

Supporting information

Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption.

Eugene S. Kadantsev, Peter G. Boyd, Tom D. Daff, and Tom K. Woo*

Centre for Catalysis Research and Innovation, Department of Chemistry, University of Ottawa, 10 Marie Curie, Ottawa, Canada K1N 6N5

* Corresponding Author: twoo@uottawa.ca

Table of Contents:

S1. Charge Equilibration (QEq) Methods

S2. MEPO-QEq Fitting Procedure

S3. GCMC Simulations.

S4. Structure Generation

S5. Inorganic Validation Set.

S6. Alternative Presentation of the Validation Set Results

S1. Charge Equilibration (QEq) Methods. The QEq equations postulate the relationship between the atomic parameters (electronegativity and hardness for each atomic specie) and N_{ATOM} atomic charges. The N_{ATOM} atomic charges can be obtained by solving QEq equations. These equations can be written in the matrix form as

$$\mathbf{G} \begin{pmatrix} \mathbf{Q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\chi} \\ 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{J} & \mathbf{1} \\ \mathbf{1}^\dagger & 0 \end{pmatrix}, \quad (\text{S1.1})$$

where the charges are represented as a $N_{ATOM} \times 1$ column vector $\mathbf{Q}=(Q_1, Q_2, \dots)^\dagger$, λ is the Lagrange multiplier that stems from the charge conservation condition, $\boldsymbol{\chi}=(\chi_1, \chi_2, \dots)^\dagger$ is the electronegativity column vector and \mathbf{J} is the interaction matrix.

In the standard QEq method with periodic boundary conditions, the interaction matrix element between two atoms is given by the lattice summation of the electrostatic interaction between distributed unit charges ρ_A, ρ_B or, in the case of the self-interaction, by the idempotential

$$J_{AB} = \sum_{\vec{T}} (1 - \delta_{A,B} \delta_{\vec{T},0}) \iint \frac{\rho_A(\vec{r}_1; \vec{R}_A) \rho_B(\vec{r}_2; \vec{R}_B + \vec{T})}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 + \delta_{A,B} J_{AA}^0, \quad (\text{S1.2})$$

where the individual unit cells are referenced by lattice vectors \vec{T} , the atoms within unit cell are referenced by vectors \vec{R}_A, \vec{R}_B , and the unit charges are described by a single normalized s-type Slater orbital (STO). For an atom A with an outer valence orbital belonging to n th shell, the STO's is $\rho_A(\vec{r}; \vec{R}_A) = N_n |\vec{r} - \vec{R}_A|^{n-1} \exp\{-\xi_A |\vec{r} - \vec{R}_A|\}$. The decay parameter ξ_A is tabulated, for example, in the Universal Force Field for each atom's type and is inversely proportional to the atom's covalent radius. Throughout this work, we employ the decay parameters which were taken from the OpenBabel software¹. The same decay parameters were used by Sholl *et.al.* to screen 500 real MOFs for CO₂ uptake². Finally, we mention that in the standard QEq method, the QEq equations have to be solved self-consistently if hydrogen is present in the system. The self-consistency is introduced due to the fact that the hydrogen idempotential is charge-dependent³.

We have implemented the standard QEq method in conjunction with periodic boundary conditions in a stand-alone software package. This package was developed in C programming language. In our calculations, we use a simple linear mixing scheme with a mixing parameter of 0.3 and terminate the self-consistent procedure if the change in the charges per atom is less than 10^{-5} . Once the QEq equations are solved and the charges are determined, one can obtain the electrostatic potential at any point in the solid through Ewald summations.

In the case of "extended" QEq method (EQEq)⁴, we use implementation by the authors of Ref. 4 of the Supporting Information. The EQEq method differs from the standard QEq in 3 respects: 1) The QEq energy expression is modified by considering Taylor series expansion around charge-positive metal cations; 2) The self-consistency for hydrogen is dropped and the Coulomb pair interactions are screened by a finite dielectric constant; 3) The form of distributed charges is taken as 1s Gaussians instead of Slater type orbitals (STOs). The two ad-hoc parameters required for the EQEq method were set to the values recommended in Ref 4. The dielectric constant was taken to be $\epsilon_R = 1.67$ and the electron affinity for hydrogen was taken as $I_0 = -2.0$ eV. A 5x5x5 supercell for the reciprocal and real space was used to perform Ewald sums. The charge centers assigned to each metal were as follows: Cu = +2, Zn = +2, V = +4. The ionization potentials and electron affinities for the Taylor series expansion were the experimentally derived values centered on the above ionic species.

S2. MEPO-QEq Fitting Procedure. The MEPO-QEq was developed by varying the atomic electronegativity and hardness in the standard QEq scheme to minimize the "distance" between QEq- and *ab initio* derived quantities. The decay parameters (exponents of STOs) are not varied. Our QEq parameterization (MEPO-QEq) was trained to reproduce *ab initio* electrostatic potentials in a large and diverse set of MOFs. The first step in development of a new parameterization is the generation of the "target" data. In our case, for a given set of MOFs, we perform self-consistent-field DFT^{5,6} calculations of electrostatic potential with PBE exchange-correlation functional^{7,8} and PAW *ab initio* pseudopotentials as implemented in the VASP package⁹⁻¹¹. A plane wave cutoff of 500 eV was used. The energy tolerance of 10^{-5} eV was used to terminate the self-consistent

cycle. The reciprocal space was sampled using Γ -point only. Then, for each MOF, the monopole representation of this *ab initio* electrostatic potential is derived using the REPEAT method¹². The set of all *ab initio* charges and MOF geometries constitute the input ("target") data.

For each MOF, our code accepts the structure's geometry, *ab initio* derived charges ($\{q^{DFT}\}$), and QEq parameters (electronegativity and hardness for each atomic type) as an input and calculates new charges using QEq formalism: $\{q^{QEq}\}$. For each MOF we then calculate the difference between *ab initio* and QEq charges: $\{\Delta q = q^{QEq} - q^{DFT}\}$ and construct the electrostatic potential due to this charge difference: $V(\vec{r})[\Delta q]$. If the norm of this potential is small, then the given QEq parameterization is deemed as a good description of the *ab initio* potential. In practice, the electrostatic potential due to the charge difference is computed on the real space grid made up of N points with space discretization step of $h = 0.25$ Bohr. The points at which the electrostatic potential is computed lie within 3 Å outside of the van der Waals radii of the framework atoms. The calculation of the electrostatic potential is a time-consuming task as it involves an Ewald sum for each grid point. To overcome this computational bottleneck, our code employs shared memory multi-processing. As the electrostatic potential in systems with translational invariance is defined up to an additive constant, the potential is further modified to remove this gauge dependence

$$V(\vec{r}_i)[\Delta q] \rightarrow V(\vec{r}_i)[\Delta q] - \sum V(\vec{r}_i)[\Delta q] / N \quad (\text{S2.1}).$$

The average (per grid point) value of the gauge-modified electrostatic potential due to the *ab initio* and QEq charge difference constitutes the value of the "fitness" for the MOF and QEq parameterization under consideration. The small value of the fitness function corresponds to small potential difference or, in other words, the fitness function is small when *ab initio* and QEq charges yield the same electrostatic potential. The total fitness (per training set of MOFs) is the sum of fitness function values computed for individual MOFs. The QEq parameterization is obtained by minimizing fitness for the whole training set. We employ a number of schemes to optimize the total fitness including full scans for subsets of MOFs, global optimization with custom-built genetic algorithm, and, as a final step, the steepest descent method. In the case of the gradient-based

optimization, the derivatives of the fitness function with respect to the hardness and electronegativity parameters is computed using 2-point central difference scheme with discretization step of 0.1 eV.

S3. GCMC simulations. The details of the GCMC simulations used to calculate the uptake and HOAs of CO₂ performed in this work are as follows. We perform 10 x 10⁶ GCMC production steps following the initial 10⁷ equilibration GCMC steps. The sampling of configurational energies in the GCMC simulations was carried out with a modified version of DL_POLY code with a classical force field. The non-covalent interactions were described by the Lennard-Jones pair potential with parameters taken from the Universal Force Field (UFF) in conjunction with the Lorentz-Berthelot mixing rules for the cross-terms. The non-covalent parameters for CO₂ gas guest molecule were taken from Garcia-Sanchez¹³. 12.5 Å cutoff distance with an isotropic tail correction was used in evaluation of the Lennard-Jones potential. The guest CO₂ gas molecule had a fixed CO bond length of 1.16 Å and partial charges of 0.645 and -0.3325 on carbon and oxygen, respectively. The GCMC simulations were performed at temperature $T=298\text{K}$ and 0.15 bar CO₂ pressure.

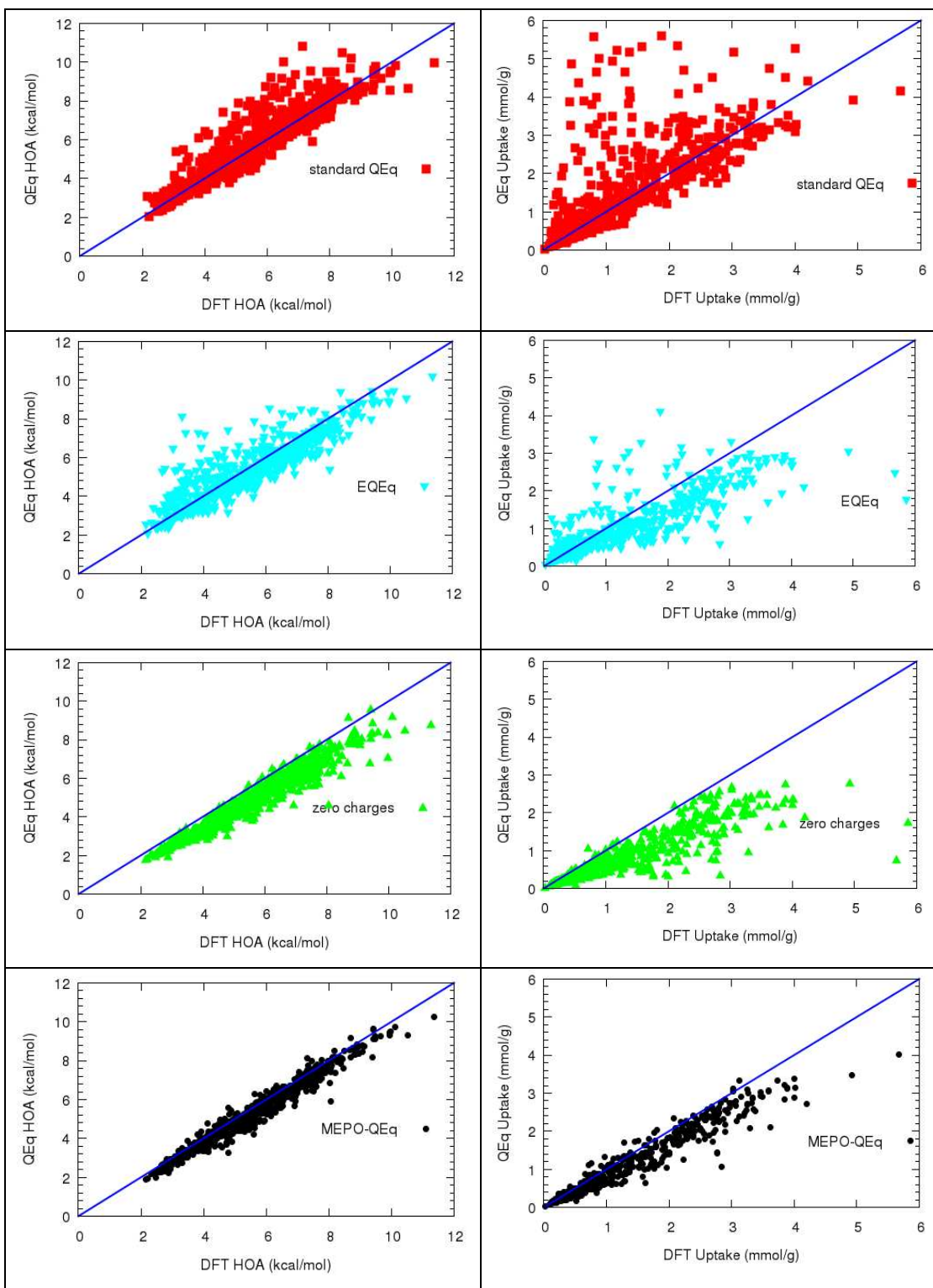
S4. Structure Generation. Each hypothetical MOF in the training and validation sets was constructed using an algorithm similar to the method presented by Wilmer *et al.*¹⁶ using an in-house written code.¹⁹ The structures and lattice parameters were optimized in GULP¹⁷ using the UFF force field¹⁸ where the fractional coordinates of the metal centers and the atoms directly bonded to them were fixed.

S5. Inorganic Validation Set. The additional inorganic SBUs were selected based on their different coordination chemistry from the inorganic SBUs used in the training and validation sets. ZIFs were extracted from Ref. 14 including ZIF-60, ZIF-64, ZIF-68, ZIF-70, ZIF-72, ZIF-73, ZIF-74, and ZIF-79. ZIFs contain Zn (II) ions coordinated tetrahedrally to four imidazolate anions, which was not represented in the training set. Additional ZIFs were created by functionalizing ZIF-74, ZIF-64 and ZIF-79 with F and Cl for a total of 11 ZIFs. Each of the ZIFs were optimized at the DFT level prior to evaluating their CO₂ uptake with the VASP package⁹⁻¹¹, the PBE exchange-correlation

functional^{7, 8} and PAW pseudopotentials using an energy cutoff of 400 eV. The ions were considered in an energy minimum once the forces were below 0.02 eV/Å.

Vanadium MOFs derived from MIL-100¹⁵ were constructed. The inorganic unit of MIL-100 is a discrete SBU coordinating 6 organic SBUs with a trigonal prism-type connectivity. This is different from the V-MOFs included in the training set, which were continuous 1-dimensional chains of V₂O₂ subunits. A total of 16 MOFs were generated using organic building units A, B, C, D, E, F, G, I, L, N, P, S, T, U, V, and X from Figure 1.

S6. Alternative Presentation of the Validation Set Results. The following 8 figures represent results for HOA and CO₂ gas uptake at 0.15 bar, 298 K for the validation set of MOFs obtained from the GCMC simulations using standard QEq charges, zero charges, MEPO-QEq charges, and extended QEq (EQEq) charges as compared to those obtained using DFT-derived charges.



References

1. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*, 33.
2. Haldoupis, E.; Nair, S.; Sholl, D. S. Finding MOFs for Highly Selective CO₂/N₂ Adsorption Using Materials Screening Based on Efficient Assignment of Atomic Point Charges. *Journal of the American Chemical Society* **2012**, *134*, 4313-4323.
3. Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991**, *95*, 3358-3363.
4. Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An Extended Charge Equilibration Method. *The Journal of Physical Chemistry Letters* **2012**, *3*, 2506-2511.
5. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864-B871.
6. Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133-A1138.
7. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865-3868.
8. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Physical Review Letters* **1997**, *78*, 1396-1396.
9. Kresse, G.; Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **1996**, *54*, 11169-11186.
10. Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **1993**, *47*, 558-561.
11. Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium. *Physical Review B* **1994**, *49*, 14251-14269.
12. Campana, C.; Mussard, B.; Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *Journal of Chemical Theory and Computation* **2009**, *5*, 2866-2878.
13. Garcia-Sanchez, A.; Ania, C. O.; Parra, J. B.; Dubbeldam, D.; Vlugt, T. J. H.; Krishna, R.; Calero, S. Transferable Force Field for Carbon Dioxide Adsorption in Zeolites. *The Journal of Physical Chemistry C* **2009**, *113*, 8814-8820.
14. Banerjee, R.; Phan, A.; Wang, B.; Knobler, C.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. High-Throughput Synthesis of Zeolitic Imidazolate Frameworks and Application to CO₂ Capture. *Science* **2008**, *319*, 939-943.
15. Lieb, A.; Leclerc, H.; Devic, T.; Serre, C.; Margiolaki, I.; Mahjoubi, F.; Lee, J. S.; Vimont, A.; Daturi, M.; Chang, J.-S. MIL-100(V) - A mesoporous vanadium metal organic framework with accessible metal sites. *Microporous and Mesoporous Materials* **2012**, *157*, 18-23.
16. Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nat Chem* **2011**, *4*, 83-89.

17. Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *Journal of the Chemical Society, Faraday Transactions* **1997**, *93*, 629-637.
18. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.
19. Boyd, P. G.; Williams, L. A.; Woo, T. K. Generation of a Diverse Chemical and Structural database of MOFs *in silico. in preparation.*