_{CA}$	1.27	1.27	1.27	1.26	1.26
$C_{PC}—C_{PH}$	1.40	1.41	1.41	1.40	1.41
$C_{PH}—C_{PH}$	1.39	–	1.39	1.39	1.39
$C_{PH}—H_{PH}$	1.08	1.08	1.08	1.08	1.08
$H_{HY}—O_{HY}$	–	0.96	0.91	0.96	0.96
MD			-3.7×10^{-3}	-1.5×10^{-2}	-1.6×10^{-2}
MAD			8.8×10^{-3}	1.5×10^{-2}	1.6×10^{-2}
STD			2.0×10^{-2}	2.1×10^{-2}	2.6×10^{-2}

Two options are investigated in QuickFF: inclusion of out-of-plane distance terms (oop) or not (nooop). For clarity, the mean deviation (MD), mean absolute deviation (MAD), and standard deviation (STD) of each FF column are given with respect to the average of linker and oxide (or only one of the clusters in case the IC is absent in the other cluster).

Finally, the energy profile along the interdiagonal angle θ is computed. This energy profile is characteristic for a breathing material, because the motion related to a variation in θ represents this breathing motion. The result is plotted in Figure 5. According to Walker et al.,[81] the energy difference between np and lp should be in the range 33–42 kJ mol^{-1} depending on which functional is used in the DFT-D method. The force field, presented in our previous work,[15] predicts an energy difference of 60 kJ mol^{-1}, which was outside the range suggested by Walker. At that point, we argued that a scaling of the MM3-vdW parameters had a large impact on the quantitative energy differences between the large and narrow pore forms. The results generated by QuickFF in the two options (oop and nooop) demonstrate that the energy profile is very sensitive to the method of how dihedral forces are treated and the inclusion of oop motions. The QuickFF/nooop force field predicts an energy difference of about 39 kJ mol^{-1} per unit cell, which is consistent with the range of Walker, while a slightly different parametrization due to incorporation of oop terms, increases the np–lp barrier to 49 kJ mol^{-1}, which more consistent with our previous force field. In our earlier force field (Ref. [15]), the force constants corresponding to the $Al—O_{CA}—C_{CA}—C_{PC}$ and $Al—O_{CA}—C_{CA}—O_{CA}$ dihedrals had to be reparametrized to reproduce the periodic behavior more accurately. Therefore, an *ab initio* scan of the periodic structure along the interdiagonal angle θ was necessary at that point and thus the procedure required some manual interventions

Table 2. Comparison of the unit cell predicted by the force field with the experimental cell parameters (a) Ref. [77] (b) Ref. [15].

	Narrow pore				Large pore			
		QuickFF				QuickFF		
	Exp[a]	oop	Nooop	**Prev**[b]	**Exp**[a]	oop	nooop	**Prev**[b]
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

and additional generation of *ab initio* data. In this case, using QuickFF/oop, such manual interventions are no longer necessary and thus the procedure is more transparent to use to a broader set of MOFs. For MIL-53(Al), the prototype example of a flexible MOF, QuickFF successfully passed the validation and thus the protocol may be used to other hybrid flexible materials.

Validation for MOF-5: QuickFF versus MOF-FF. MOF-FF is the force field protocol developed by Schmid and coworkers, and is very successful in generating force fields for a variety of hybrid materials, but as we mentioned earlier, manual interventions may prove necessary during the fitting procedure of the force field. For MOF-5, an assessment is made on the force field generated by QuickFF and by MOF-FF. Both methodologies have some common features but also some differences in the analytical FF expression for the description of the covalent part of the force field. In MOF-FF, a Morse potential is used to describe the coordination bonds in which the metal is involved. Additionally, some stretch–stretch and stretch–bend cross terms are taken into account. We generated a force field using QuickFF starting from the same reference data of a single cluster (Fig. 6), used for the MOF-FF force field.[49] Instead of point charges the same spherical Gaussian charge distributions are used in the two FF protocols.

A first validation of the force fields concerns the reproduction of the equilibrium geometry and the vibrational frequencies of the model system. Bond lengths are geometrical parameters, that are the most sensitive to variations of the FF parameters. They are tabulated in Table 3. There is a slight preference for MOF-FF in reproducing the *ab initio* reference data. The largest deviation is noticed in the metal–oxygen bond distance O_{ce}—Zn, which is slightly underestimated by QuickFF. The remaining geometrical parameters are all accurately reproduced; the whole list is reported in the Supporting Information. The final goal of the FF is its adequacy to predict properties of the MOF-5 framework. Both force fields were able to reproduce the cubic nature of the unit cell as all three

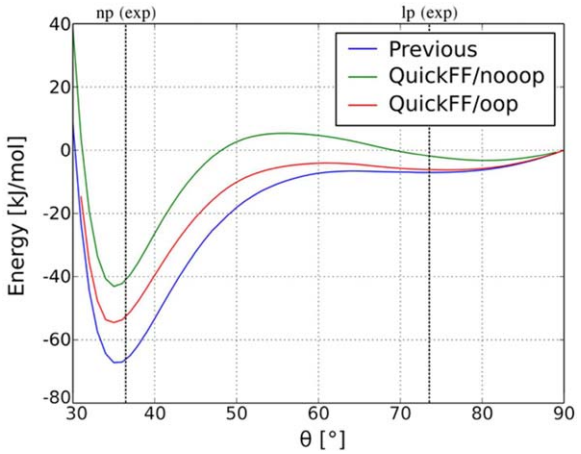

Figure 5. Breathing profile of Mil-53(Al) according to the new QuickFF force field (with the two options oop and nooop) and the previous force field (with unscaled MM3-vdW).

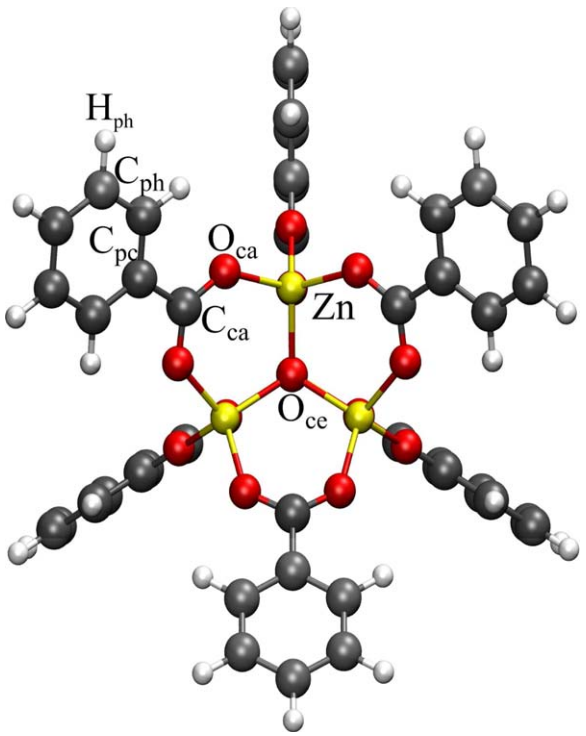

Figure 6. Structure of basic zinc benzoate as model system for MOF-5. All atomic types are given. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 3. Comparison of bond lengths in the benzoate cluster as predicted by DFT, QuickFF, and MOF-FF.

ATYPES	DFT	QuickFF	MOF-FF
C_{ca}—C_{pc}	1.497	1.510	1.497
C_{ca}—O_{ca}	1.271	1.274	1.271
C_{pc}—C_{ph}	1.404	1.413	1.402
C_{ph}—C_{ph}	1.398	1.402	1.399
C_{ph}—H_{ph}	1.090	1.090	1.093
O_{ca}—Zn	1.960	1.968	1.962
O_{ce}—Zn	1.963	1.921	1.967
MD		-7.1×10^{-4}	1.1×10^{-3}
MAD		1.1×10^{-2}	1.7×10^{-3}
STD		1.9×10^{-2}	2.0×10^{-3}

For clarity, the mean deviation (MD), mean absolute deviation (MAD) and standard deviation (STD) of each FF column are given with respect to the average of linker and oxide (or only one of the clusters in case the IC is absent in the other cluster).

angles are 90° and all three cell dimensions equal. The cubic lattice constant has a value of 26.173 Å for QuickFF, 26.080 Å for MOF-FF, and 25.885 Å from experiment.[76] Hence, both force fields reasonably succeed in reproducing the experimental estimates, with a slight preference to MOF-FF. A serious test for the force fields is the prediction of the (low) frequencies of the normal modes. They are tabulated in the Supporting Information, but a one-to-one correspondence of the data, resulting from the two force fields, is not meaningful due to the degeneracies corresponding to each normal mode, which are obviously different for each FF (different terms). A correlation diagram is, therefore, more instructive and transparent (Fig. 7). In this diagram, both the QuickFF and MOF-FF frequencies are sorted numerically and plotted against each other. The stepwise increase of the frequencies in the correlation diagram (best visualized in the inset of the figure) points toward a different degeneracy observed in the QuickFF frequencies, but the global trend of the two sets of data is similar at least for frequencies below 1400 cm^{-1}.

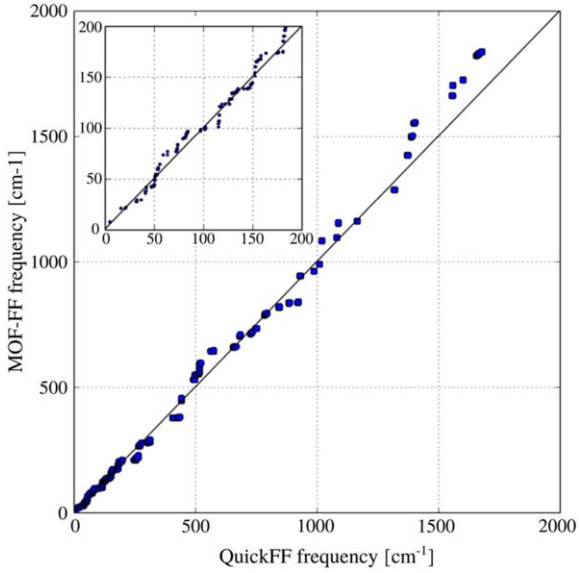

Figure 7. Comparison of frequencies of MOF-5 as calculated by QuickFF and MOF-FF. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Conclusions

Quick FF is a new and fast protocol to derive force fields for isolated and complex molecular systems from *ab initio* calculations. The development of the new software is inspired by the quest from the MOF community to determine in a fast, transparent and easy way force fields for these new type of hybrid materials. The input data for QuickFF consists of *ab initio* equilibrium geometries and a Hessian on smaller building units. The mathematical expression for the covalent energy terms is kept simple using harmonic terms for bond lengths, bending angles, and oop distances and single cosine functions for dihedral angles. Such an approach was preferred to ensure robustness and to avoid fitting deficiencies as much as possible. The parameters of the electrostatic and van der Waals interactions are assumed to be known *a priori* and can be taken from population schemes that are available in literature. An option is built in to spread the atomic point charge over a spherical Gaussian distribution function centered on the nucleus. The resulting force fields are intended for direct use in molecular simulations. If they are too simple to describe more complex systems, they still provide a first generation force field for further fine tuning.

A new methodology was implemented that relies on the generation of perturbation trajectories around the equilibrium. For each internal coordinate, a trajectory was constructed to minimize the strain along the trajectory generated by the other IC's. Finally, all force constants were refined using a least-square cost function that measures the error between the force field and *ab initio* Hessian expressed in Cartesian coordinates. In QuickFF, there is no need for complicated cost functions nor for an introduction of weight factors.

The QuickFF procedure was illustrated by applying it on a large set of organic molecules, for which the new force fields succeeded in reproducing the *ab initio* geometry and Hessian in equilibrium. Furthermore, QuickFF performed well in comparison with general force fields such as UFF and GAFF. Especially the introduction of a term containing oop distances was shown to be very valuable.

To show the validity of the protocol for generation of force fields on MOFs, two materials (MIL-53(Al) and MOF-5) were studied for which force fields are available in literature but which were constructed using several manual interventions. The force fields for the periodic structures are constructed on basis of a building block concept as originally introduced by Schmid and coworkers. *Ab initio* calculations are performed on smaller building units, from which force-field parameters are deduced using the QuickFF methodology. Afterward, the parameters are merged in an appropriate way to generate a force field for the periodic structure. For MIL-53(Al), QuickFF fully complies with the expectations, the force field was able to reproduce the geometries, unit cell dimensions, and relative stabilities of the large and narrow pore phases. Furthermore, the new force field was able to predict a correct breathing profile for this flexible material.

In addition, QuickFF and MOF-FF were compared for generating a force field for MOF-5. The two FF methodologies succeed in achieving the expected accuracy. The more elaborated MOF-FF turned out to be slightly superior to the present version of QuickFF in reproducing properties of the framework. QuickFF was developed within the scope of offering a fast and easy recipe to the MOF community to construct a valuable FF for any MOF structure.

QuickFF has been implemented in a user-friendly Python code and is available via http://github.com/molmod/QuickFF. The program is clearly valuable for the screening of large databases of MOFs and for the derivation of their properties based on extensive MD or Monte Carlo simulations.

Acknowledgment

Computational resources (Stevin Supercomputer Infrastructure) and services were provided by Ghent University.

Keywords: QuickFF · automated software · force-field development · metal-organic frameworks · molecular simulation

How to cite this article: L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, V. Van Speybroeck. *J. Comput. Chem.* **2015**, DOI: 10.1002/jcc.23877

Additional Supporting Information may be found in the online version of this article.

[1] P. Charifson, Practical Application of Computer-aided Drug Design; Marcel Dekker: New York, **1997**.
[2] A. Leach, Molecular Modelling: Principles and Applications, 2 ed.; Prentice-Hall: Harlow, England, **2001**.
[3] C. Cramer, Essentials of Computational Chemistry: Theories and Models; Wiley: New York, **2002**.
[4] P. Kollman, D. Case, In Burger's Medicinal Chemistry and Drug Discovery, Vol. 1, 6th ed.; Wiley: New York, **2003**.
[5] T. Düren, L. Sarkisov, O. M. Yaghi, R. Q. Snurr, *Langmuir* **2004**, *20*, 2683.
[6] Y. D. Kuang, S. Q. Shi, P. K. L. Chan, C. Y. Chen, *Int. J. Nonlinear Sci. Numer. Simul.* **2010**, *11*, 121.
[7] W. Somers, A. Bogaerts, A. van Duin, S. Huygh, K. Bal, E. Neyts, *Catal. Today* **2013**, *211*, 131.
[8] Q. Yang, C. Zhong, *J. Phys. Chem. B* **2005**, *109*, 11862.
[9] G. Garberoglio, A. Skoulidas, J. Johnson, *J. Phys. Chem. B* **2005**, *109*, 13094.
[10] F. Salles, A. Ghoufi, G. Maurin, R. G. Bell, C. Mellot-Draznieks, G. Ferey, *Angew. Chem. Int. Ed.* **2008**, *47*, 8487.
[11] S. Amirjalayer, M. Tafipolsky, R. Schmid, *Angew. Chem. Int. Ed.* **2007**, *46*, 463.
[12] M. Wehring, J. Gascon, D. Dubbeldam, F. Kapteijn, R. Q. Snurr, F. Stallmach, *J. Phys. Chem. C* **2010**, *114*, 10527.
[13] F. Salles, H. Jobic, T. Devic, V. Guillerm, C. Serre, M. M. Koza, G. Ferey, G. Maurin, *J. Phys. Chem. C* **2013**, *117*, 11275.
[14] B. Liu, B. Smit, *Langmuir* **2009**, *25*, 5918.
[15] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, V. Van Speybroeck, *J. Chem. Theory Comput.* **2012**, *8*, 3217.
[16] A. Ghysels, L. Vanduyfhuys, M. Vandichel, M. Waroquier, V. Van Speybroeck, B. Smit, *J. Phys. Chem. C* **2013**, *117*, 11540.
[17] A. Rappé, C. Casewit, K. Colwell, W. Goddard, W. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.
[18] J. Wang, R. Wolf, J. Caldwell, P. Kollman, D. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
[19] S. Mayo, B. Olafson, W. Goddard, *J. Phys. Chem.* **1990**, *94*, 8897.
[20] W. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657.
[21] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
[22] S. Weiner, P. Kollman, D. Nguyen, D. Case, U. Singh, C. Ghio, G. Alagona, Jr., S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765.
[23] S. Weiner, P. Kollman, D. Nguyen, D. Case, *J. Comput. Chem.* **1986**, *7*, 230.
[24] W. Cornell, P. Cieplak, C. Bayly, I. Gould, Jr., K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, P. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
[25] A. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrac, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
[26] A. MacKerell, N. Banavali, N. Foloppe, *Biopolymers* **2001**, *56*, 257.
[27] A. MacKerell, M. Feig, C. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.
[28] N. Allinger, Y. Yuh, J. Lii, *J. Am. Chem. Soc.* **1989**, *111*, 8551.
[29] N. Allinger, K. Chen, J. Lii, *J. Comput. Chem.* **1996**, *17*, 642.
[30] I. Cacelli, G. Prampolini, *J. Chem. Theory Comput.* **2007**, *3*, 1803.
[31] V. Barone, I. Cacelli, N. De Mitri, D. Licari, S. Monti, G. Prampolini, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736.
[32] S. K. Burger, M. Lacasse, T. Verstraelen, J. Drewry, P. Gunning, P. W. Ayers, *J. Chem. Theory Comput.* **2012**, *8*, 554.
[33] L. Huang, B. Roux, *J. Chem. Theory Comput.* **2013**, *9*, 3543.
[34] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell Jr., *J. Comput. Chem.* **2010**, *31*, 671.
[35] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, J. C. Gumbart, *J. Comput. Chem.* **2013**, *34*, 2757.
[36] S. Grimme, *J. Chem. Theory Comput.* **2014**, *10*, 4497.
[37] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[38] C. Wilmer, M. Leaf, C. Lee, O. Farha, B. Hauser, J. Hupp, R. Q. Snurr, *Nat. Chem.* **2012**, *4*, 83.
[39] Y. Colon, R. Snurr, *Chem. Soc. Rev.* **2014**, *43*, 5735.
[40] F. Salles, H. Jobic, A. Ghoufi, P. L. Llewellyn, C. Serre, S. Bourelly, C. Serre, D. Vincent, S. Loera-Serna, Y. Filinchuk, G. Férey, *Angew. Chem. Int. Ed.* **2009**, *48*, 8335.
[41] N. A. Ramsahye, G. Maurin, S. Bourrelly, P. L. Llewellyn, C. Serre, T. Loiseau, T. Devic, G. Ferey, *J. Phys. Chem. C* **2008**, *112*, 514.
[42] J. R. Karra, K. S. Walton, *J. Phys. Chem. C* **2010**, *114*, 15735.
[43] B. Arstad, H. Fjellvag, K. O. Kongshaug, O. Swang, R. Blom, *Adsorption* **2008**, *14*, 755.
[44] E. Stavitski, E. Pidko, S. Couck, T. Remy, E. J. Hensen, B. M. Weckhuysen, J. Denayer, J. Gascon, F. Kapteijn, *Langmuir* **2011**, *27*, 3970.
[45] P. L. Llewellyn, P. Horcajada, G. Maurin, T. Devic, N. Rosenbach, S. Bourelly, C. Serre, D. Vincent, S. Loera-Serna, Y. Filinchuk, et al., *J. Am. Chem. Soc.* **2009**, *131*, 13002.
[46] M. Pera-Titus, T. Lescouet, S. Aguado, D. Farrusseng, *J. Phys. Chem. C* **2011**, *116*, 9507.
[47] F. X. Coudert, C. Mellot-Draznieks, A. H. Fuchs, A. Boutin, *J. Am. Chem. Soc.* **2009**, *131*, 3442.
[48] P. L. Llewellyn, G. Maurin, T. Devic, S. Loera-Serna, N. Rosenbach, C. Serre, S. Bourelly, P. Horcajada, Y. Filinchuk, G. Ferey, *J. Am. Chem. Soc.* **2008**, *130*, 12808.
[49] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, R. Schmid, *Phys. Status Solidi B* **2013**, *250*, 1128.
[50] J. Bristow, D. Tiana, A. Walsh, *J. Chem. Theory Comput.* **2014**, *10*, 4644.
[51] M. Addicoat, N. Vankova, I. Akter, T. Heine, *J. Chem. Theory Comput.* **2014**, *10*, 880.
[52] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, et al., Gaussian 09 Revision A.02, Gaussian Inc., Wallingford, CT, **2009**.
[53] J. Maple, U. Dinur, A. Hagler, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 5350.
[54] F. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*, 129.
[55] P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbo-Dorca, *J. Chem. Phys.* **2007**, *126*, 144111.

[56] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, M. Waroquier, *J. Chem. Theory Comput.* **2013**, *9*, 2221.
[57] C. Bayly, P. Cieplak, W. Cornell, P. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269.
[58] S. K. Burger, P. W. Ayers, J. Schofield, *J. Comput. Chem.* **2014**, *35*, 1438.
[59] D. A. Kraft, Software package for sequential quadratic programming, Technical Report DFVLR-FB 88-28, DLR German Aerospace Center—Institute for Flight Mechanics, Koln, Germany, **1988**.
[60] E. Bolton, Y. Wang, P. A. Thiessen, S. H. Bryant, *Annu. Rep. Comput. Chem.* **2008**, *4*, 217.
[61] N. O'Boyle, M. Banck, C. James, C. Morley, T. Vandermeersch, G. Hutchison, *J. Cheminform.* **2011**, *3*, 33.
[62] A. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[63] A. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[64] C. Lee, W. Yang, R. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[65] S. Zygmunt, R. Mueller, L. Curtiss, L. Iton, *Theochem* **1998**, *430*, 9.
[66] S. Sousa, P. Fernandes, M. Ramos, *J. Phys. Chem. A* **2007**, *111*, 10439.
[67] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.
[68] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639.
[69] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265.
[70] T. Verstraelen, S. Vandenbrande, M. Chan, F. H. Zadeh, C. González, P. A. Limacher, A. Malek, Horton 1.2.1, Available at: http://theochem.github.com/horton/,**2014**.
[71] T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande, Yaff, yet another force field, Available at: http://molmod.ugent.be/software/.
[72] A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier, V. Van Speybroeck, *J. Chem. Inf. Model.* **2010**, *50*, 1736.
[73] Y. Zhao, D. Truhlar, *Theor. Chem. Acc* **2008**, *120*, 215.
[74] Y. Zhao, D. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.
[75] T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille, G. Ferey, *Chem. Eur. J.* **2004**, *10*, 1373.
[76] H. Li, M. Eddaoudi, M. O'Keeffe, O. Yaghi, *Nature* **1999**, *402*, 276.
[77] Y. Liu, J. Her, A. Dailly, A. Ramirez-Cuesta, D. Neumann, C. Brown, *J. Am. Chem. Soc.* **2008**, *130*, 11813.
[78] P. Yot, Z. Boudene, J. Macia, D. Granier, L. Vanduyfhuys, T. Verstraelen, V. V. Speybroeck, T. Devic, C. Serre, G. Férey, et al., *Chem. Commun. (Camb.)* **2014**, *50*, 9462.
[79] M. Tafipolsky, S. Amirjalayer, R. Schmid, *J. Comput. Chem.* **2007**, *28*, 1169.
[80] G. Piccini, J. Sauer, *J Chem Theory Comput* **2013**, *9*, 5038, doi:10.1021/ct4005504.
[81] A. Walker, B. Civalleri, B. Slater, C. Mellot-Draznieks, F. Corà, C. Zicovich-Wilson, G. Román-Pérez, J. Soler, J. Gale, *Angew. Chem. Int. Ed.* **2010**, *49*, 7501.

Received: 25 November 2014
Revised: 6 January 2015
Accepted: 9 January 2015
Published online on 00 Month 2015