JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

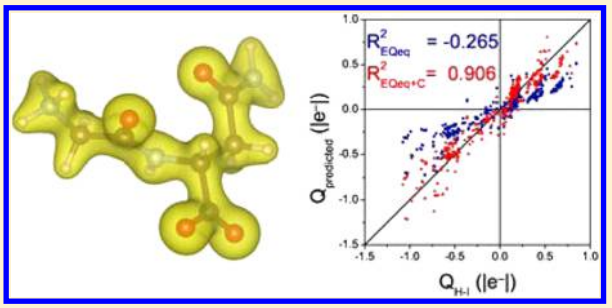
EQeq+C: An Empirical Bond-Order-Corrected Extended Charge Equilibration Method

Geoffrey C. Martin-Noble, David Reilley, Luis M. Rivas, Matthew D. Smith,[†] and Joshua Schrier*

Department of Chemistry, Haverford College, Haverford, Pennsylvania 19041, United States

Supporting Information

ABSTRACT: The extended charge equilibration (EQeq) scheme computes atomic partial charges using the experimentally measured ionization potentials and electron affinities of atoms. However, EQeq erroneously predicts constant (environment independent) charges for high-oxidation-state transition metals in amine-templated metal oxide (ATMO) compounds, contrary to the variation observed in iterative Hirshfeld (Hirshfeld-I) charges, bond-valence sum calculations, and formal oxidation state calculations. To fix this problem, we present a simple, noniterative empirical pairwise correction based on the Pauling bond-order/distance relationship, which we denote EQeq+C. We parametrized the corrections to reproduce the Hirshfeld-I charges of ATMO compounds and REPEAT charges of metal organic framework (MOF) compounds. The EQeq+C correction fixes the metal charge problem and significantly improves the partial atomic charges compared to EQeq. We demonstrate the transferability of the parametrization by applying it to a set of unrelated dipeptides. After an initial parametrization, the EQeq+C correction requires minimal computational overhead, making it suitable for treating large unit cell solids and performing large-scale computational materials screening.



1. INTRODUCTION

Understanding the charge distribution in large molecules and solids is important in fields as diverse as crystal design,[1] gas storage/separation,[2] pharmaceutical design,[3] and prediction of molecular solubility.[4] A common simplification is to describe the electron distribution as a set of atom-centered partial charges; this provides both a simple conceptual framework and is useful for classical force-field simulations. However, assigning atomic partial charges is computationally challenging. The most rigorous approaches start with the electron density obtained from electronic-structure calculations and then attempt to reproduce the electrostatic potential (e.g., REPEAT[5]) or partition the charge density (e.g., Hirshfeld-I[6]) or try to do both simultaneously (e.g., DDEC[7]). The high computational cost of electronic-structure calculations makes the preliminary step intractable for very large systems or for screening large numbers of hypothetical compounds. Thus, it is desirable to have approximate methods that assign partial charges from a molecular geometry without performing detailed quantum mechanical calculations.

The simplest approximation strategy for assigning partial atomic charges is to use the charges for similar types of chemical systems. One approach is to assign the charges based on the type of atom and the number and types of atoms immediately covalently bonded to it (e.g., the connectivity-based atomic charges (CBAC) approach for metal organic frameworks (MOFs)[8] and covalent organic frameworks (COFs)[9] of Zhong and co-workers). A more sophisticated strategy is to use machine learning to assign partial charges using the geometric similarity to a database of known molecules, geometries, and charges.[10] However, these approaches require a trade-off between either having a very large number of empirical parameters or having a limited ability to describe the partial atomic charge variations that occur in diverse environments.

A compromise between the *ab initio* and completely empirical approach are charge equilibration techniques, which can broadly be divided into split charge equilibration (SQE) and electronegativity equalization (EEM) methods; Verstraelen et al. provide a critical overview of both of these approaches.[11] In SQE, each type of covalent chemical bond results in a charge transfer, and the atomic partial charge is the sum of the credits and debits for each bond in which the atom participates. SQE has three disadvantages: (i) empirical parameters are required for each type of chemical bond, (ii) bond types must be assigned, requiring an additional set of information besides the geometry, and (iii) only nearest-neighbor interactions occur between covalently bonded atoms, neglecting long-range interaction between different parts of the system. In EEM methods, such as the Qeq method of Rappé and Goddard,[12] the total energy of the molecular system is expressed as a function of the charge on each atom, specifically as per-atom energies and electrostatic interactions between the atoms, and is minimized to obtain the partial charges of the atoms. EEM methods have three disadvantages: (i) empirical parameters are

Received: January 15, 2015
Published: June 11, 2015

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

typically required for the atom energies as a function of charge and thus, like SQE, a large number of parameters must be fitted, (ii) iterative solutions are typically required to obtain the charges on each site, adding an additional computational cost, and (iii) only the atomic charging energies and electrostatics are taken into account, neglecting covalent bonding effects. Wilmer et al. described an improved EEM scheme, denoted EQeq, which addresses the first two of these disadvantages.[13] In EQeq, the per-atom energies are obtained from experimental ionization energies and electron affinities, leaving only two empirical parameters: an empirical dielectric constant and a modified per-atom energy for hydrogen. Additionally, the total energy is minimized noniteratively, further reducing the computational cost. Wilmer et al. showed that EQeq and REPEAT partial charges yielded comparable predictions for gas adsorption on MOFs using grand canonical Monte Carlo (GCMC) calculations.

We began by comparing EQeq charges to the Hirshfeld-I charges of amine-templated metal oxide (ATMO) compounds studied in our past work.[14−18] Despite the very different chemical environment of our densely packed solids (versus the porous MOFs for which EQeq was first developed), modifying the few empirical parameters in EQeq leads to only very minor improvements in the agreement with the Hirshfeld-I charges, suggesting that it is not necessary to refit those parameters to apply EQeq to different classes of compounds. However, EQeq erroneously predicts constant (environment independent) charges on high-oxidation-state transition metals, contrary to the variations observed in the Hirshfeld-I charges, bond-valence sum calculations, and formal oxidation state calculations.

In this article, we show how to fix this problem with a simple, noniterative empirical correction based on the Pauling bond-order, denoted EQeq+C. Conceptually, EQeq, like other EEM methods, considers the interaction between atoms to be purely electrostatic. Our empirical correction adds a covalent contribution, analogous to SQE methods, but uses only the geometric distance between atoms rather than requiring explicit bond-type assignments. Using existing tables of atomic covalent radii, our correction requires only a few additional parameters to reproduce either Hirshfeld-I or REPEAT charges. EQeq+C improves the agreement with these charges for ATMOs and MOFs, and we demonstrate that our correction is transferable by comparing to Hirshfeld-I charges for a variety of dipeptides using our ATMO parametrization. The EQeq+C correction involves a negligible amount of computer time beyond the initial EQeq calculation, making it suitable for large systems and computational screening projects.

2. THEORY

EQeq is described by Wilmer et al. in ref 13, and the Supporting Information of that paper contains an extended pedagogical derivation, so we summarize it only briefly here. The energy of an isolated (gas phase) atom k, as a function of charge on that atom, Q_k, is assumed to be a second-order Taylor expansion about a reference partial charge, Q_k^*, given by

$$E_{Ak}(Q_k) = E_{Ak}(Q_k^*) + \chi_{Q_k^*}(Q_k - Q_k^*) + \frac{1}{2}J_{Q_k^*}(Q_k - Q_k^*)^2 \quad (1)$$

The coefficients of this Taylor expansion are the Mulliken electronegativity, $\chi_{Q_k^*}$, and the chemical hardness, $J_{Q_k^*}$, where the subscripts indicate that these properties are functions of both the reference partial charge and the particular atom being considered. In EQeq, the electronegativities and hardness are obtained from the experimental ionization potentials and electron affinities of the atoms, with the exception of hydrogen, where it was found to be necessary to empirically set the first ionization energy to −2 eV. For most elements, the reference partial charge is set to $Q_k^* = 0$, but for transition metals and other atoms having large positive charges, Q_k^* is set to an expected reference oxidation state. The total energy of the system is the sum of the above isolated atom energies and the Coulombic interactions between the atoms

$$E_{sys}(Q_1, Q_2, ..., Q_N) = \sum_{k=1}^{N} (E_{Ak}(Q_k) + E_{coul,k}) \quad (2)$$

where $E_{coul,k}$ is the electrostatic interaction of atom k with all of the other $N - 1$ atoms in the system. This can be evaluated directly for finite systems or evaluated as an Ewald sum for periodic systems. Wilmer et al. also include a pairwise damping term to prevent infinite charge separation when the atom pairs are brought very close together. An empirical dielectric constant, $\varepsilon_r = 1.67$, is used in the Coulomb energy calculation to reduce the energy of charge separation. The partial charges, Q_k, are obtained by minimizing E_{sys}, relative to the constraint of conserved total charge.

The qualitative relationship between (increasing) bond order and (decreasing) bond length is taught in introductory chemistry classes. Theorists have sought many ways to quantify this relationship.[19] Of these, one of the earliest, simplest, and most widely used is Linus Pauling's "chemist's bond order"[20]

$$n = \exp[-(r - R_0)/b] \quad (3)$$

where n is the bond order, R_0 is the equilibrium bond length of a single bond, r is the actual observed bond length for which we are trying to obtain the bond order, and b is an empirical parameter. In Pauling's model, and in subsequent data fitting by Lendvay,[19] R_0 and b are constants specific to each pair of elements. The same functional form is also used to determine the bond valence in the bond valence sum (BVS) model of Brown.[21]

Pauling used this bond-order relationship to determine the atomic radii of metals.[20] Alternatively, using a known set of radii, one can compute the bond order and in turn make a SQE-type charge correction. This is the approach taken by Marenich et al. in their CM5 charge correction model,[22] where an empirical correction is added to (noniterative) Hirshfeld atomic partial charges, Q_k^H, to improve the agreement with experimental dipole and high-multipole moments of molecules. The CM5 correction has the form

$$Q_k^{CM5} = Q_k^H + \sum_{k' \neq k} T_{kk'}B_{kk'} \quad (4)$$

where $T_{kk'}$ describes the relative charge transfer between atoms k and k' (requiring $T_{kk'} = -T_{k'k}$ and $T_{kk} = 0$) and $B_{kk'}$ is a Pauling bond-order like interactions between the two atoms

$$B_{kk'} = \exp[-\alpha(r_{kk'} - R_{Z_k} + R_{Z_{k'}})] \quad (5)$$

where Z_k is the atomic number of atom k, R_{Z_k} is the covalent radius of an atom with that atomic number, and $r_{kk'}$ is the distance between atoms k and k'. Note how this correction has the same type of functional form as SQE, with $T_{kk'}$ playing the

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

role of the charge transfer and $B_{kk'}$ serving as a more general way to include bonds (rather than only present/absent as in SQE schemes). In CM5, Marenich et al. have two different ways to define the $T_{kk'}$; for cases involving H, C, N, O, they define pair-specific parameters

$$T_{kk'} = D_{Z_k Z_{k'}} \tag{6}$$

and for other cases, these are defined in terms of differences of single-atom parameters

$$T_{kk'} = D_{Z_k} - D_{Z_{k'}} \tag{7}$$

which limits the number of free parameters.

For our EQeq+C scheme, we propose applying a CM5-like correction to the EQeq charges

$$Q_k^{\mathrm{EQeq+C}} = Q_k^{\mathrm{EQeq}} + \sum_{k' \neq k} T_{kk'} B_{kk'} \tag{8}$$

with all of the parameters described according to eq 7. This introduces a minimal number of new parameters. If we use existing sets of the covalent radii, R_{Z_k}, then the only new parameters are a single value of α and a set of D_{Z_k} equal to the number of elements minus one. (Since the charge transfer is relative, we chose $D_C = 0$ as our reference value, as was done in CM5.) We hypothesize that these parameters are transferable to different chemical environments. However, since the goal of the parametrization is to reproduce a particular charge assignment scheme, different charge assignment schemes (e.g., REPEAT, H-I) may have different parametrizations. For the ATMO/H-I parametrization, we have considered only H, C (set to zero), N, O, V, Se, and Te. For the MOF/REPEAT parametrization, we consider only H, C (set to zero), N, O, Mg, V, Co, Ni, Cu, Zn, and Pd. In principle, the whole periodic table can be parametrized, given a set of partial charges derived from electronic structure calculations on an appropriate set of reference compounds.

3. COMPUTATIONAL METHODS

3.1. Electronic Structure Calculations. Solid-state density functional theory (DFT) electronic structure calculations were performed using ABINIT, v6.4.1 and v6.6.12,[23,24] with the Perdew–Burke–Ernzerhof generalized gradient approximation (PBE-GGA) exchange-correlation functional,[25,26] norm-conserving Troullier–Martins pseudopotentials,[27] planewave basis sets with an energy cutoffs of 25 hartree for the vanadium tellurite ATMOs, vanadium oxide ATMOs, and dipeptides and 35 hartree for the vanadium selenite ATMOs. A 6 × 6 × 6 and 4 × 4 × 4 Monkhorst–Pack grid was used to sample the Brillouin zone for ATMOs and dipeptides, respectively. Experimental crystal structures were used without further geometry optimization, except for the BEQJEN and XUDVOH dipeptides (*vide infra*). Complete lists including references to the source crystal structures are shown in Tables 1 and 3. REPEAT charges for the MOFs in Table 2 were derived from ultrasoft pseudopotential density functional theory calculations, taken from the Supporting Information of Wilmer et al.[13]

3.2. Geometry Optimizations. Hydrogen atom positions are poorly described by X-ray crystallography due to the small number of electrons and large distortions of the electron density away from the hydrogen nucleus. Consequently, the hydrogen positions reported experimentally are often the result of various empirical corrections, discussed by Lusi and Barbour.[28] In the course of our calculations described below, we found that the BEQJEN and XUDVOH dipeptide crystals had poor agreement between the Hirshfeld-I charges and the EQeq+C charges. Suspecting that this was due to incorrect hydrogen atom positions, we performed geometry optimizations of all of the hydrogen atoms in these structures, leaving the heavy atoms fixed. Geometry optimizations were performed using VASP 4.6.35[29−32] to perform PBE-GGA projector augmented wave (PAW)[33,34] calculations. A 4 × 4 × 4 Monkhorst–Pack grid was used to sample the Brillouin zone. The forces on the hydrogen atoms were converged to below 0.01 eV/Å on each atom. The optimized geometries are contained in the Supporting Information.

3.3. Iterative-Hirshfeld Partial Atomic Charges. Iterative-Hirshfeld (Hirshfeld-I or H-I)[6] partial atomic charge determinations were performed on the self-consistent valence electron density obtained from the ABINIT calculations in conjunction with promolecular all-electronic atomic charge densities generated using the HF96 Hartree–Fock code,[35] as described in our previous work.[14] A benchmarking study by Vanpoucke et al.,[36] as well as our own calculations, indicate that H-I partial charges converge with increasing planewave basis set, similar to the behavior seen with atomic orbital basis sets.[6] Verstraelen et al.[37] described how the H-I charges are too negative for oxygen atoms in silicates, which they attribute to unphysical dianion states that are used to construct the weight functions; consequently, they propose that H-I charges $< -1|e^-|$ are suspect. Although we did not artificially impose this constraint on our H-I calculation, we found this not to be a problem. None of the ATMOs or dipeptides that we computed H-I charges for in this study (or in our previous work[14−17]) have H-I charges less than −1.08. In the ATMO data set, there are 24 nitrogen and 6 oxygen atoms with charges between −1 and −1.05 distributed across 7 different compounds. In the dipeptide data set, there are 16 nitrogen and 4 oxygen atoms between −1 and −1.08 in 5 different compounds. These relatively small incursions are not as extreme as those observed by Verstraelen et al.[37] and hence we have used them as-is. Miliordos and Harrison described how H-I partial charges for transition metal-containing diatomic and triatomic molecules depend on the specific electronic configuration (term symbol) used for the promolecular atomic charge density.[38] Here (and in our previous work[14−17]) we used the orbital occupancies and corresponding ground-state term symbols tabulated by the U.S. National Institute of Standards and Technology (NIST).[39] The only exception is for the vanadium cation, where we used a $[Ar]4s^1 3d^3$ configuration rather than the ground-state $[Ar]3d^4$ configuration because the latter gives rise to an unphysical negative vanadium partial atomic charge in the H-I calculations.

3.4. EQeq and EQeq+C Partial Atomic Charges. EQeq partial atomic charges were computed for the compounds listed in Tables 1–3 using the EQeq C++ program developed by Wilmer et al. available in the Supporting Information of their paper.[13] We observed that small numbers of unit cells in the periodic image (e.g., 3 × 3 × 3) resulted in EQeq charges on symmetry-identical atoms in the unit cell having different charges. To ensure convergence of the charges with respect to symmetry, we performed the EQeq calculations using periodic images contained in a system of 41 × 41 × 41 unit cells (for both real and reciprocal space).

The "+C" correction calculation (eq 8) was implemented in Python, v.2.7.3, and source code is available in the Supporting Information of this article. The atomic covalent radii are from

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

the most recent CRC Handbook[40] and are shown in Table 4. Because $B_{kk'}$ falls off exponentially with increasing distance, these terms are evaluated in real-space, using one unit cell in each direction. The D_Z parameters were determined by minimizing the least-squares difference from the reference charges, Q_k^{ref} (H-I charges in the case of the ATMOs and REPEAT charges in the case of the MOFs)

$$\chi = \sum_k (Q_k^{\mathrm{EQeq+C}} - Q_k^{\mathrm{ref}})^2 \tag{9}$$

The parametrization of eq 8 can be made into a linear optimization problem by grouping the terms for each atom into sets of types of atoms. The goal of the parametrization is to determine the EQeq+C parameters for each atom, k, so that the first equality

$$Q_k^{\mathrm{H\text{-}I}} = Q_k^{\mathrm{EQeq+C}} = Q_k^{\mathrm{EQeq}} + \sum_{k'\neq k} B_{kk'}T_{kk'} \tag{10}$$

is obeyed as closely as possible. This can be achieved by finding values of the parameters D_{Z_k} that satisfy

$$Q_k^{\mathrm{H\text{-}I}} - Q_k^{\mathrm{EQeq}} = \sum_{k'\neq k} (D_{Z_k} - D_{Z_{k'}})B_{kk'} \tag{11}$$

Letting $y_k \equiv Q_k^{\mathrm{H\text{-}I}} - Q_k^{\mathrm{EQeq}}$ and expanding the sum on the right-hand side into two separate terms yields

$$y_k = \sum_{k'\neq k} D_{Z_k}B_{kk'} - \sum_{k'\neq k} D_{Z_{k'}}B_{kk'} \tag{12}$$

Since D_{Z_k} is independent of the identity of atom k', it can be moved outside the first sum, yielding

$$y_k = D_{Z_k}\sum_{k'\neq k} B_{kk'} - \sum_{k'\neq k} D_{Z_{k'}}B_{kk'} \tag{13}$$

Let $Z(k) \equiv Z_k$. Then, the atoms can be separated based on their element types

$$y_k = D_{Z(k)}\sum_{k'\neq k} B_{kk'} - \sum_Z D_Z \sum_{k'\neq k, Z(k')=Z} B_{kk'} \tag{14}$$

The sums of $B_{kk'}$ are based only on geometry and the value of α, which remain fixed throughout the optimization. Therefore, for each atom, k, there is a linear equation for y_k in terms of the element specific correction parameters, D_Z, where there is a D_Z value for each element present in the compound. Again, as all of these parameters are relative, we set $D_C \equiv 0$. From here, we can solve for the optimal D_Z parameters for the other elements by finding the least-squares solution to the system of n linear equations, where n is the number of atoms in the training set of compounds. To determine the optimal α parameter value, we performed this linear optimization over a range of potential values for α and selected the parameters that gave the lowest R^2 value. (Using fixed values of α avoids having to perform a nonlinear optimization.) The parametrization code, written in Python, is available in the Supporting Information. The final optimized parameters are in Table 4.

4. RESULTS AND DISCUSSION

4.1. Application of EQeq to Amine-Templated Metal Oxides. We applied the EQeq method to three different sets of extended structures: amine-templated metal oxides (ATMO), metal organic frameworks (MOF), and dipeptides crystals. The specific crystal structures investigated in each system are provided in Tables 1−3.

Table 1. Amine-Templated Metal Oxide (ATMO) Compounds Used in This Study and Mean Absolute Deviations (MAD) from H-I Charges[a]

compound	EQeq MAD	EQeq+C MAD	ref
$[1\text{-mpipH}_2][(VO)_2(C_2O_4)(SeO_3)_2]\cdot 2H_2O$*	0.266	0.068	18
$[2,5\text{-dmpipH}_2][(VO)_2(C_2O_4)(SeO_3)_2]\cdot 2H_2O$	0.227	0.083	18
$[\text{bappH}_2][VO_3]_2\cdot 2H_2O$*	0.185	0.073	16
$[\text{S-3-apipdH}_2][VO_3]_2$	0.181	0.071	16
$[\text{R-3-apipdH}_2][VO_3]_2$	0.185	0.068	16
$[\text{trenH}_3][VO_3]_3\cdot H_2O$	0.213	0.094	41
$[\text{R-3-aqnH}_2][V_2Te_2O_{10}]$	0.260	0.082	14
$[\text{S-3-aqnH}_2][V_2Te_2O_{10}]$	0.257	0.084	14
$[2,5\text{-dmpipH}_2][V_2Te_2O_{10}]$	0.217	0.102	15
$[1,6\text{-dahH}_2][V_2Te_2O_{10}]$	0.270	0.065	17
$[\text{pipH}_2][V_2Te_2O_{10}]$	0.251	0.046	42
$[\text{tmedH}_2][V_2Te_2O_{10}]$*	0.245	0.084	17
$[1,4\text{-dabH}_2][V_2Te_2O_{10}]$	0.289	0.082	17
$[2\text{-mpipH}_2][VO(SeO_3)(HSeO_3)]_2\cdot 2H_2O$*	0.315	0.066	43
$[2,5\text{-dmpipH}_2][VO(SeO_3)(HSeO_3)]_2\cdot 2H_2O$	0.301	0.059	43
$[\text{S-2-mpipH}_2][(VO)_3(SeO_3)_2(HSeO_3)_4]$	0.356	0.081	18
$[\text{R-2-mpipH}_2][(VO)_3(SeO_3)_2(HSeO_3)_4]$	0.356	0.089	18

[a]Compounds marked with (*) were used for cross-validation shown in Figure 6.

Table 2. Metal Organic Framework (MOF) Compounds Used in This Study and Mean Absolute Deviation (MAD) from REPEAT Charges[a]

compound	metal	EQeq MAD	EQeq+C MAD	ref
Co-MOF-74	Co	0.137	0.061	44
HKUST-1	Cu	0.142	0.028	45
IRMOF-1*	Zn	0.162	0.070	46
IRMOF-3	Zn	0.238	0.117	47
Mg-MOF-74	Mg	0.193	0.069	48
MIL-47	V	0.112	0.054	49
Ni-MOF-74	Ni	0.146	0.056	50
Pd-2-pymo	Pd	0.241	0.083	51
UMCM-150	Cu	0.109	0.056	52
UMCM-150-N2*	Cu	0.151	0.103	53
ZIF-8*	Zn	0.134	0.206	54
Zn-MOF-74	Zn	0.141	0.084	55

[a]Compounds marked with (*) were used for cross-validation shown in Figure 6.

Wilmer et al.[13] used a 5 × 5 × 5 supercell in both real and reciprocal spaces in their study of MOFs, but we wanted to make sure that this fully converges the calculation, since truncating the Ewald summation of $E_{\mathrm{coul},k}$ can cause two possible errors in the EQeq charges. First, Q_k^{EQeq} should approach an asymptotic value with increasing numbers of real-space and reciprocal-space cells. Second, symmetrically identical positions in the crystal should have the same Q_k^{EQeq}, but it is possible for them to not be the same due to truncation of the Ewald sum. Figure 1 shows the convergence of the MAD of the ATMO EQeq charges with respect to the 41 × 41 × 41 supercell, which itself is converged with respect to symmetry. The MAD values of the EQeq charges asymptotically approach

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

Table 3. Dipeptides Used in This Study and Mean Absolute Deviation (MAD) from H-I Charges[a]

compound	description	EQeq MAD	EQeq+C MAD	ref
APALTY	F-Y	0.155	0.084	56
BEQJEN	W-V	0.179	0.102	57
BEQJEN (opt)	W-V	0.183	0.061	57
BUVKEJ	E-E-	0.187	0.122	58
CGLYTP	G-W-	0.156	0.062	59
ETONIK	I-F	0.184	0.069	60
EYIVAJ	S-V	0.181	0.113	61
GLLASP	G-N	0.275	0.114	62
GLYTRE03	G-T	0.235	0.066	63
MAPKOE	T-A	0.195	0.113	64
TEKNAY	H-A	0.208	0.092	65
WIRYEB	V-A	0.181	0.119	66
XUDVOH	A-V	0.205	0.156	67
XUDVOH (opt)	A-V	0.203	0.085	67

[a]Entries marked with (opt) had hydrogen positions optimized, as described in Sections 3.2 and 4.4.

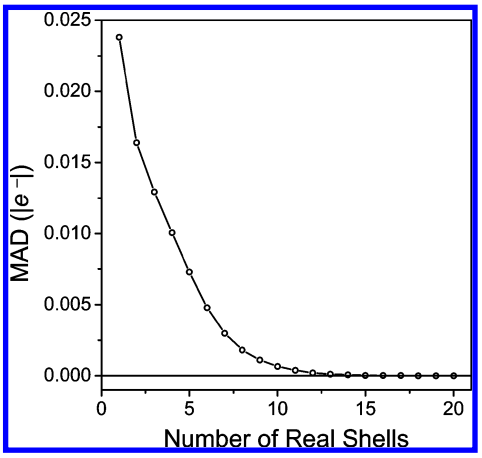


Figure 1. Convergence of the mean absolute deviation (MAD) of EQeq charges ($|e^-|$) versus the number of real shells used in the Ewald summation of the Coulombic interaction energy term for ATMOs, relative to the fully converged 41 × 41 × 41 case.

zero and are within $0.001|e^-|$ of the 41 × 41 × 41 supercell case for supercells larger than 8 × 8 × 8.

Is the dielectric screening parameter, ε_r, introduced as an empirical parameter for describing MOFs in ref 13 equally suitable for the ATMO compounds? Figure 2 shows the R^2 between the H-I and EQeq charges for the ATMO compounds. The maximum occurs at $\varepsilon_r = 1.65$, which is slightly less than the $\varepsilon_r = 1.67$ found to be optimal in ref 13. However, the difference between these two choices is small. This supports the general transferability of the EQeq method, but it is nevertheless surprising because the ATMO compounds are densely packed metal oxides that are very different from the porous MOFs. For compatibility with previous work, we have chosen to use $\varepsilon_r = 1.67$ for all of the subsequent calculations.

How well does EQeq reproduce the ATMO/H-I charges? The blue points in Figure 3a show the EQeq charges; the diagonal line indicates the bisectrix ($Q_k^{\text{H-I}} = Q_k^{\text{EQeq}}$). The EQeq and H-I charges are correlated ($R^2 = 0.809$), but several systematic discrepancies can be seen. The charges for the V, Se, and Te atoms deviate the most, and as shown in Figure 3b, the EQeq charges (circles) are both much larger than the H-I charges and essentially constant. In contrast, the H-I charges for

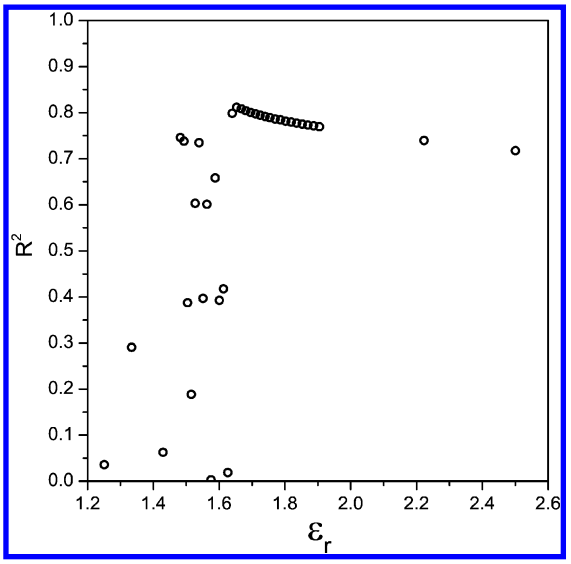


Figure 2. R^2 coefficient of EQeq to the ATMO/H-I charges, as a function of ε_r. The maximum (indicating the best fit) is at $\varepsilon_r = 1.65$.

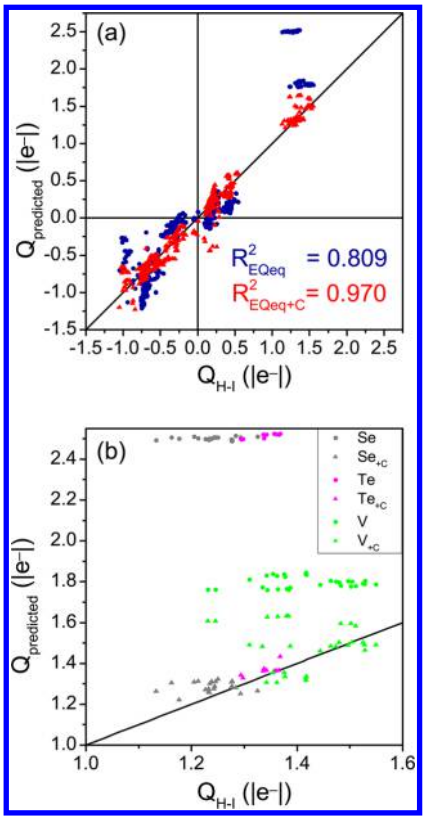


Figure 3. (a) EQeq (blue circles) and EQeq+C (red triangles) charges versus H-I charges for ATMOs, using $\varepsilon_r = 1.67$, $\alpha = 2.10$ Å^{-1}. The diagonal line indicates the bisectrix. (b) The same data, plotted to emphasize the Se (gray), Te (pink), and V (green) charges; circles and triangles again indicate EQeq and EQeq+C, respectively.

these elements vary significantly, reflecting the diverse chemical environments in which they are present. For example, the vanadium atoms occur in two oxidation states (IV and V), with three (tetrahedral, square pyramidal, and distorted octahedral) distinct coordination geometries, and coordinated to two possible ligands (O^{2-} and $C_2O_4^{2-}$). BVS calculations also yield large (>1 valence unit) differences in the total valence bonding

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

of vanadium atoms in these different sites. Therefore, the EQeq prediction that vanadium atoms exhibit essentially identical partial atomic charge is unreasonable and suggests that EQeq does not treat metallic elements with high (≥4) formal oxidation states such as these correctly.

4.2. EQeq+C Parameterization. Optimization of the D_Z parameters in eq 8 (with fixed α) to satisfy eq 9 was performed separately for ATMOs with H-I charges (denoted EQeq+C/ATMO/H-I) and MOFs with REPEAT charges (denoted EQeq+C/MOF/REPEAT). We performed optimizations with multiple values of α in increments of 0.1 Å^{-1}, between 1.0 and 5.0 Å^{-1}. We performed additional, finer scans with increments of 0.01 Å^{-1} within 0.1 Å^{-1} of the maxima from these initial scans (α = 2.00 to 2.20 Å^{-1} for the ATMOs and 2.60 to 2.80 Å^{-1} for the MOFs). The results for the ATMO/H-I and MOF/REPEAT data sets are shown in Figure 4. Both vary smoothly

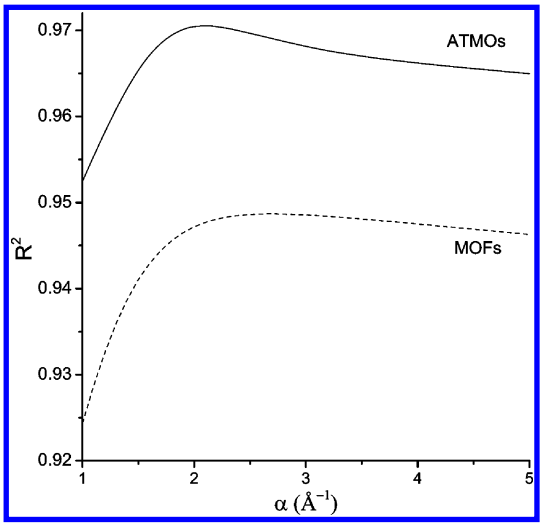


Figure 4. Coefficient of determination (R^2) values of the EQeq+C method as a function of the α parameter for ATMO (solid) and MOF (dashed) data sets.

and have a single maximum, with maximum values of R^2 at α = 2.10 and 2.69 Å^{-1}, respectively. Consequently, we used these two different parameters for the different correction schemes described below. However, we note that the scales for the R^2 values exaggerate the dependence on α, and the use of a common intermediate value of α would give essentially the same results. Finally, we note that these values are comparable to, but slightly smaller than, the α = 2.833 Å^{-1} value used by Marenich et al. in their CM5 parametrization[22] and the value of α = 1/0.37 = 2.703 Å^{-1} commonly used in BVS calculations.[21]

The evolution of the D_Z parameters with respect to α for both the ATMO and MOF parametrizations is shown in Figure 5. The magnitude of the D_Z parameters increases with increasing α until $\alpha \approx 2.5$, and it then slowly decreases in magnitude. Since increasing α makes the $B_{kk'}$ term smaller, the increased D_{Z_k} term compensates to make the product have the same magnitude. Since the ATMO/H-I and MOF/REPEAT training sets are based on different charge assignment schemes, it is unsurprising that the common parameters D_H, D_N, D_O, and D_V in Figure 5, panels a and b, are different. However, their relative order remains the same, with $D_H > D_N > D_O > D_V$, suggesting that the parametrization reflects underlying physical structures that are independent of the types of compounds or the charge assignment scheme.

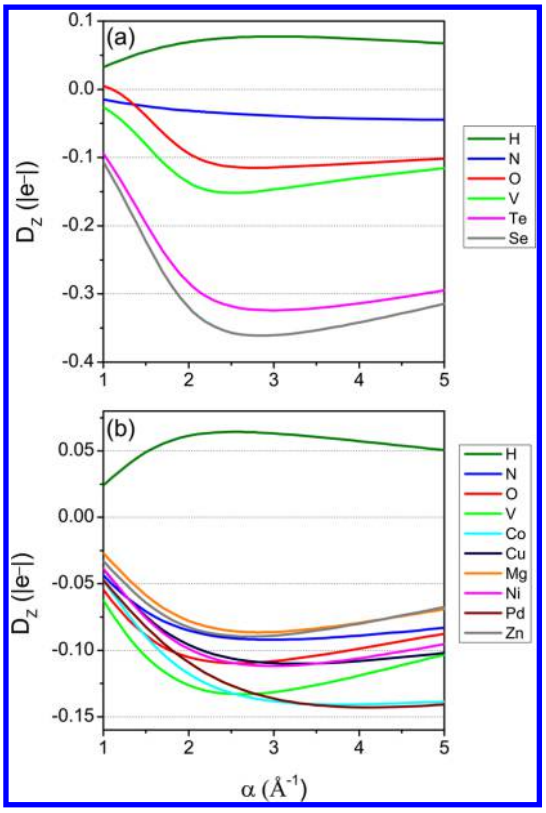


Figure 5. Dependence of the D_Z parameters on α. (a) ATMO/H-I; (b) MOF/REPEAT.

To ensure that we have not overfit the data, we cross-validated the EQeq+C/ATMO/H-I and MOF/REPEAT D_Z parameters. Four compounds from the ATMO data set and three from the MOF data set, marked by (*) in Tables 1 and 2, were removed to reduce the size of the ATMO and MOF training sets by 23.5 and 25.0%, respectively. The D_Z parameters were then reoptimized using the reduced training set and used to calculate EQeq+C charges for the excluded atoms. The resulting charges were compared to the EQeq+C charges based on the complete data set parametrization, shown in Figure 6. The cross-validated charges are in extremely good agreement (R^2 > 0.98) for both ATMO/H-I and MOF/REPEAT data sets (Table 4). This demonstrates that the EQeq+C parametrization is stable to the removal of >20% of the training set and therefore that we have not overfit the data.

4.3. Application of EQeq+C to ATMOs and MOFs. How well does EQeq+C reproduce the ATMO/H-I charges? The blue points in Figure 3a indicate the EQeq+C charges. The EQeq+C/ATMO/H-I charges are more correlated to the H-I charges than EQeq, indicated by the increase of the R^2 coefficient from 0.809 to 0.970. The mean absolute deviation (MAD) of the EQeq and EQeq+C charges from the H-I charges for each ATMO compound are shown in Table 1. The MAD of the EQeq+C charges is less than half of the EQeq charge MAD for every ATMO compound in our data set, and seven compounds have EQeq+C charges with a quarter of the MAD of the EQeq charges. Overall, the +C correction reduces the MAD by 71% on the entire data set. The 3D framework vanadium selenite compounds had the largest improvement, which is due to the overall improvement in the V, Se, (and Te) charges by EQeq+C. As seen in Figure 3b, EQeq+C (filled symbols) does not suffer from the problem of constant metal

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

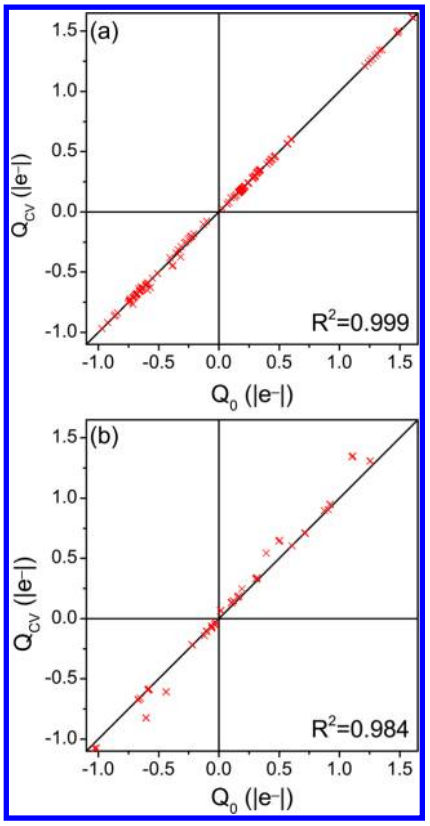


Figure 6. Cross-validated EQeq+C charges versus original EQeq+C charges for a test set of compounds. (a) ATMO/H-I data set with ε_r = 1.67, α = 2.10 Å^{-1}; (b) MOF/REPEAT data set with ε_r = 1.67, α = 2.69 Å^{-1}.

Table 4. Parameters Used for the EQeq+C Correction[a]

element	R_Z/Å	D_Z/|e⁻| (ATMO/H-I)	D_Z/|e⁻| (MOF/REPEAT)
H	0.32	0.07107	0.06422
C	0.75	0	0
N	0.71	−0.03228	−0.09167
O	0.64	−0.10059	−0.10957
Mg	1.40		−0.08637
V	1.44	−0.14276	−0.13300
Co	1.18		−0.13518
Ni	1.17		−0.11097
Cu	1.22		−0.10803
Zn	1.20		−0.08968
Se	1.18	−0.33139	
Pd	1.30		−0.13101
Te	1.37	−0.29440	

[a]Covalent radii, R_Z, are from ref 40.

charges noted earlier for EQeq, but, instead, it predicts the same qualitative variations in atomic charge as H-I. Across the entire ATMO data set, the MAD for V, Se, and Te decreased by an average of 0.29|e^-|, 1.20|e^-|, and 1.15|e^-|, respectively, compared to EQeq. In addition, the description of charges on the nonmetallic atoms is also improved by EQeq+C: the MAD of the H,C, N, O atoms decreased by 0.12|e^-|, 0.08|e^-|, 0.12|e^-|, and 0.15|e^-|, respectively, compared to EQeq. Since the ATMO compounds are templated by amines, we also examined how the N atom charges in primary, secondary, and tertiary amines behaved in the two methods. In general, the mean (signed) deviations become more negative as the degree of substitution increases. Moreover, the EQeq+C charges are smaller than the EQeq charges for all N atoms. For primary, secondary, and tertiary amine N atoms, the mean (signed) deviations of the EQeq charges were 0.71 ± 0.03e^-, 0.19 ± 0.06e^-, and −0.25 ± 0.07e^-, respectively. The mean deviations for the EQeq+C charges were 0.30 ± 0.03e^-, −0.16 ± 0.06e^-, and −0.51 ± 0.12e^-, respectively. In other words, the +C correction makes the nitrogen atoms more negative by −0.41|e^-|, −0.35|e^-|, and −0.25|e^-| for primary, secondary, and tertiary amines. This downward shift greatly improves the primary amine charges, leaves the MAD of the secondary amine charges about the same (but now overshoots the correct value making these too negative), and exacerbates the errors of the tertiary amine charges.

How well does EQeq+C reproduce the MOF/REPEAT charges? We parametrized an EQeq+C correction for the REPEAT charges on the 12 MOFs from the paper by Wilmer et al.,[13] listed in Table 2. Figure 7 shows a comparison between

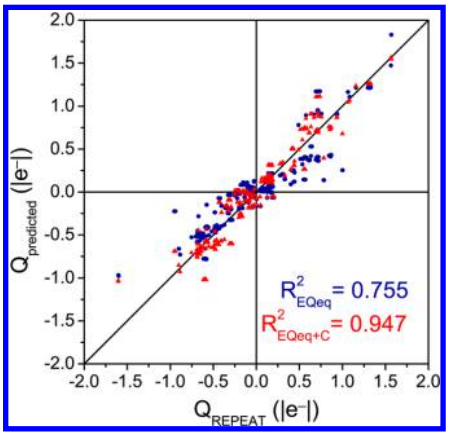


Figure 7. EQeq (blue circles) and EQeq+C (red triangles) charges versus REPEAT charges for MOFs, using ε_r = 1.67, α = 2.69 Å^{-1}. The diagonal line indicates the bisectrix.

the REPEAT charges (horizontal axis) and the EQeq (blue) and EQeq+C (red) charges. EQeq+C significantly improves upon the EQeq charges, increasing R^2 from 0.755 to 0.947. The MAD values for each MOF are shown in Table 2. With the exception of ZIF-8, the MAD of the EQeq+C charges is significantly smaller than that of the uncorrected EQeq charges. The best improvement is for HKUST-1, where the EQeq+C MAD is only 20% of the EQeq MAD; in other cases, the MAD is 34−68% as large. However, for ZIF-8, the MAD for the EQeq+C charges actually increases by more than 0.07e^-, or 54% compared to EQeq. This is the only zeolitic imidazolate framework (ZIF) in the data set, so the 2-methylimidazole linker is an outlier compared to the other compounds used for the parametrization. While the penalty here is small, this highlights the need for the training set to adequately represent the types of compounds to which the EQeq+C correction will be applied.

4.4. Application of EQeq+C to Peptides. How transferable are the EQeq+C parameters? We tested this by comparing EQeq and EQeq+C/ATMO/H-I against H-I charges for the set of dipeptides shown in Table 3. This is an extreme test of transferability, since the chemistry is very different from that of the ATMOs. In the dipeptides, the oxygen and nitrogen atoms are primarily present in peptide

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

bonds, in contrast to the highly nucleophilic oxide anions and the simple amines contain in the ATMO data set. As shown in Figure 8, EQeq (black symbols) poorly reproduces the H-I

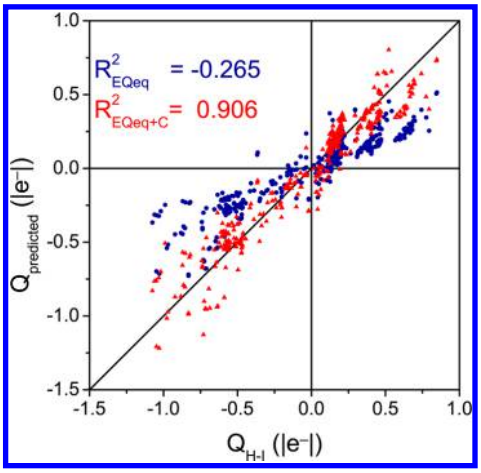


Figure 8. Comparison of EQeq (blue circles) and EQeq+C/ATMO/H-I (red triangles) to H-I charges for the dipeptide data set, using ε_r = 1.67, α = 2.10 Å^{-1}.

charges (R^2 = −0.266). The negative R^2 indicates that agreement between $Q_{predicted}$ and Q_{H-I} is worse than assuming the mean value, which is evident in the large deviations from the bisectrix of the blue EQeq data points. In contrast, the EQeq+C charges (green symbols) are strongly correlated to the H-I charges (R^2 = 0.906). The MAD for each of these methods is shown in Table 3. The MAD for all EQeq+C charges is 0.090|e^-|, which is less than half the MAD for all EQeq charges (0.196|e^-|).

The EQeq and EQeq+C charges for BEQJEN and XUDVOH initially had a poor agreement with the H-I charges, particularly for the charges on the methyl group carbons on the alanine and valine residues in these structures. Since hydrogen atom positions are poorly described by X-ray crystallography, it is typical to place the hydrogens in idealized positions. A cursory examination of the crystal structures showed no obviously unreasonable atomic geometries: in BEQJEN (XUDVOH), the C−H bonds were 0.98 Å (0.94−1.00 Å), the C−C bonds were 1.53 Å (1.5 Å), and the H−C−H and C−C−H bond angles were 109.5° (109.5°), with no unreasonable steric clashes. However, we suspected that the idealized geometries were the cause of the poor agreement with the H-I charges. Keeping the unit cell dimensions and the positions of the non-hydrogen atoms fixed, we used PAW/PBE-GGA calculations to optimize the hydrogen atom positions. The optimized geometry had slightly longer C−H bond lengths (averaging 1.10 Å) and narrower H−C−H bond angles (107.3−108.3°). We then recalculated the EQeq, EQeq+C, and H-I charges for the optimized structures. The final optimized structures are contained in the Supporting Information, and the charge calculations based on these structures are indicated as (opt) in Table 3. The hydrogen atom geometry optimization caused the H-I charges in BEQJEN and XUDVOH to change by a relatively small average of 0.0475|e^-| per atom; the general trend was to reduce the magnitude of the charges. Using the optimized hydrogen atom positions left the discrepancy between EQeq and H-I charges essentially unchanged, but it reduced the discrepancy between EQeq+C and H-I charges by nearly a factor of 2, making the MAD for these compounds comparable to those seen in the other dipeptides in Table 3. The EQeq and EQeq +C charges change by 0.1363|e^-| and 0.157|e^-| per atom following the geometry optimization. The largest changes were for carbon atoms that had unreasonably large negative charges in the idealized structure. Carbons and hydrogens that were positive in the unoptimized structure became somewhat less positive, and the nitrogen and oxygen charges were essentially unchanged by the geometry optimization. This confirmed our hypothesis that the idealized geometries in the reported crystal structures were the cause of the problem for these two peptides.

Returning to the performance of the different charge methods, and in particular the transferability of the ATMO parametrization to the very different types of chemical bonding found in the peptide crystals, we performed an analysis of the mean absolute deviation from the H-I charges for the different atom types found in the peptide crystals (see Table 5). The +C

Table 5. Mean Absolute Deviation (MAD) of EQeq and EQeq+C from H-I Charges for Atom Types of Dipeptides Described in Table 3[a]

atom type	EQeq MAD	EQeq+C MAD	atoms	compounds
all atoms	0.196	0.090	1320	12
all nitrogen atoms	0.456	0.182	94	12
ammonium ion N	0.669	0.273	34	9
primary amide N	0.742	0.480	4	1
secondary amine and amide N	0.307	0.105	56	12
all oxygen atoms	0.271	0.063	166	12
peptide O	0.232	0.025	50	12
water O	0.436	0.064	24	4
carboxylate anion O	0.377	0.057	68	9
carboxylic acid hydroxyl O	0.114	0.287	6	2
carboxylic acid carbonyl O	0.144	0.088	6	2
alcohol O	0.175	0.225	12	4
all hydrogen atoms	0.129	0.057	674	12
all carbon atoms	0.217	0.136	386	12

[a]The latter two columns indicate the number of atoms and compounds represented by each type.

correction substantially reduces the MAD for all atom types, with the exception of a small increase for oxygen atoms in alcohols. Despite the initial concerns that oxygen atoms would be overly parametrized for metal oxides, EQeq+C gives better predictions even for types of compounds for which it has not been explicitly parametrized. The improved nitrogen atom charges also support transferability of the ATMO parametrization. As described above, the ATMO data set contains nitrogens in simple organic amines, whereas the dipeptide nitrogen atoms are predominantly ammonium counterions and amides. The largest MAD for both EQeq and EQeq+C are observed for primary amides; this is unsurprising because of the electron-withdrawing carbonyl group in the delocalized amide bond. The pairwise models underlying both EQeq and the +C correction do not account for this delocalization. Nonetheless, the +C correction improves the charges for all of the nitrogen atom types, and, overall, the MAD for the diverse nitrogen atom types found in the dipeptides is only slightly greater than the 0.12|e^-| for the ATMO data set. This transferability suggests that our approach corrects an underlying physical deficiency of EQeq in a systematic way rather than being merely an *ad hoc* system-dependent correction.

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

5. CONCLUSIONS

The EQeq partial charge scheme is physically motivated and computationally inexpensive, but it has difficulties describing the charges on high-oxidation-state transition metals. Introducing an empirical, pairwise bond-valence correction, denoted EQeq+C, solves this problem with negligible additional computational cost. We parametrized EQeq+C model parameters to reproduce ATMO H-I charges and MOF REPEAT charges to demonstrate how the parametrizations can be applied to several compound classes and charge partitioning schemes. In principle, the +C correction needs to be reparameterized for each compound class, although we demonstrate that these corrections are transferable by comparing to the H-I calculations on a set of dipeptides. The computational efficiency of our correction combined with the improvement in accuracy makes the EQeq+C method suitable for treating large unit cell solids and large-scale computational screening of materials. In future work, we intend to extend our approach to metal organic frameworks.

■ ASSOCIATED CONTENT

Supporting Information

Python programs implementing the EQeq+C fitting procedure and calculation. Optimized geometries of BEQJEN and XUDVOH dipeptide crystals as well as comparisons of charges pre- and postgeometry optimization. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00037.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jschrier@haverford.edu.

Present Address

[†](M.D.S.) Department of Chemistry, Stanford University, Palo Alto, California 94305, United States.

Funding

M.D.S. and D.R. were supported by a grant to Haverford College from the Howard Hughes Medical Institute and by the Research Corporation for Science Advancement's Cottrell Scholar grant. G.C.M.-N. was supported by the Brian Kovaric Memorial Fellowship. L.M.R. was supported by a grant to the John P. Chesick Scholars program at Haverford College by the San Francisco Foundation. J.S. also acknowledges the Henry Dreyfus Teacher-Scholar Award program.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Malia Wenny for a careful reading of the manuscript. This work used the resources of the National Energy Research Scientific Computing Center (NERSC), which supported by the U.S. Department of Energy under contract no. DE-AC02-05CH11231, and the Extreme Science and Engineering Discovery Environment (XSEDE)[68], supported by National Science Foundation grant no. ACI-1053575.

■ REFERENCES

(1) Férey, G. Microporous Solids: From Organically Templated Inorganic Skeletons to Hybrid Frameworks...Ecumenism in Chemistry. *Chem. Mater.* **2001**, *13*, 3084−3098.

(2) Zheng, C.; Liu, D.; Yang, Q.; Zhong, C.; Mi, J. Computational Study on the Influences of Framework Charges on CO_2 Uptake in Metal-Organic Frameworks. *Ind. Eng. Chem. Res.* **2009**, *48*, 10479−10484.

(3) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.

(4) Mobley, D. L.; Dumong, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.

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

(6) Bultinck, P.; van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical Analysis and Extension of the Hirshfeld Atoms in Molecules. *J. Chem. Phys.* **2007**, *126*, 144111.

(7) Manz, T. A.; Sholl, D. S. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Non-periodic Materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455−2468.

(8) Xu, Q.; Zhong, C. A General Approach for Estimating Framework Charges in Metal-Organic Frameworks. *J. Phys. Chem. C* **2010**, *114*, 5035−5042.

(9) Zheng, C.; Zhong, C. Estimation of Framework Charges in Covalent Organic Frameworks Using Connectivity-Based Atom Contribution Method. *J. Phys. Chem. C* **2010**, *114*, 9945−9951.

(10) Rai, B. K.; Bakken, G. A. Fast and Accurate Generation of ab Initio Quality Atomic Charges Using Nonparametric Statistical Regression. *J. Comput. Chem.* **2013**, *34*, 1661−1671.

(11) Verstraelen, T.; Bultinck, P.; van Speybroeck, V.; Ayes, P. W.; van Neck, D.; Waroquier, M. The Significance of Parameters in Charge Equilibration Models. *J. Chem. Theory Comput.* **2011**, *7*, 1750−1764.

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

(13) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.* **2012**, *3*, 2506−2511.

(14) Glor, E. C.; Blau, S. M.; Yeon, J.; Zeller, M.; Halasyamani, P. S.; Schrier, J.; Norquist, A. J. $[R\text{-}C_7H_{16}N_2][V_2Te_2O_{10}]$ and $[S\text{-}C_7H_{16}N_2]$-$[V_2Te_2O_{10}]$; New Polar Templated Vanadium Tellurite Enantiomers. *J. Solid State Chem.* **2011**, *184*, 1445−1450.

(15) Smith, M. D.; Blau, S. M.; Chang, K. B.; Zeller, M.; Schrier, J.; Norquist, A. J. Beyond Charge Density Matching: The Role of C−H···O Interactions in the Formation of Templated Vanadium Tellurites. *Cryst. Growth Des.* **2011**, *11*, 4213.

(16) Smith, M. D.; Blau, S. M.; Chang, K. B.; Tran, T. T.; Zeller, M.; Halasyamani, P. S.; Schrier, J.; Norquist, A. J. Inducing Polarity in $[VO_3]_n^{n-}$ Chain Compounds Using Asymmetric Hydrogen-Bonding Networks. *J. Solid State Chem.* **2012**, *195*, 86−93.

(17) Chang, K. B.; Smith, M. D.; Blau, S. M.; Glor, E. C.; Zeller, M.; Schrier, J.; Norquist, A. J. Steric-Induced Layer Flection in Templated Vanadium Tellurites. *Cryst. Growth Des.* **2013**, *13*, 2190−2197.

(18) Koffer, J. H.; Olshansky, J. H.; Smith, M. D.; Hernandez, K. J.; Zeller, M.; Ferrence, G. M.; Schrier, J.; Norquist, A. J. Formation Principles for Templated Vanadium Selenite Oxalates. *Cryst. Growth Des.* **2013**, *13*, 4504−4511.

(19) Lendvay, G. On the Correlation of Bond Order and Bond Length. *J. Mol. Struct.: THEOCHEM* **2000**, *501*, 389−393.

(20) Pauling, L. Atomic Radii and Interatomic Distances in Metals. *J. Am. Chem. Soc.* **1947**, *69*, 542−553.

(21) Brown, I. D. Recent Developments in the Methods and Applications of the Bond Valence Model. *Chem. Rev.* **2009**, *109*, 6858−6919.

(22) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.

(23) Gonze, X.; Amadon, B.; Anglade, P.-M.; Beuken, J.-M.; Bottin, F.; Boulanger, P.; Bruneval, F.; Caliste, D.; Caracas, R.; Cote, M.; et al. ABINIT: First-Principles Approach to Material and Nanosystem Properties. *Comput. Phys. Commun.* **2009**, *180*, 2582−2615.

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

(24) Bottin, F.; Leroux, S.; Knyazev, A.; Zerah, G. Large Scale ab Initio Calculations Based on Three Levels of Parallelization. *Comput. Mater. Sci.* **2008**, *42*, 329−336.
(25) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(26) Perdew, J. P.; Burke, K.; Ernzerhof, M. Errata: Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1997**, *78*, 1396.
(27) Fuchs, M.; Scheffler, M. Ab Initio Pseudopotentials for Electronic Structure Calculations of Poly-atomic Systems Using Density-Functional Theory. *Comput. Phys. Commun.* **1999**, *119*, 67−98.
(28) Lusi, M.; Barbour, L. J. Determining Hydrogen Atom Positions for Hydrogen Bonded Interactions: A Distance-Dependent Neutron-Normalized Method. *Cryst. Growth. Des.* **2011**, *11*, 5515−5521.
(29) Krese, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47*, 558.
(30) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49*, 14251.
(31) Kresse, G.; Furthmüller, J. Efficiency of ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15.
(32) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169.
(33) Blochl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953.
(34) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758.
(35) Fischer, C. F.; Gaigalas, G. Extension of the HF Program to Partially Filled f-Subshells. *Comput. Phys. Commun.* **1996**, *98*, 255−264.
(36) Vanpoucke, D. E. P.; Bultinck, P.; Van Driessche, I. Extending Hirshfeld-I to Bulk and Periodic Materials. *J. Comput. Chem.* **2013**, *34*, 405−417.
(37) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. Hirshfeld-E Partitioning: AIM Charges with an Improved Trade-off between Robustness and Accurate Electrostatics. *J. Chem. Theory Comput.* **2013**, *9*, 2221−2225.
(38) Miliordos, E.; Harrison, J. F. Hirshfeld Density Partitioning Technique: A First Application to the Transition Metal Compounds, HScO, TiO, VO. *J. Chem. Phys.* **2013**, *138*, 184305.
(39) *Ground Levels and Ionization Energies for the Neutral Atoms*; United States National Institute of Standards and Technologies Physical Measurement Laboratory: Gaithersburng, MD. http://www.nist.gov/pml/data/ion_energy.cfm.
(40) *CRC Handbook of Chemistry and Physics*, 94th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2013; Section 9, No. 49.
(41) Chang, K. B.; Smith, M. D.; Zeller, M.; Norquist, A. J. catena-Poly[2,2′,2″-nitrilotris(ethan-aminium) [tri-μ-oxido-tris-[dioxidovanadate(V)]] monohydrate]. *Acta Crystallogr., Sect. E* **2013**, *69*, m570−m571.
(42) Feng, M.-L.; Mao, J.-G. Syntheses and Crystal Structures of the First Organically Templated Metal Tellurites. *J. Solid State Chem.* **2005**, *178*, 2256−2261.
(43) Olshansky, J. H.; Tran, T. T.; Hernandez, K. J.; Zeller, M.; Halasyamani, P. S.; Schrier, J.; Norquist, A. J. Role of Hydrogen-Bonding in the Formation of Polar Achiral and Nonpolar Chiral Vanadium Selenite Frameworks. *Inorg. Chem.* **2012**, *51*, 11040−11048.
(44) Dietzel, P. D. C.; Morita, Y.; Blom, R.; Fjellvag, H. An in Situ High-Temperature Single-Crystal Investigation of a Dehydrated Metal-Organic Framework Compound and Field-Induced Magnetization of One-Dimensional Metaloxygen Chains. *Angew. Chem., Int. Ed.* **2005**, *44*, 6354−6358.
(45) Chui, S. S. Y.; Lo, S. M.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. A Chemically Functionalizable Nanoporous Material $[Cu_3(TMA)_2(H_2O)_3]_N$. *Science* **1999**, *283*, 1148−1150.
(46) Li, H.; Eddaoudi, M.; O'Keefe, M.; Yaghi, O. M. Design and Synthesis of an Exceptionally Stable and Highly Porous Metal-Organic Framework. *Nature* **1999**, *402*, 276−279.
(47) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keefe, M.; Yaghi, O. M. Systematic Design of Pore Size and Functionality in Isoreticular MOFs and their Application in Methane Storage. *Science* **2002**, *295*, 469−472.
(48) Caskey, S. R.; Wong-Foy, A. G.; Matzger, A. J. Dramatic Tuning of Carbon Dioxide Uptake via Metal Substitution in a Coordination Polymer with Cylindrical Pores. *J. Am. Chem. Soc.* **2008**, *130*, 10870−10871.
(49) Barthelet, K.; Marrot, J.; Riou, D.; Ferey, G. A Breathing Hybrid Organic-Inorganic Solid with Very Large Pores and High Magnetic Characteristics. *Angew. Chem., Int. Ed.* **2001**, *41*, 281−284.
(50) Dietzel, P. D. C.; Panella, B.; Hirscher, M.; Blom, R.; Fjellvag, H. Hydrogen Adsorption in a Nickel Based Coordination Polymer with Open Metal Sites in the Cylindrical Cavities of the Desolvated Framework. *Chem. Commun.* **2006**, 959−961.
(51) Navarro, J. A. R.; Barea, E.; Salas, J. M.; Masciocchi, N.; Galli, S.; Sironi, A.; Ania, C. O.; Parra, J. B. H_2, N_2, CO, and CO_2 Sorption Properties of a Series of Robust Sodalite-type Microporous Coordination Polymers. *Inorg. Chem.* **2006**, *45*, 2397−2399.
(52) Wong-Foy, A. G.; Lebel, O.; Matzger, A. J. Porous Crystal Derived from a Tricarboxylate Linker with Two Distinct Binding Motifs. *J. Am. Chem. Soc.* **2007**, *129*, 15740−15741.
(53) Park, T.-H.; Cychosz, K. A.; Wong-Foy, A. G.; Dailly, A.; Matzger, A. J. Gas and Liquid Phase Adsorption in Isostructural Cu_3[biaryltricarboxylate]$_2$ Microporous Coordination Polymers. *Chem. Commun.* **2011**, *47*, 1452−1454.
(54) Park, K. S.; Ni, Z.; Cote, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keefe, M.; Yaghi, O. M. Exceptional Chemical and Thermal Stability of Zeolitic Imidazolate Frameworks. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10186−10191.
(55) Rowsell, J. L. C.; Millward, A. R.; Park, K. S.; Yaghi, O. M. Hydrogen Sorption in Functionalized Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2004**, *126*, 5666−5667.
(56) Stenkamp, R. E.; Jensen, L. H. Crystal-Structure of a Pepsin Substrate: *N*-Acetyle-L-phenylalanyl-L-tyrosine. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1973**, *29*, 2872−2878.
(57) Sun, H.; Oldfield, E. Tryptophan Chemical Shift in Peptides and Proteins: A Solid State Carbon-13 Nuclear Magnetic Resonance Spectroscopic and Quantum Chemical Investigation. *J. Am. Chem. Soc.* **2004**, *126*, 4726−4734.
(58) Mendham, A. P.; Palmer, R. A.; Potter, B. S.; Dines, T. J.; Snowden, M. J.; Withnall, R.; Chowdhry, B. Z. Vibrational Spectroscopy and Crystal Structure Analysis of Two Polymorphs of the Di-amino Acid Peptide Cyclo(L-Glu-L-Glu). *J. Raman Spectrosc.* **2010**, *41*, 288−302.
(59) Morris, A.; Geddes, A.; Sheldrick, B. Cyclo-glycyl-tryptophyl. *Cryst. Struct. Commun.* **1974**, *3*, 345.
(60) Görbitz, C. L-Isoleucyl-L-phenylalanine dihydrate. *Acta Crystallogr., Sect. C* **2004**, *60*, o371.
(61) Moen, A.; Frøseth, M.; Görbitz, C.; Dalhus, B. A Non-planar Peptide Bond in L-Seryl-L-valine. *Acta Crystallogr., Sect. C* **2004**, *60*, o564.
(62) Pasternak, R. A.; Katz, L.; Corey, R. B. The crystal structure of glycyl-L-asparagine. *Acta Crystallogr.* **1954**, *7*, 225−236.
(63) Benabicha, F.; Pichon-Pesme, V.; Jelsch, C.; Lecomte, J.; Khmou, A. Experimental Charge Density and Electrostatic Potential of Glycyl-L-threonine Dihydrate. *Acta Crystallogr., Sect. B* **2000**, *56*, 155−165.
(64) Görbitz, C. H. L-Threonyl-L-alanine. *Acta Crystallogr., Sect. E* **2005**, *61*, o2012−o2014.
(65) Steiner, T. L-Histidyl-L-alanine Dihydrate. *Acta Crystallogr., Sect. C* **1996**, *52*, 2554−2556.
(66) Görbitz, C.; Gundersen, E. L-Valyl-L-alanine. *Acta Crystallogr., Sect. C* **1996**, *52*, 1764−1767.
(67) Görbitz, C. An Exceptionally Stable Peptide Nanotube System with Flexible Pores. *Acta Crystallogr., Sect. B* **2002**, *58*, 849−854.

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374

(68) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62−74.

DOI: 10.1021/acs.jctc.5b00037
J. Chem. Theory Comput. 2015, 11, 3364−3374