MOLECULAR PHYSICS, 2017
VOL. 115, NO. 24, 3155–3163
https://doi.org/10.1080/00268976.2017.1351628

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

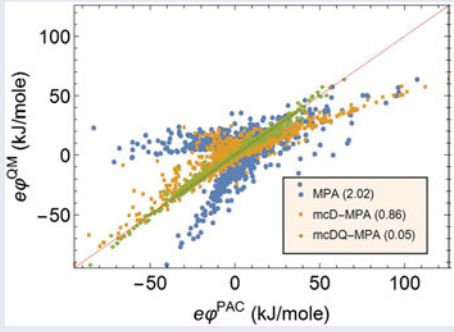
Minimally corrected partial atomic charges for non-covalent electrostatic interactions

Rebecca Efrat Hadad and Roi Baer

Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem, Israel

ABSTRACT

We develop a new scheme for determining molecular partial atomic charges (PACs) with external electrostatic potential (ESP) closely mimicking that of the molecule. The PACs are the 'minimal corrections' to a reference set of PACs necessary for reproducing exactly the tensor components of the Cartesian zero-, first- and second- molecular electrostatic multipoles. We evaluate the quality of ESP reproduction when 'minimally correcting' (MC) Mulliken, Hirshfeld or iterative-Hirshfeld reference PACs. In all these cases, the MC-PACs significantly improve the ESP while preserving the reference PACs' invariance under the molecular symmetry operations. When iterative-Hirshfeld PACs are used as reference, the MC-PACs yield ESPs of comparable quality to those of the ChElPG charge fitting method.



ARTICLE HISTORY
Received 6 December 2016
Accepted 23 June 2017

KEYWORDS
Partial atomic charges; force fields; non-covalent electrostatic interactions; molecular charge moments

1. Introduction

Partial atomic charges (PACs), i.e. point charges placed on the nuclei position of a molecule, are often used in large-scale molecular mechanics calculations to replace the detailed quantum mechanical charge distributions [1–7]. The model is extremely useful, since by using them, the long-range electrostatic forces acting *between* molecules can be expressed as a sum of pairwise interactions, enabling a fast computation, important especially as molecules jiggle around and rotate quite a lot during the course of the simulation. The question of just how to determine PACs for this purpose is critical. We argue that the most important constraint is the exact reproduction of the low-order electrostatic moments (ESM), the monopole $Q = e\int \rho(\boldsymbol{r})\, d\boldsymbol{r}$, which is the total charge of the system, the dipole $\mu_i = e\int \rho(\boldsymbol{r})\, r_i d\boldsymbol{r}$ $(i = x, y, z)$ and the quadrupole moment $\Theta_{ij} = e\int \rho(\boldsymbol{r})\left(3r_i r_j - \delta_{ij} r^2\right) d\boldsymbol{r}$, where $e\rho(\boldsymbol{r})$ is the charge distribution within the molecule.[1] These moments are of critical importance as they determine the far-field potential produced by the molecule, as evident from the monopole expansion:[2]

$$4\pi\epsilon_0\varphi(\boldsymbol{r}) \equiv e\int \rho(\boldsymbol{r}')\left|\boldsymbol{r}-\boldsymbol{r}'\right|^{-1} d\boldsymbol{r}' \tag{1}$$

$$= \frac{Q}{r} + \frac{\mu_i r_i}{r^3} + \frac{1}{2}\frac{r_i \Theta_{ij} r_j}{r^5} + \cdots . \tag{2}$$

These low-order ESMs also control the electrostatic interaction energy W_{es} between the molecule (and through it the forces) with a weakly non-constant potential $\varphi^{other}(\boldsymbol{r})$ resulting from the other molecules or distant charged sources [8]:

$$W_{es} = Q\varphi^{other} + \mu_i \varphi_i^{other} + \frac{1}{6}\Theta_{ij}\varphi_{ij}^{other} + \cdots \tag{3}$$

where $\varphi_i^{other} = \frac{\partial \varphi^{other}}{\partial r_i}$ and $\varphi_{ij}^{other} = \frac{\partial^2 \varphi^{other}}{\partial r_i \partial r_j}$ (estimated at a central point within the molecule), etc. This pivotal

CONTACT Roi Baer roi.baer@huji.ac.il

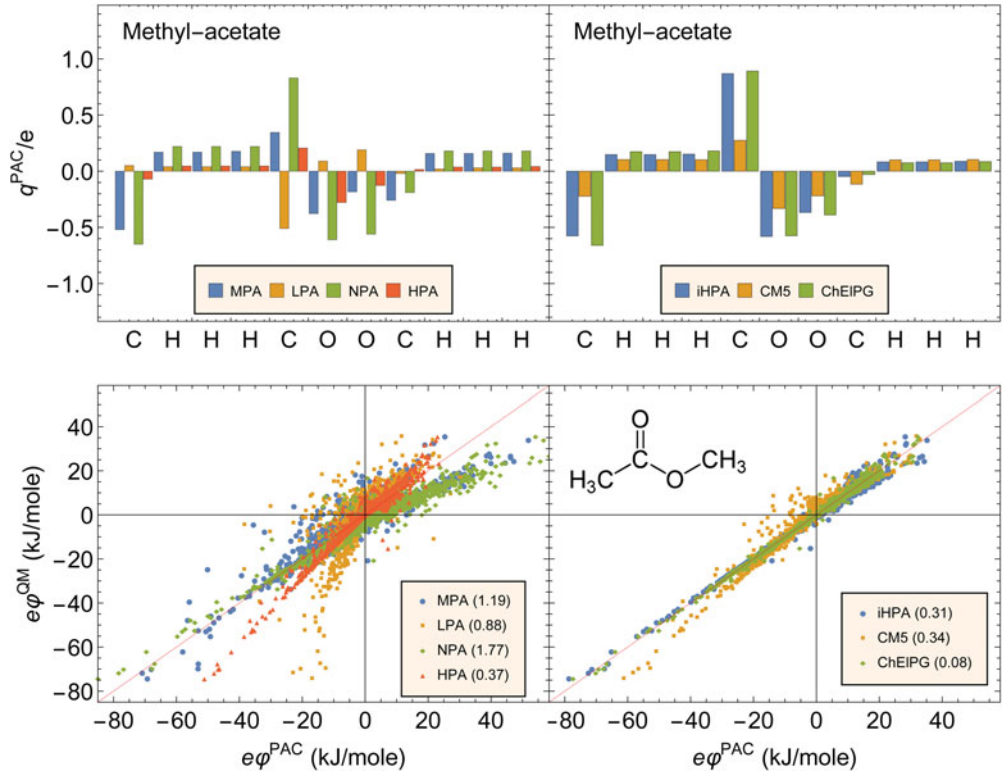


Figure 1. (Colour online) The PACs (top panels) and the ESP correlation (bottom panels) for the methyl-acetate molecule, using standard population analysis methods (left panels) and other PAC methods (right panels). The mean absolute relative deviation (MARD) of Equation (5) appears in parenthesis near each PAC method. Each point in the ESP correlation plot describes a pair of potentials $[\varphi^{PAC}(\boldsymbol{r}), \varphi^{QM}(\boldsymbol{r})]$, the abscissa is the PAC potential (Equation (4)) and the ordinate is the quantum potential (Equation (1)) where $\boldsymbol{r}$ is taken from a subset of grid points of spacing $\Delta x = 0.3$ Å around the molecule (see description of the grid in Section 3).

dual role of ESMs is what drives the requirement that the charge distribution of the PACs reproduces exactly low-lying molecular ESMs (MOL-ESMs). This point was discussed at length in [9] where the importance of adherence to the ESMs was demonstrated. An efficient elegant method for achieving this in as many as possible moments has been developed [10] although inapplicable for large molecule charges due to numerical instabilities [11].

Another source of PACs is the quantum mechanical population analysis (PA) techniques, such as the Mulliken (MPA) [12], Loewdin (LPA) [13], Hirshfeld (HPA) [14] and natural population (NPA) [15] analyses. These PAs reflect not only the charge distribution but also aspects of the quantum mechanical wave function. In Figure 1 (top left), we show as bar-plot the PACs produced by these methods applied to the methyl-acetate molecule. It is seen that the different methods produce sometimes significantly different sets of PACs, even PAC signs are not preserved! For example, the LPA assigns positive charges to oxygen atoms, which seems awkward given their high electronegativity. Furthermore, standard PAs do not reproduce the MOL-ESPs closely, as shown in the ESP correlation plot of Figure 1 (bottom left), where several PAC-ESPs,

$$\varphi^{PAC}(\boldsymbol{r}) = \frac{e}{4\pi\epsilon_0}\sum_a \frac{q_a}{|\boldsymbol{r}-\boldsymbol{R}_a|}, \tag{4}$$

are plotted vs. the MOL-ESP $\varphi(\boldsymbol{r})$ calculated from the QM density (Equation (1)) at a grid point $\boldsymbol{r}$. The thin red line in the plot corresponds to the perfectly correlated condition $\varphi^{PAC} = \varphi^{QM}$. In order to quantify the quality of $\varphi^{PAC}(\boldsymbol{r})$, we define the mean absolute relative deviation (MARD) from $\varphi(\boldsymbol{r})$ as

$$MARD\left(\varphi^{PAC}, \varphi\right) = \left\langle \left| \frac{\varphi^{PAC}(\boldsymbol{r}) - \varphi(\boldsymbol{r})}{\varphi(\boldsymbol{r})} \right| \right\rangle, \tag{5}$$

where an average is taken over all grid points $\boldsymbol{r}$ for which: (1) $\boldsymbol{r}$ is 'outside of the molecule', i.e. its distance from any nucleus a is larger than the atomic van der Waals radius R_a^{vdW} [16]) and (2) $\boldsymbol{r}$ is not too far from the molecule, so that its potential $|\varphi(\boldsymbol{r})|$ is not smaller than the threshold value of $e\varphi_{thresh} = 0.3$ eV.[3] PACs obtained by 'standard' PAs have large MARDs, ranging from 0.37 for HPA up to a whopping 1.76 for NPA. On the right panel of the figure, we show data concerning the same molecule, but

using the iterative-Hirshfeld method (iHPA) [17–19], the CM5 method [20], which is a parameterised database correction to HPA charges, and the ChElPG method [21], which selects PACs that reconstruct the *ab initio* ESP on a set of grid points as close as possible. The latter approach is taken here as representative of a class of methods routinely used for PACs determination. Other members of this method class are the 'charge from ESPs' (ChElP) [22], the Merz–Kollman [16,23,24], the charge-restraint ESPs (RESP) [25,26], atomic multipoles ESPs [27], in combination with molecular multipoles [28] (related to the method proposed here), the dynamical RESP (D-RESP) [29] and Hu-Yang fitting [30].The iHPA, CM5 and ChElPG methods yield much improved description of the ESP with MARD going from 0.3 for iHPA and CM5 down to 0.08 for ChElPG. Despite the close ESP fit, ChElPG produces PACs that are usually not invariant under transformations preserving the point symmetry of the molecule or under rotations or translations of the nuclei with respect to the real space grid used to perform the fit. Furthermore, in larger molecules, the PACs of atoms distant from the molecular surface can become unwieldy large. Both of these issues are discussed in the literature [30,31]. This instability is likely linked to the fact that the number of parameters derivable from the ESP in a statistically significant way is considerably less than the number of atoms [32]. Therefore, iHPA and CM5 are often considered preferred approaches for PACs, although as seen in the figure, both methods leave ample room for improvement. Note that the iHPA charges for this molecule are close to the ChElPG PACs.

Here, we study a new idea: take PACs which are as close as possible to a reference set, for example the MPA, HPA or iHPA PACs, but insist that they reproduce exactly the components of the lowest ESM tensors (dipole and quadrupole) characterising the molecular charge distribution. We formulate a straightforward method to determine such 'minimally corrected PACs' (Section 2) and then benchmark the results using a subset of molecules taken from the database of [20] (Section 3). Final conclusions are summarised in Section 4. All MPA, HPA and iHPA PACs, as well as the associated MOL-ESPs and MOL-ESMs, were computed using developer versions of Q-Chem 4.3 and 4.4 [33] at the M06-L DFT level [34] and using the MG3 semi-diffuse (MG3S) basis set [35]. This functional/basis set combination was used for developing the CM5 approach. The CM5, NPA and LPA results were taken from [20].

2. Method

Consider a molecule having A nuclei at given Cartesian positions $\boldsymbol{R}^a=\left(R^a_x, R^a_y, R^a_z\right)$ $(a = 1, \ldots, A)$, for which a QM calculation has determined the charge density $\rho(\boldsymbol{r})$ of the molecule and from it, low order moments the charge Q, the dipole μ_i and the symmetric traceless quadrupole moment tensor Θ_{ij}. Note that below, we use the notation $\Theta^D_i \equiv \Theta_{ii}$ for the diagonal elements of Θ and $\Theta^{OD}_i \equiv \Theta_{jk}$ where $i = x, y, z$ and ijk is a cyclic permutation of xyz. For any set of PACs $\boldsymbol{q} = (q_1, \ldots, q_A)$, we define the PAC-ESMs as the monopole (total charge) $Q^{PAC} \equiv e\Sigma_a q_a$, the dipole $\mu^{PAC}_i \equiv e\sum_a q_a R^a_i$ and the quadrupole $\Theta^{PAC}_{ij} \equiv e\sum_a q_a(3R^a_i R^a_j - \delta_{ij}(R^a)^2)$, $(i, j = x, y, z)$. Given a set of reference PACs $\boldsymbol{q}^{ref}$, we seek to determine a 'minimally corrected' set of PACs $\boldsymbol{q}^{mc} = \boldsymbol{q}^{ref} + \Delta\boldsymbol{q}$ such that the size of the correction $\|\Delta\boldsymbol{q}\|^2 = \Delta\boldsymbol{q}\cdot\Delta\boldsymbol{q}$ is *minimal* but the multipoles are equal to the QM-determined multipoles, i.e. the following constraints are satisfied:

$$\begin{aligned} c &= Q - Q^{PAC} = 0, \\ c_i &= \mu_i - \mu^{PAC}_i = 0, \\ c_{ij} &= \Theta_{ij} - \Theta^{PAC}_{ij} = 0. \end{aligned} \tag{6}$$

Note that the number of constraints (denoted C) in Equation (6) is 9 and not 13 since the electric quadrupole tensor is symmetric and traceless. Point symmetries can reduce this number of constraints further. If, for example, both positive and negative charge densities are symmetric against the reflection through a plane (the x–y plane, for example) then there are three constraint less (one from the z component of the dipole and two from the XZ and YZ components of the quadrupole, which are zero by symmetry). Only when the number of atoms A in the molecule is greater than the number of constraints C can we hope to reproduce the constraints exactly. We therefore demand that $A > 9$ and use the $A - C$ additional 'degrees of freedom' to minimise the deviance $\Delta\boldsymbol{q}$. When $2 \le A \le 9$, we avoid the quadrupole moment constraint and use only the dipole moment constraint.

We are led to consider the Lagrangian

$$\begin{aligned} L_{mcDQ} = \tfrac{1}{2}\sum_a \left(q^{mc}_a - q^{ref}_a\right)^2 - \lambda c - \lambda_i c_i \\ - \sum_{xy,yz,zx} \lambda^{OD}_{ij} c_{ij} - \sum_{x,y,z} \lambda^{D}_{ii} c_{ii} \end{aligned} \tag{7}$$

as a function of the A q's and the ten Lagrange multipliers: one λ, three λ_i's , three diagonal λ^D_{ii} and three off-diagonal λ^{OD}_{jk} where $i = x, y, z$ and ijk is a cyclic permutation of xyz. Taking derivatives with respect to these variables and equating to zero leads to the following set of $(10 + A)$ linear equations in $(10 + A)$ unknowns, given here in

block-matrix/vector form[4]:

$$S\begin{pmatrix} q_1^{mc} \\ \vdots \\ q_A^{mc} \\ \lambda_{1\times 1} \\ \lambda_{3\times 1} \\ \lambda^{D}_{3\times 1} \\ \lambda^{OD}_{3\times 1} \end{pmatrix} = \begin{pmatrix} q_1^{ref} \\ \vdots \\ q_A^{ref} \\ Q_{1\times 1} \\ \mu_{3\times 1} \\ \Theta^{D}_{3\times 1} \\ \Theta^{OD}_{3\times 1} \end{pmatrix}. \tag{8}$$

The $(A+10)\times(A+10)$ matrix S is of the following form:

$$S = \begin{pmatrix} \lceil & \cdots & \rceil & -1 & \lceil & & \rceil & \lceil & & \rceil & \lceil & & \rceil \\ \vdots & I_{A\times A} & \vdots & \vdots & \vdots & -D_{A\times 3} & \vdots\vdots & -T^{D}_{A\times 3} & \vdots\vdots & -T^{OD}_{A\times 3} & \vdots \\ \lfloor & \cdots & \rfloor & -1 & \lfloor & & \rfloor & \lfloor & & \rfloor & \lfloor & & \rfloor \\ 1 & \cdots & 1 & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lceil & \cdots & \rceil & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ & D_{3\times A} & & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lfloor & \cdots & \rfloor & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lceil & \cdots & \rceil & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ & T^{D}_{3\times A} & & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lfloor & \cdots & \rfloor & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lceil & \cdots & \rceil & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ & T^{OD}_{3\times A} & & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \\ \lfloor & \cdots & \rfloor & 0 & 0 & 0 & 0\,0 & 0 & 0\,0 & 0 & 0 \end{pmatrix} \tag{9}$$

and depends *only* on the location of the atomic nuclei. The matrix is composed of blocks: the $I_{A\times A}$ block is a $A\times A$ unit matrix, $D_{A\times 3}$, $T^{D}_{A\times 3}$ and $T^{OD}_{A\times 3}$ are matrices of dimension $A\times 3$ (three columns each of length A) of matrix elements: $D_{ai}=eR_i^a$ and $T^{D}_{ai}=e\left(3R_i^aR_i^a-\|R^a\|^2\right)$ for $a=1,\ \ldots, A$ and $i=x,y,z$ and $T^{OD}_{ai}=3eR_j^aR_k^a$ (where the ordered set i,j,k is a cyclic permutation of x,y,z). The $D_{3\times A}$, $T^{D}_{3\times A}$, and $T^{OD}_{3\times A}$ blocks are, respectively, the transposed matrices. The $A+10$ column-vector on the left-hand side of Equation (8) includes the unknowns, the A partial charges q_a^{mc} and the $\lambda's$, the ten Lagrange multipliers for the ten constraints. The $10+A$ column-vector on the right-hand-side has A values of the reference charges q_a^{ref}, followed by the total charge on the molecule Q, then the three values of the QM dipole moment μ_i followed by the three values of the diagonal elements of the given QM quadrupole tensor $\Theta_i^D=\Theta_{ii}$ and finally the QM values of the three off-diagonal elements $\Theta_i^{OD}=\Theta_{jk}$ where $i=x$, y, z and ijk is a cyclic permutation of xyz. A similar equation holds for the mcD method, where the last six rows are erased from S and from the column vectors and the six right columns are erased from S as well. This leaves us with a $(A+4)\times(A+4)$ system of equations.

The structural matrix S may become singular or rank deficient. One trivial source for singularity is the use of three diagonal constraints while their sum is composed to be zero. The use of the singular-value-decomposition pseudo-inverse [36] for solving Equation (8) helps to bypass such a singularity. A more delicate source of singularities may arise from symmetry. For example, when the molecule is perfectly planar (or has a plane of symmetry) in the x–y plane, then the row corresponding to the dipole in the z direction $D_{az}=eR_z^a$ must be identically zero and the matrix S will be rank deficient. In this case, the T^{OD}_{ai} with $i=x$ and y must also be zero. In these cases, the SVD pseudo-inverse will automatically eliminate constraints that cannot be met due to this kind of symmetry. But for near-symmetrical configurations, instabilities may exist. In cases such as these, we can still spot problems by examining the values of the Lagrange multipliers λ, λ_i and λ_{ij} in the solution vector of Equation (8). The Lagrange multiplier is equal to the derivative of the minimal value of the Lagrangian L with respect to the constraint value (Q, μ_i and Θ_{ij}, respectively). Thus, if the *ab initio* dipole moment μ_x is given to precision $\delta\mu_x$, the product $|\lambda_x\delta\mu_x|$ is expected to be the error in the minimal value of L. Clearly, the minimising procedure is meaningless unless this error is much smaller than 1. Hence, it is important to eliminate 'offending' constraints from the matrix equation (the corresponding row and column in the matrix and the entry in the column vectors) for those having large Lagrange multipliers. We know that *ab initio* multipole properties are usually given to three digits, and hence we eliminate constraints corresponding to Lagrange multipliers large than 1000. The reduced equation is then solved and the remaining Lagrange multipliers are examined again. We repeat such elimination until all Lagrange multipliers have proper magnitudes. This pruning procedure helps avoid cases where small inaccuracies of the input data dominate the final result. Within the molecules studied here, such a pruning procedure was used only for few cases of molecules having a near-plane symmetry.

When symmetry is active, our procedures reduce the number of constraints and hence the number of *independent* q_a's (called number of degrees of freedom (NDOFs)). For example, the water molecule has three nuclei, but due to symmetry, the two H nuclei will have the same PACs and so NDOF = 2. Due to the symmetry, only the dipole moment in the direction of the C_2 axis is a constraint (the components perpendicular to the C2 axis are zero by symmetry). Together with the charge of water (0), we already have two constraints so one must give up imposing the quadrupole moment for water.

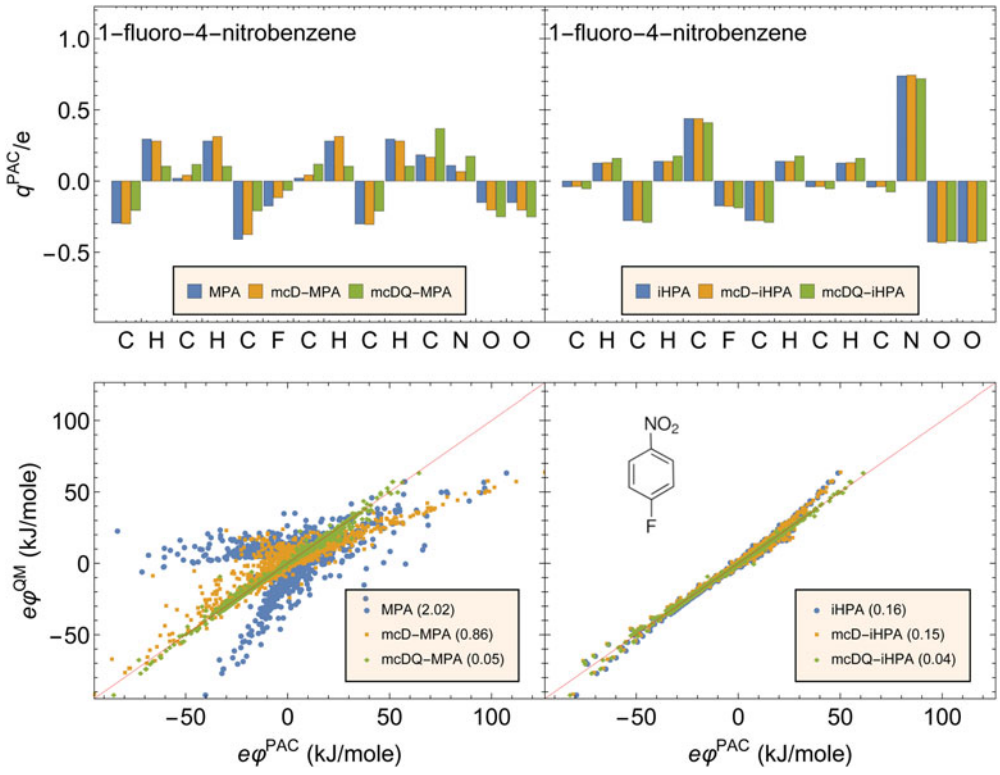


Figure 2. (Colour online) The PACs (top panels) and ESP correlation plots (bottom panels) for MPA, mcD-MPA and mcDQ-MPA (left panels) and iHPA, mcD-iHPA and mcDQ-iHPA (right panels) in the 1-fluoro-4-nitrobenzene molecule. The mean absolute relative deviations (MARD) appear in parenthesis near each PAC method.

3. Results

To demonstrate the efficacy of the method, we show in Figure 2 the MPA and iHPA PACs and their ESP correlation plots before and after applying the minimal corrections required for imposing dipole and quadrupole moments (denoted mcD/mcDQ-MPA and mcD/mcDQ-iHPA respectively).[5]

Note that the MPA-ESP has low correlation with the MOL-ESP, as can be evident visually and also by the reported MARD of 2. The mcD corrections improve the ESP but only mcDQ corrections show high-quality ESP (with MARD of 0.05). In accordance with previous reports [19], the iHPA ESP already correlates nicely with the MOL-ESP (MARD of 0.16) but the mcDQ-iHPA improves the correlation significantly and the MARD reduces by a factor of 4. For this molecule, both mcDQ-MPA and mcDQ-iHPA have similar MARDs but this is not typical; for most molecules, the mcDQ-iHPA MARDs are much smaller than those of mcDQ-MPA (see Figure 3). The mcDQ-MPA PACs are not drastically different from the MPA PACs, yet their MARDs are considerably lower. This shows the power of the minimally corrected PACs, where a small change in PACs can improve the PAC-based ESP considerably.

In Figure 3, we display a log-scale bar-plot of MARDs of several PAC-based potentials on selected molecules containing 10–18 atoms. Each PAC method can be characterised by a pair of numbers (shown in parenthesis within the legend box) indicating the median/maximal MARD taken over the given set of molecules. The PACs obtained by minimally correcting the $q \equiv 0$ reference (called 0PA) are actually the minimal PACs that give the dipole and quadrupole of the molecules. It is seen that their correlation with the exact ESP is considerably higher than that of MPA and HPA, somewhat similar to that of mcDQ-MPA and mcDQ-HPA, and close to that of CM5. This goes to show that the fit of just the dipole and quadrupole, keeping the charges as small as possible, gives a reasonably behaved ESP, although in general, for very large molecules the mcDQ-0PA performance may degrade with size compared to the PA methods. We see that MPA and HPA have similar MARDs while iHPA seems to give considerably smaller MARDs (by a factor of 2–3). The minimally corrected (mcDQ) to MPA and HPA yield smaller MARDs by a factor of 4 and for iHPA by a factor of 2. Altogether, the mcDQ significantly improves the ESP. The mcDQ-iHPA median MARD is 7% which is similar to that of ChElPG (5%).

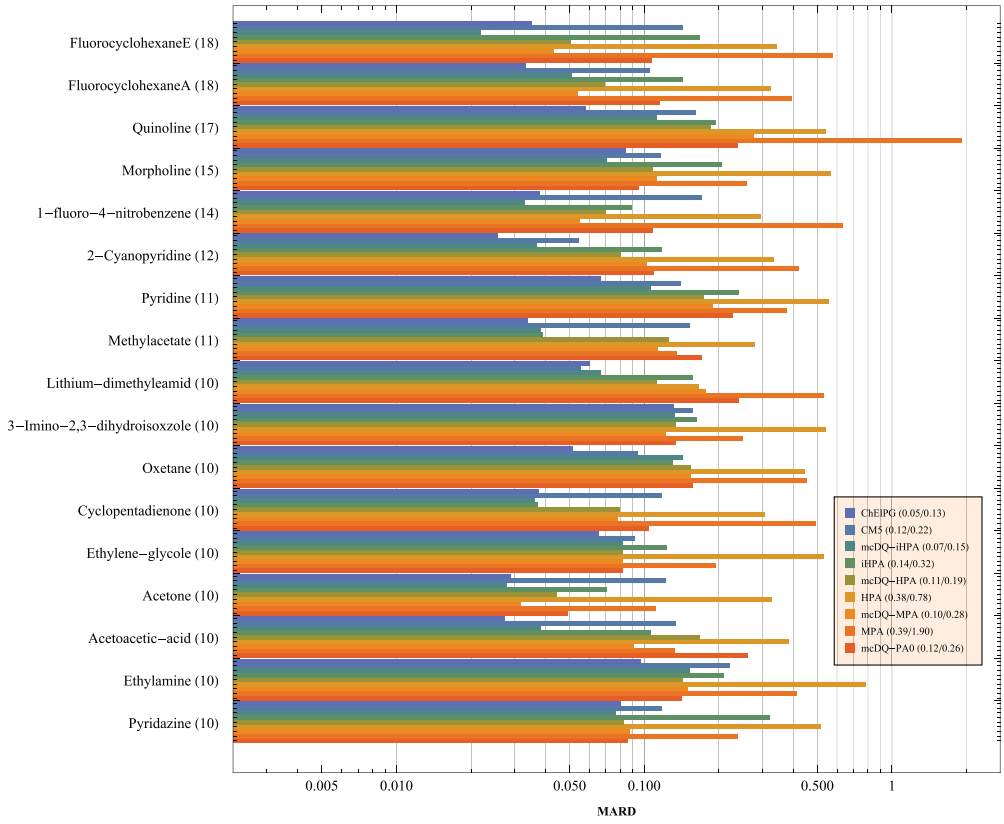


Figure 3. (Colour online) The MARD (Equation (5)) of various PACs schemes for a subset of molecule taken from [20]. Numbers in parenthesis appearing near the molecule names indicate the number of atoms in that molecule. The pair of numbers (median/max) appearing in the legend box near each scheme is, respectively, the median and maximum of the relative deviance taken over the shown set molecules.

It is worthwhile to examine the sensitivity of the MARD estimation with respect to the distance of grid points from the nearest nuclei. In Figure 3, all sampling grid points were at a distance larger than $1.5 \times r_{vdW}$ from any atom. When MARD is estimated using points further way (distance larger than a value of $2 \times r_{vdW}$), the iHPA MARD dropped from 0.14 to 0.09 and mcDQ-iHPA MARD dropped from 0.07 to 0.03. ChElPG MARD also reduced from 0.05 to 0.03. This finding is consistent with the fact that the MCDQ methods provide an asymptotically exact far-field ESP resulting from their reconstruction of the molecular dipole and quadrupole moments.

In Table 1, we show, for each set of PACs, the magnitude of the charge correction $\|\Delta q\|_\infty$. For a given molecule, the mcD correction is largest for 0PA and then for MPA and HPA and it is smallest for iHPA. mcDQ corrections are in general considerably larger than mcD but in both methods $\|\Delta q\|_\infty$ decreases as the number of atoms in the molecule grows. This is due to the fact that in large systems, even small charge shifts have a large affect on the dipole and the quadrupole moments.

In Table 2, we summarise the MARD statistics (median and maximal) for four sets of reference charges: 0PA (reference charges are equal to zero) and MPA, HPA, iHPA. The efficiency of the mc procedure is apparent for MPA, HPA and iHPA, where the mcD reduces the median/maximal MARD by about a factor of 2. mcDQ reduces the MARD further, by a factor of 3 for 0PA and ∼2 for MPA and HPA and only 1.1 for iHPA. We thus see that iHPA reconstruction of the ESP strongly benefits from a dipole correction and, interestingly, much less a quadrupole correction.

PACs are sometimes used when molecules distort. In this case, it is important that they remain continuous under the distortion, so as to enable force calculations. The MPA/HPA/iHPA do not show non-smooth behaviour and the mcDQ which is a minimisation procedure does not show it as well.[6] In Figure 4, we show the MPA, mcDQ-MPA and ChElPG PACs of the oxygen atom in *N*-methylethanamide [30] as a function of the dihedral angle ϕ. It is seen that as the angle increases from 0, the mcDQ-MPA PAC slightly decreases and then increases rapidly followed by a rapid yet continuous drop near $\phi_c = 0.25$ from a value of $q_O = -0.25$ to $q_O \approx -0.64$. An additional very sharp feature is seen near $\phi = \pi$. We have checked that this sharp feature is not discontinuous (see inset in Figure 4) and that the matrix S of Equation (9) does not become rank deficient. A similar behaviour is seen for the PACs of other atoms. We thus

Table 1. The PAC change $\|\Delta q\|_\infty = max_{1 \le a \le A}|\Delta q_a|$ induced by mcD and mcDQ for 0PA (where the reference PACs are all zero), MPA, HPA and iHPA for the set of molecules of Figure 3. Also shown are the number of atoms A, the number of degrees of freedom F and the number of constraints C for each molecule.

					$\|\Delta q\|_\infty$ (mcD)				$\|\Delta q\|_\infty$ (mcDQ)			
Molecule	Sym	A	F	C	0PA	MPA	HPA	iHPA	0PA	MPA	HPA	iHPA
Pyridazine	C_{2v}	10	7	6	0.11	0.02	0.06	0.04	0.27	0.29	0.17	0.12
Ethylamine	C_s	10	7	6	0.04	0.01	0.03	0.01	1.17	1.61	1.22	1.28
Acetoacetic acid	C_1	10	9	6	0.07	0.02	0.03	0.01	0.54	0.26	0.25	0.21
Acetone	C_{2v}	10	5	5	0.17	0.02	0.05	0.01	0.52	0.48	0.34	0.08
Ethylene-glycol	C_1	10	10	9	0.09	0.03	0.05	0.02	0.67	0.88	0.64	0.54
Cyclopentadienone	C_{2v}	10	6	6	0.09	0.04	0.03	0.00	0.19	0.22	0.05	0.02
Oxetane	C_s	10	7	6	0.08	0.05	0.03	0.01	1.09	1.29	1.09	1.09
3-Imino-2,3-dihydroisoxzole	C_s	10	8	6	0.05	0.02	0.02	0.01	0.79	0.83	0.69	0.41
Lithium-dimethylamine	C_{2v}	10	6	4	0.31	0.16	0.05	0.04	0.31	0.16	0.05	0.04
Methyl-acetate	C_s	11	9	6	0.10	0.02	0.02	0.01	0.53	0.22	0.26	0.01
Pyridine	C_{2v}	11	7	4	0.06	0.02	0.03	0.01	0.15	0.23	0.09	0.03
2-Cyanopyridine	C_s	12	12	6	0.12	0.05	0.04	0.01	0.19	0.19	0.07	0.03
1-Fluoro-4-nitrobenzene	C_{2v}	14	9	6	0.05	0.06	0.01	0.01	0.13	0.20	0.05	0.04
Morpholine	C_s	15	9	6	0.03	0.01	0.02	0.01	0.47	0.04	0.33	0.08
Quinoline	C_s	17	17	6	0.03	0.12	0.02	0.01	0.10	0.26	0.05	0.02
Fluorocyclohexane (A)	C_s	18	12	6	0.05	0.03	0.02	0.01	0.17	0.06	0.05	0.02
Fluorocyclohexane (E)	C_s	18	12	6	0.04	0.03	0.01	0.01	0.17	0.08	0.07	0.04
Median					**0.07**	**0.03**	**0.03**	**0.01**	**0.31**	**0.23**	**0.17**	**0.04**
Max					**0.31**	**0.16**	**0.06**	**0.04**	**1.17**	**1.61**	**1.22**	**1.28**

Table 2. The median/maximal MARD (for the set of molecules used above) determined for each PAC reference found for non-corrected and minimally corrected schemes, mcD and mcDQ.

	0PA	MPA	HPA	iHPA
No correction	NA	0.39/1.90	0.38/0.78	0.14/0.32
mcD	0.41/0.67	0.18/0.60	0.18/0.53	0.08/0.17
mcDQ	0.12/0.26	0.10/0.28	0.11/0.19	0.07/0.15

conclude that the charges change continuously although sometimes very rapid charge fluctuations can occur.

4. Summary and conclusions

We have studied a new scheme for minimally correcting reference PACs so that they reproduce the exact dipole and quadrupole moments of a molecule and we found that such a minimal correction greatly improves the correlation of the PAC-ESP with respect to the MOL-ESP. The minimal correction scheme does not alter symmetry properties of the reference PACs. Hence, minimally corrected PACs (mc-PACs) based on MPA, HPA and iHPA fully respect the point-symmetry and rotational/translational symmetries of the molecule.

An additional benefit of the mc-PACs is their stability for inner (or buried) atoms of large molecules. This rises from the stability of the standard population schemes themselves and the fact that mc-PACs involve rather small corrections. As an example, consider the 2-(dimethylamino)-2-propanol molecule:

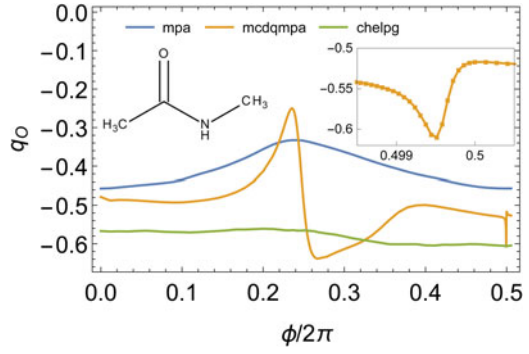
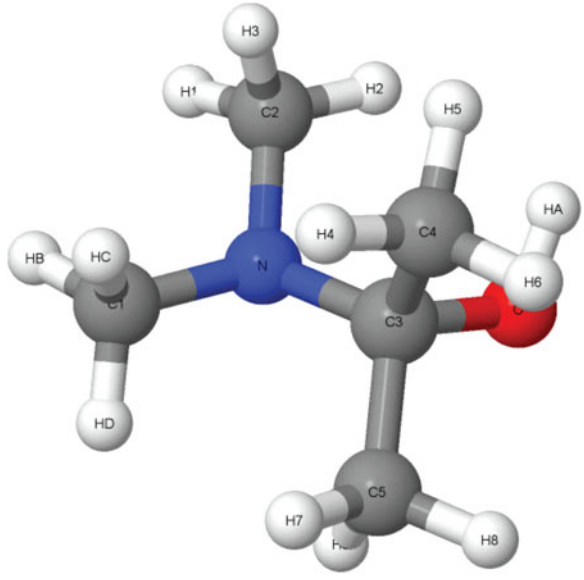


Figure 4. (Colour online) The partial charge, determined by MPA, mcDQ-MPA and ChElPG on the oxygen atom as a function of the O–C–N–H dihedral angle ϕ in *N*-methylethanamide (NMA). Inset on the right shows the sharp feature near $\phi = \pi$.



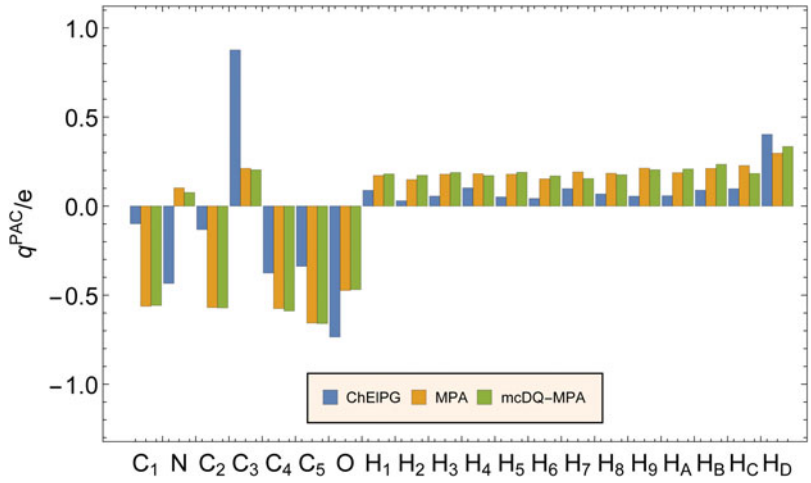


Figure 5. (Colour online) The ChElPG, MPA and mc-MPA PACs for the 2-(dimethylamino)-2-propanol.

for which the ChElPG, HPA and mcDQ-HPA PACs are shown in Figure 5. Here, ChElPG tends to polarise the molecule: the oxygen and nitrogen share between them a negative unit charge and this is counteracted by the positive unit charge of the central carbon atom C3. On the other hand, MPA assigns a low charge for C3 and spreads rather evenly the remaining positive charge on the 12 terminal hydrogen atoms. mc-MPA charges are very close to those of MPA and thus yet they improve significantly the ESP description for this molecule: the MPA MARD is 0.35 while that of the mc-MPA is 0.1. It is worthwhile to note that the PACs assigned by ChElPG also have a MARD of 0.1.

When the underlying reference is the iHPA set of PACs, the resulting ESP is of similar quality to that of the ChElPG set of PACs resulting from a best-fit to ESPs. The dependence of the PACs on the molecular distortion was demonstrated to have sometimes very sharp features; however, all the changes were smooth, and hence forces can be calculated on the atoms of the molecule.

The method here bears a similarity to the optimal point-charge model of [10] which determines PACs that reproduce as many low-order moments as possible. The crucial difference is best seen when systems grow; the model of [10] would target increasingly higher ESMs as more atoms are included while the present method targets multipoles up to second order and not beyond, thereby avoiding the numerical instabilities described in see [11]. The present approach avoids this problem by capping the highest reconstructed moment order to 2. This type of order-capping was first proposed in [37], although its implementation was based on a different approach than here. On the other hand, the optimal point-charge model treats the multipole constraints in a more systematic way by minimising the error over unused moments in the last incomplete spherical shell.

Notes

1. When defining the moments, it is customary to take the origin in the centre of the positive charge distribution.
2. See [8]; we use the Einstein convention by which repeated Cartesian indices are summed over.
3. Note that the expression in Equation (5) cannot become singular due to this requirement.
4. Since the matrix is dominated by zeros, one can formulate the linear equation in a more concise way. However, this form is straightforward to derive and manipulate when there are instabilities, discussed later.
5. Minimally corrected PACs that reproduce only the dipole moment are designated mcD and those that reproduce the components of the dipole and the quadrupole ESMs are designated mcDQ.
6. We cannot rule out possible issues if the matrix S of Equation (9) becomes rank deficient. However, we believe this is an unlikely or quite rare event.

Acknowledgments

Authors express special thanks to Dr Yihan Shao from Q-CHEM Inc. for his advice and critical assistance in performing the iHPA calculations.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The study was supported by the Israel Science Foundation [grant number 189/14].

References

[1] S. Lifson and A. Warshel, J. Chem. Phys. **49**, 5116 (1968).
[2] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[3] N.L. Allinger, Y.H. Yuh, and J.H. Lii, J. Am. Chem. Soc. **111**, 8551 (1989).
[4] M.J. Field, P.A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[5] E.M. Duffy and W.L. Jorgensen, J. Am. Chem. Soc. **122**, 2878 (2000).
[6] P. Politzer and D.G. Truhlar, *Chemical Applications of Atomic and Molecular Electrostatic Potentials: Reactivity, Structure, Scattering, and Energetics of Organic, Inorganic, and Biological Systems* (Springer Science & Business Media, New York, 2013).
[7] Y. Mei, A.C. Simmonett, F.C. Pickard, IV, R.A. DiStasio, Jr, B.R. Brooks, and Y. Shao, J. Phys. Chem. A **119**, 5865 (2015).
[8] J.D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999).
[9] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh,L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, and P.W. Ayers, *arXiv preprint arXiv:1608.05556* 2016.
[10] A.C. Simmonett, A.T. Gilbert, and P.M. Gill, Mol. Phys. **103**, 2789 (2005).
[11] A. Gilbert and P. Gill, Mol. Simul. **32**, 1249 (2006).
[12] R.S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
[13] P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).
[14] F. Hirshfeld, Theor. Chim. Acta **44**, 129 (1977).
[15] J. Foster and F. Weinhold, J. Am. Chem. Soc. **102**, 7211 (1980).
[16] U.C. Singh and P.A. Kollman, J. Comput. Chem. **5**, 129 (1984).
[17] P. Bultinck, C. Van Alsenoy, P.W. Ayers, and R. Carbó-Dorca, J. Chem. Phys. **126**, 144111 (2007).
[18] P. Bultinck, D.L. Cooper, and D. Van Neck, Phys. Chem. Chem. Phys. **11**, 3424 (2009).
[19] S. Van Damme, P. Bultinck, and S. Fias, J. Chem. Theory Comput. **5**, 334 (2009).
[20] A.V. Marenich, S.V. Jerome, C.J. Cramer, and D.G. Truhlar, J. Chem. Theory Comput. **8**, 527 (2012).
[21] C.M. Breneman and K.B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[22] L.E. Chirlian and M.M. Francl, J. Comput. Chem. **8**, 894 (1987).
[23] F.A. Momany, J. Phys. Chem. **82**, 592 (1978).
[24] B. Besler, K. Merz, Jr, and P. Kollman, J. Comput. Chem. **11**, 431 (1990).
[25] C.I. Bayly, P. Cieplak, W. Cornell, and P.A. Kollman, J. Phys. Chem. **97**, 10269 (1993).
[26] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollmann, J. Am. Chem. Soc. **115**, 9620 (1993).
[27] D.E. Williams, J. Comput. Chem. **9**, 745 (1988).
[28] E. Sigfridsson and U. Ryde, J. Comput. Chem. **19**, 377 (1998).
[29] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Phys. Chem. B **106**, 7300 (2002).
[30] H. Hu, Z. Lu, and W. Yang, J. Chem. Theory Comput. **3**, 1004 (2007).
[31] M.M. Francl and L.E. Chirlian, Rev. Comput. Chem. **14**, 1 (2000).
[32] S. Jakobsen and F. Jensen, J. Chem. Theory Comput. **12**, 1824 (2016).
[33] Y. Shao, Z. Gan, E. Epifanovsky, A.T. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P.R. Horn, L.D. Jacobson, I. Kaliman, R.Z. Khaliullin, T. Kus, A. Landau, J. Liu, E.I. Proynov, Y.M. Rhee, R.M. Richard, M.A. Rohrdanz, R.P. Steele, E.J. Sundstrom, III H. Lee Woodcock, P.M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G.J.O. Beran, Y.A. Bernard, E. Berquist, K. Brandhorst, K.B. Bravaya, S.T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S.H. Chien, K.D. Closser, D.L. Crittenden, M. Diedenhofen, R.A.D. Jr., H. Do, A.D. Dutoi, R.G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M.W. Hanson-Heine, P.H. Harbach, A.W. Hauser, E.G. Hohenstein, Z.C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R.A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C.M. Krauter, K.U. Lao, A.D. Laurent, K.V. Lawler, S.V. Levchenko, C.Y. Lin, F. Liu, E. Livshits, R.C. Lochan, A. Luenser, P. Manohar, S.F. Manzer, S.-P. Mao, N. Mardirossian, A.V. Marenich, S.A. Maurer, N.J. Mayhall, E. Neuscamman, C.M. Oana, R. Olivares-Amaya, D.P. O'Neill, J.A. Parkhill, T.M. Perrine, R. Peverati, A. Prociuk, D.R. Rehn, E. Rosta, N.J. Russ, S.M. Sharada, S. Sharma, D.W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A.J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M.A. Watson, J. Wenzel, A. White, C.F. Williams, J. Yang, S. Yeganeh, S.R. Yost, Z.-Q. You, I.Y. Zhang, X. Zhang, Y. Zhao, B.R. Brooks, G.K. Chan, D.M. Chipman, C.J. Cramer, III William A. Goddard, M.S. Gordon, W.J. Hehre, A. Klamt, III Henry F. Schaefer, M.W. Schmidt, C.D. Sherrill, D.G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A.T. Bell, N.A. Besley, J.-D. Chai, A. Dreuw, B.D. Dunietz, T.R. Furlani, S.R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D.S. Lambrecht, W. Liang, C. Ochsenfeld, V.A. Rassolov, L.V. Slipchenko, J.E. Subotnik, T.V. Voorhis, J.M. Herbert, A.I. Krylov, P.M. Gill and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).
[34] Y. Zhao and D.G. Truhlar, J. Chem. Phys. **125**, 194101 (2006).
[35] B.J. Lynch, Y. Zhao, and D.G. Truhlar, J. Phys. Chem. A **107**, 1384 (2003).
[36] G.H. Golub and C.F. van Loan, *Matrix Computations*, 3rd ed. (The John Hopkins University Press, Baltimore, 1996).
[37] M. Swart, P.T. van Duijnen, and J.G. Snijders, J. Comput. Chem. **22**, 79 (2001).