Density Fitting for QC/MM Interactions

Martin J. Field

Density Fitting for QC/MM Interactions

Martin J. Field*,†,‡

†*Laboratoire de Chimie et Biologie des Métaux, UMR5249, Université Grenoble I, CEA, CNRS, 17 avenue des Martyrs, 38054 Cedex 9, Grenoble, France*

‡*Theory Group, Institut Laue-Langevin, 71 avenue des Martyrs CS 20156, 38042 Cedex 9, Grenoble, France*

E-mail: martin.field@cea.fr,field@ill.fr

Abstract

Density fitting is a standard technique in quantum chemistry as it helps to accelerate certain parts of a calculation, such as the computation of the electron repulsion energy, without significant loss of accuracy. This paper explores the effectiveness of this technique when it is extended to treat interactions with external electrostatic potentials, in particular those that arise from the environment in hybrid quantum chemical/molecular mechanical calculations. It is found that fitted densities are able to well reproduce the energies, forces and properties obtained using un-fitted densities, as long as a suitable operator is employed for the fitting. It is expected that this precision would be improved by the development of basis sets specifically designed to treat these types of interactions, and that the use of this approximation could lead to substantial speed-ups in large hybrid potential simulations.

Introduction

Density fitting is a standard tool in quantum chemistry. Some of the first references to it are in the works of Baerends, Ellis and Ros,[1] who used it to speed up the calculation of the electron repulsion energy in Hartree-Fock-Slater calculations, and of Whitten and Jafri,[2,3] who were interested in approximating Coulombic potential energy integrals. Other significant early works include that of Sambe and Felton[4,5] who extended the fitting idea to approximating the exchange potential in Slater's SCF-Xα equations, and of Dunlap, Connolly and Sabin[6] who introduced robust variational Coulomb fitting. Since these early studies, developments of the method have mostly concentrated on the area in which it provides the most benefit, namely acceleration of the two-electron contributions to the electronic energy for a range of QC approaches, although there have been exceptions (see, for example, Köster et al.[7]).

In this article the use of density fitting approximations is examined within the context of hybrid potential quantum chemical/molecular mechanical (QC/MM) potentials[8] for the

calculation of the electrostatic interactions between the QC and MM atoms. The rationale for doing this is two-fold. First, it could potentially lead to significant speed-ups in QC/MM calculations with large numbers of MM atoms if relatively small fit basis sets of sufficient precision could be developed and, second, it would bridge the obvious, and needless, discrepancy that exists between the sophistications of the electrostatic representations of the QC and MM regions, especially when QC methods with large basis sets are being employed. In spirit the approach taken in this paper can be seen to be most akin to existing schemes that also approximate the QC electron density "on-the-fly" using, for example, multipolar representations.[9,10]

Methods

This section starts with a brief overview of the density fitting method, before going on to describe its use in the calculation of the electrostatic interactions between QC and MM atoms in hybrid potential calculations.

Density Fitting

In standard QC methods, the electron density, ρ, is expressed in terms of orbital basis functions, χ_μ, as follows:

$$\rho(\boldsymbol{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu(\boldsymbol{r})\chi_\nu(\boldsymbol{r}) \tag{1}$$

where P_{μ_ν} are elements of the density matrix.

The idea of density fitting is to approximate the full density by a linear expansion in terms of a set of auxiliary functions, η_f:

$$\tilde{\rho}(\boldsymbol{r}) = \sum_f a_f \eta_f(\boldsymbol{r}) \tag{2}$$

and then to determine the expansion coefficients, a_f, by minimizing the function:

$$\mathcal{F} = \langle \rho(\boldsymbol{r}_1) - \tilde{\rho}(\boldsymbol{r}_1) | \hat{\mathcal{O}}_{12} | \rho(\boldsymbol{r}_2) - \tilde{\rho}(\boldsymbol{r}_2) \rangle \tag{3}$$

where the angled brackets indicate integration over the two-electron operator $\hat{\mathcal{O}}_{12}$. The most commonly used operators are the Coulomb operator, $1/r_{12}$,[2,3,6] and the overlap operator, $\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$,[1] although others, such as the anti-Coulomb operator, $-r_{12}$, of Gill et al.[11] and the attenuated Coulomb operator, $\mathrm{erfc}\left(\omega r_{12}\right)/r_{12}$, of Jung et al.,[12] have also been proposed. The Coulomb, overlap and anti-Coulomb operators minimize the energy, overlap and potential of the charge distribution, respectively, whereas the attenuated Coulomb operator was designed to mimic the Coulomb operator at short ranges and the overlap operator at longer range.

The minimization of $\mathcal{F}$ is often performed subject to a series of constraint conditions on the density which, to be tractable, should be linear functions of the expansion coefficients of the auxiliary functions, a_f. The most common constraint is on the density's charge, Q, which can be expressed as:

$$\Lambda_Q = \int \rho(\boldsymbol{r}) d\boldsymbol{r} - \int \tilde{\rho}(\boldsymbol{r}) d\boldsymbol{r} = Q - \tilde{Q} = 0 \tag{4}$$

Fitting is typically done with respect to full electron densities whose charges are constant. However, it is also possible to fit partial densities whose charges may vary (see, for example, Reine et al.[13]), although these cases are not considered further here.

It is convenient to recast equations 3 and 4 in matrix-vector form as follows:

$$\mathcal{F} = \langle \rho | \hat{\mathcal{O}}_{12} | \rho \rangle + \boldsymbol{A}^T \mathbf{T} \boldsymbol{A} - 2\boldsymbol{A}^T \boldsymbol{B} \tag{5}$$

$$\boldsymbol{B} = \mathbf{U}\boldsymbol{P} \tag{6}$$

$$Q = \boldsymbol{P}^T \boldsymbol{S} \tag{7}$$

$$\tilde{Q} = \boldsymbol{A}^T \boldsymbol{F} \tag{8}$$

In these equations, $\boldsymbol{A}$ is the vector of auxiliary function expansion coefficients, $\boldsymbol{P}$ is the vector of density matrix elements, $\boldsymbol{F}$ is the vector of auxiliary function self-overlaps, $\boldsymbol{S}$ is the orbital function overlap vector, $\mathbf{T}$ is the matrix of auxiliary function fitting integrals, and $\mathbf{U}$ is the matrix of auxiliary function-orbital function fitting integrals. The expressions for the integrals are:

$$F_f = \int \eta_f(\boldsymbol{r}) d\boldsymbol{r} \tag{9}$$

$$S_{\mu\nu} = \int \chi_\mu(\boldsymbol{r})\chi_\nu(\boldsymbol{r}) d\boldsymbol{r} \tag{10}$$

$$T_{fg} = \langle \eta_f(\boldsymbol{r}) | \hat{\mathcal{O}}_{12} | \eta_g(\boldsymbol{r}) \rangle \tag{11}$$

$$U_{f,\mu\nu} = \langle \eta_f(\boldsymbol{r}) | \hat{\mathcal{O}}_{12} | \chi_\mu(\boldsymbol{r})\chi_\nu(\boldsymbol{r}) \rangle \tag{12}$$

Using the method of Lagrange multipliers, minimization of $\mathcal{F}$ from equation 5 with respect to $\boldsymbol{A}$, but subject to the charge constraint condition, Λ_Q, leads to a set of linear equations of the form:

$$\begin{bmatrix} \mathbf{T} & \boldsymbol{F} \\ \mathbf{F}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{A} \\ -\frac{\lambda}{2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B} \\ Q \end{bmatrix} \tag{13}$$

where λ is the constraint's Lagrange multiplier. The explicit solutions for the variables $\boldsymbol{A}$ and λ are:

$$\frac{\lambda}{2} = \frac{\left(Q - \boldsymbol{F}^T \mathbf{T}^{-1} \boldsymbol{B}\right)}{\boldsymbol{F}^T \mathbf{T}^{-1} \boldsymbol{F}} \tag{14}$$

$$\boldsymbol{A} = \mathbf{T}^{-1}\left(\boldsymbol{B} + \frac{\lambda}{2}\boldsymbol{F}\right) \tag{15}$$

In what follows, it is convenient to rewrite equation 13 more compactly as:

$$\mathbf{M}\boldsymbol{A}_\lambda = \boldsymbol{B}_\lambda \tag{16}$$

where the extents of the $\boldsymbol{A}$ and $\boldsymbol{B}$ vectors have been extended to accommodate the appro-

priate number of constraint condition quantities.

Coulomb Density Fitting

Dunlap et al.[6] introduced what is now the commonest type of density fitting which approximates the full density using the Coulomb operator for the density fit. This scheme has a number of formal and practical advantages, including stability, accuracy and efficiency as the integrals required for the fitting can also be employed for determining the electron repulsion energy.

In this scheme the electron repulsion energy is approximated as:

$$\begin{aligned} E_{\mathrm{C}} &= \frac{1}{2}\langle\rho|\frac{1}{r_{12}}|\rho\rangle \\ &\sim \langle\rho|\frac{1}{r_{12}}|\tilde{\rho}\rangle - \frac{1}{2}\langle\tilde{\rho}|\frac{1}{r_{12}}|\tilde{\rho}\rangle \end{aligned} \tag{17}$$

In matrix form this becomes:

$$E_{\mathrm{C}} = \boldsymbol{A}^T\boldsymbol{B} - \frac{1}{2}\boldsymbol{A}^T\mathbf{T}\boldsymbol{A} \tag{18}$$

$$= \frac{1}{2}\left(\boldsymbol{A}^T\boldsymbol{B} - \frac{\lambda}{2}Q\right) \tag{19}$$

$$= \frac{1}{2}\boldsymbol{A}_\lambda^T\boldsymbol{B}_\lambda \tag{20}$$

In calculations with density fitting, the derivatives of the energy with respect to both $\boldsymbol{P}$ and the atomic coordinates, $\boldsymbol{R}$, are normally required. In the first case, this is E_{C}'s contribution to the Fock matrix, which can be derived by differentiating equation 20 and then determining the derivatives of $\boldsymbol{A}$ with respect to $\boldsymbol{P}$ by differentiating equation 16, remembering that $\mathbf{M}$ is independent of $\boldsymbol{P}$. This gives:

$$\frac{dE_{\mathrm{C}}}{d\boldsymbol{P}} = \boldsymbol{A}_\lambda^T\left(\frac{\partial\boldsymbol{B}_\lambda}{\partial\boldsymbol{P}}\right) \tag{21}$$

In this equation, the derivatives of the vector $\boldsymbol{B}$ are elements of the matrix $\mathbf{U}$ (equation 6), whereas the derivative of Q will be zero if the charge is constant, but equal to the orbital function overlap, $\boldsymbol{S}$, otherwise (equation 7).

In a similar fashion, the coordinate derivatives are:

$$\frac{dE_{\mathrm{C}}}{d\boldsymbol{R}} = \frac{\partial E_{\mathrm{C}}}{\partial \boldsymbol{R}} + \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{R}}\right)^{T} \frac{\partial E_{\mathrm{C}}}{\partial \boldsymbol{P}} + \left(\frac{\partial \boldsymbol{A}_{\lambda}}{\partial \boldsymbol{R}}\right)^{T} \frac{\partial E_{\mathrm{C}}}{\partial \boldsymbol{A}_{\lambda}} \tag{22}$$

In this equation, the second term on the right-hand side is E_{C}'s contribution to the standard Pulay gradient term, which integrates seamlessly with the Pulay terms of the other contributions to the total energy and so is handled automatically, whereas the third term can be determined in exactly the same way as the derivatives of $\boldsymbol{A}$ with respect to $\boldsymbol{P}$. Gathering all terms other than the Pulay contribution gives:

$$\frac{dE_{\mathrm{C}}}{d\boldsymbol{R}} = \boldsymbol{A}_{\lambda}^{T} \left(\frac{\partial \boldsymbol{B}_{\lambda}}{\partial \boldsymbol{R}}\right) - \frac{1}{2} \boldsymbol{A}_{\lambda}^{T} \left(\frac{\partial \mathbf{M}}{\partial \boldsymbol{R}}\right) \boldsymbol{A}_{\lambda} \tag{23}$$

Similar to before, the derivatives of the vector $\boldsymbol{B}$ involve products of $\boldsymbol{P}$ and derivatives of $\mathbf{U}$ (equation 6), whereas the derivative of Q will be zero if the charge is constant, but equal to products of $\boldsymbol{P}$ and derivatives of $\boldsymbol{S}$ otherwise (equation 7).

Non-Coulomb Density Fitting

If operators other than the Coulomb operator are employed for fitting the density, then the electron repulsion energy can be approximated as:

$$\begin{aligned} E_{\mathrm{C}} &\sim \frac{1}{2} \langle \tilde{\rho} | \frac{1}{r_{12}} | \tilde{\rho} \rangle & (24) \\ &= \frac{1}{2} \boldsymbol{A}^{T} \mathbf{T}_{\mathrm{C}} \boldsymbol{A} & (25) \\ &= \frac{1}{2} \boldsymbol{A}_{\lambda}^{T} \mathbf{M}_{\mathrm{C}} \boldsymbol{A}_{\lambda} & (26) \end{aligned}$$

where $\mathbf{T}_{\mathrm{C}}$ is the matrix of Coulomb integrals over the auxiliary functions, which is different from the $\mathbf{T}$ matrix used for the fitting, and $\mathbf{M}_{\mathrm{C}}$ is a "convenience" $\mathbf{M}$ matrix that is identical to $\mathbf{T}_{\mathrm{C}}$ but expanded with an appropriate number of zero rows and columns.

The derivatives of the energy can be evaluated in the same way as for Coulomb fitting, giving the following expressions for the $\boldsymbol{P}$ derivatives and the $\boldsymbol{R}$ derivatives without the Pulay contribution, respectively:

$$\frac{dE_{\mathrm{C}}}{d\boldsymbol{P}} = \boldsymbol{A}_\lambda^T \mathbf{M}_{\mathrm{C}} \mathbf{M}^{-1} \left(\frac{\partial \boldsymbol{B}_\lambda}{\partial \boldsymbol{P}} \right) \tag{27}$$

$$\frac{dE_{\mathrm{C}}}{d\boldsymbol{R}} = \frac{1}{2} \boldsymbol{A}_\lambda^T \left(\frac{\partial \mathbf{M}_{\mathrm{C}}}{\partial \boldsymbol{R}} \right) \boldsymbol{A}_\lambda + \boldsymbol{A}_\lambda^T \mathbf{M}_{\mathrm{C}} \mathbf{M}^{-1} \left[\left(\frac{\partial \boldsymbol{B}_\lambda}{\partial \boldsymbol{R}} \right) - \left(\frac{\partial \mathbf{M}}{\partial \boldsymbol{R}} \right) \boldsymbol{A}_\lambda \right] \tag{28}$$

Interactions with an External Potential

The interaction of the electron density with an external potential, $V(\boldsymbol{r})$, can also be approximated using a fitted density as follows:

$$\begin{aligned} E_{\mathrm{V}} &= \int \rho(\boldsymbol{r}) V(\boldsymbol{r}) d\boldsymbol{r} \\ &\sim \int \tilde{\rho}(\boldsymbol{r}) V(\boldsymbol{r}) d\boldsymbol{r} \\ &= \sum_f a_f \int \eta_f(\boldsymbol{r}) V(\boldsymbol{r}) d\boldsymbol{r} \\ &= \boldsymbol{A}_\lambda^T \boldsymbol{V}_\lambda \end{aligned} \tag{29}$$

where $\boldsymbol{V}_\lambda$ is the vector of auxiliary fitting function integrals with the potential extended with the necessary number of zeros corresponding to the number of constraint variables.

The derivatives of the energy can be evaluated in the same way as in the previous sections, giving the following expressions for the $\boldsymbol{P}$ derivatives and the $\boldsymbol{R}$ derivatives without the Pulay

contribution, respectively:

$$\frac{dE_{\mathrm{V}}}{d\boldsymbol{P}} = \boldsymbol{W}^T \left(\frac{\partial \boldsymbol{B}_\lambda}{\partial \boldsymbol{P}} \right) \tag{30}$$

$$\frac{dE_{\mathrm{V}}}{d\boldsymbol{R}} = \boldsymbol{A}_\lambda^T \frac{\partial \boldsymbol{V}_\lambda}{\partial \boldsymbol{R}} + \boldsymbol{W}^T \left[\left(\frac{\partial \boldsymbol{B}_\lambda}{\partial \boldsymbol{R}} \right) - \left(\frac{\partial \mathbf{M}}{\partial \boldsymbol{R}} \right) \boldsymbol{A}_\lambda \right] \tag{31}$$

where the vector $\boldsymbol{W}$ is equal to $\mathbf{M}^{-1}\boldsymbol{V}_\lambda$. Note that the expression for the $\boldsymbol{P}$ derivatives is independent of $\boldsymbol{P}$ and so need only be evaluated once at the start of a QC calculation.

The derivations in this section are independent of whether the Coulomb energy, E_{C}, is fitted and, if so, that it is fitted with the Coulomb operator and the same auxiliary basis set. Supposing that it is, though, then the coordinate derivatives for both energies can be conveniently combined as follows:

$$\frac{dE_{\mathrm{C+V}}}{d\boldsymbol{R}} = \boldsymbol{A}_\lambda^T \frac{\partial \boldsymbol{V}_\lambda}{\partial \boldsymbol{R}} + (\boldsymbol{A}+\boldsymbol{W})^T \left(\frac{\partial \boldsymbol{B}_\lambda}{\partial \boldsymbol{R}} \right) + \frac{1}{2} \boldsymbol{W}^T \frac{\partial \mathbf{M}}{\partial \boldsymbol{R}} \boldsymbol{W} - \frac{1}{2} (\boldsymbol{A}+\boldsymbol{W})^T \frac{\partial \mathbf{M}}{\partial \boldsymbol{R}} (\boldsymbol{A}+\boldsymbol{W}) \tag{32}$$

Hybrid Potential QC/MM Calculations

In the most common type of QC/MM calculation, a system is divided into two subsystems comprising a QC region that is embedded in a surrounding MM region. The atoms in each region interact, typically via a combination of non-bonded electrostatic and Lennard-Jones terms, and of covalent terms if there are bonds between the QC and MM atoms. The covalent and Lennard-Jones interactions are often standard MM terms, whereas the electrostatic interactions are most commonly evaluated using a point charge representation for the MM atoms:

$$E_{\mathrm{QC/MM}}^{\mathrm{Electrostatic}} = \sum_m q_m \left(\sum_\alpha \frac{Z_\alpha}{|\boldsymbol{r}_m - \boldsymbol{r}_\alpha|} - \int d\boldsymbol{r} \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}_m - \boldsymbol{r}|} \right) \tag{33}$$

In this equation, the MM atoms, m, have point charges, q_m, and positions, $\boldsymbol{r}_m$, whereas the QC atoms, α, have nuclear (core) charges, Z_α, positions, $\boldsymbol{r}_\alpha$, and electron density, $\rho(\boldsymbol{r})$.

Following equation 29, the external potential interacting with the electron density due to the MM charges, together with its energy and derivative with respect to the electron density, can be written as:

$$V(\boldsymbol{r}) = -\sum_m \frac{q_m}{|\boldsymbol{r}_m - \boldsymbol{r}|} \tag{34}$$

$$E_{\mathrm{V}} = \mathrm{Tr}\left\{\mathbf{PV}\right\} \tag{35}$$

$$\frac{dE_{\mathrm{V}}}{d\boldsymbol{P}} = \mathbf{V} \tag{36}$$

where the integrals in the matrix $\mathbf{V}$ are:

$$V_{\mu\nu} = \int \chi_\mu(\boldsymbol{r}) V(\boldsymbol{r}) \chi_\nu(\boldsymbol{r}) d\boldsymbol{r} \tag{37}$$

Note that the term in equation 36 is added to the QC one-electron matrix so that the electron density of the QC region fully adapts to the MM environment each time a QC/MM energy calculation is performed.

Normally the energy in equation 33 is calculated using the full expression for the electron density of equation 1. This has the advantage that it is "exact" for the QC and MM methods that are being employed but also disadvantages. First, it involves a double sum over QC atom centers that can become the rate-limiting step in QC/MM calculations when there are large numbers of MM atoms (see, for example, Renison et al.[14]). Second, it is a very unbalanced expression when QC methods with large basis sets are being used as the QC and MM charge distributions are treated using representations of very different sophistications.

This begs the question of whether it is worth using the "exact", expensive expression when the charge distribution on the MM atoms is so rudimentary. Instead why not use a fitted electron density of the form of equation 2 to calculate the QC/MM energy? This fitted density need not be of the same accuracy as the one used to calculate the electron repulsion energy, for example, but only sufficient to provide a reasonable representation of the QC charge distribution for calculation of the QC/MM electrostatic interactions. The

use of the fitted density has the advantage that it involves a single sum over the QC atom centers and so is cheaper to calculate than its "exact" counterpart. In addition, it would lend itself more easily to treatment by fast multipole and particle mesh Ewald methods for the computation of electrostatic interactions in extended systems. It is worth underlining here that this work is only interested in density fitting approaches that re-fit the full density every time it is recalculated (as in standard QC density fitting), and it does not consider alternative approximations such as those, for example, that might employ a fixed or constant reference density that is fit only once at the start of a QC/MM simulation.

Calculations on the s101x7 Database

To test the use of a fitted density in hybrid potential calculations, the s101x7 database developed by Wang and co-workers[15] was selected. This consists of 37 common organic monomers that are combined into 101 homo- and hetero-dimers. For each dimer, Wang *et al* determined seven interaction energies, namely those corresponding to 0.70, 0.80, 0.90, 0.95, 1.00, 1.05, and 1.10 times the equilibrium inter-monomer distance. For all dimer calculations the monomer geometries were held fixed and only the inter-monomer degrees of freedom were optimized.

In this article, all calculations that were peformed employed standard QC methods and basis sets taken from the basis set exchange,[16] and publicly-available MM parameters taken from the LigParGen OPLS/CM1A Parameter Generator for Organic Ligands website of Dodda et al.[17]. Undoubtedly the precision of the results that were obtained could be improved by optimizing some of these, especially the density fitting basis sets with the non-Coulomb operators and the inter-monomer MM parameters, but such work is left to the future.

Three QC methods were defined for the calculations. Their definitions, together with the acronyms that are employed for convenience later, are:

BLYP A density functional theory (DFT) method with the BLYP functional, the def2-

sv(p) orbital basis, the def2-sv(p)-rifit density fitting basis, and the Coulomb fitting operator.

B3LYP A DFT method with the B3LYP functional and the def2-sv(p) basis.

HF A Hartree-Fock (HF) method with the def2-sv(p) basis.

In addition, ten different representations of the electron density in a QC calculation were defined, independent of the above QC methods. These were:

Full This is the normal, default method in which the full, un-fitted electron density of equation 1 is employed.

Small/A The electron density is fit using the dgauss-a1-dftjfit (Small) basis set and the anti-Coulomb (A) operator.

Small/C The electron density is fit using the dgauss-a1-dftjfit basis set and the Coulomb (C) operator.

Small/O The electron density is fit using the dgauss-a1-dftjfit basis set and the overlap (O) operator.

Medium/A The electron density is fit using the def2-sv(p)-rifit (Medium) basis set and the anti-Coulomb operator.

Medium/C The electron density is fit using the def2-sv(p)-rifit basis set and the Coulomb operator.

Medium/O The electron density is fit using the def2-sv(p)-rifit basis set and the overlap operator.

Large/A The electron density is fit using the def2-tzvp-rifit (Large) basis set and the anti-Coulomb operator.

Large/C The electron density is fit using the def2-tzvp-rifit basis set and the Coulomb operator.

Large/O The electron density is fit using the def2-tzvp-rifit basis set and the overlap operator.

Three sets of calculations were performed with the database, one on the monomers and two on the dimers. For the monomer calculations, the geometries of the 37 monomers in the database were first optimized at the HF/6-31g$*$ level of theory. Using these geometries, a series of 30 property calculations was then carried out for each monomer employing all possible combinations of the three different QC methods and the ten different electron density representations described above. The properties comprised the atomic charges obtained by partitioning the electron density, the atomic charges obtained by fitting to the density's electrostatic potential (ESP), and the molecular dipole and quadrupole moments. With the Full density representation, charges were determined using the Mulliken method and the other properties in the usual way. For the other, fitted representations, the calculations employed the simpler, one-center expansion of equation 2. Note that in these cases, partitioning of the density into atomic charges is unambiguous and involves no approximation, Mulliken or otherwise, as each density basis function is already localized on a single center. Finally, it should be remarked that although the monomer calculations were not strictly relevant for the purposes of investigating QC/MM interactions, they were done to evaluate the behavior of the different density fitting methods.

The first set of dimer calculations consisted of 43 different interaction curves that were computed at different levels of theory for each dimer. Only single point energy calculations were carried out as the dimer geometries (seven for each dimer) were taken directly from the database. Two of the curves were computed with pure QC methods, namely the HF and BLYP approximations described above, whereas a third curve was computed using a MM force field. The remaining 40 curves were calculated with a QC/MM method, 20 of these in which the first monomer in the dimer was treated as QC and the second MM (de-

noted QC/MM), and 20 in which the second monomer was QC and the first MM (denoted MM/QC). Of each set of 20 curves, 10 were computed with the BLYP QC method, and 10 with the HF QC method. Similarly each set of 10 curves for a given QC method was determined using the same ten representations of the QC monomer electron density as described above.

The second set of dimer calculations were geometry optimizations on a selected subset of 33 dimers in the database that contained at least one water monomer. The optimizations were unconstrained so both intra- and inter-monomer degrees of freedom were allowed to change. Sixteen optimizations were carried out for each dimer that differed in the energy model that was employed. These were: MM; pure QC with the HF method; MM/QC and QC/MM with the HF QC method and interactions treated with the Full un-fitted density representation; and MM/QC and QC/MM with the HF QC method and interactions treated using densities fitted with the Small, Medium and Large basis sets, and both the anti-Coulomb and Coulomb operators. All optimizations were started from the same, 1.00 times the equilibrium inter-monomer distance dimer structures in the database so as to limit possible bias in the final results.

To finish this section, it is worth re-emphasizing that the ways in which the electron density was represented for the calculation of the QC Coulomb energy and for the QC/MM electrostatic interactions were independent, and that, where appropriate, the QC/MM density fitting method was incorporated directly into the QC method's self-consistent field (SCF) procedure. Thus, for the monomer QC property calculations, where there are no QC/MM interactions, the densities were fit post-SCF and the properties determined from the fitted density. By contrast, for QC/MM calculations with a fitted density representation, the full density was fit using the appropriate basis set (Small, Medium or Large) and operator (A, C or O), the QC/MM Fock matrix contribution of equation 30 was added into the full QC Fock matrix, and the QC/MM energy in equation 29 was calculated at <u>each</u> SCF iteration. The only QC method that employed a fitting basis set for the QC Coulomb energy was the BLYP

method which used the def2-sv(p)-rifit (Medium) basis set and the Coulomb fit operator, and so the only overlap between the two different density representations occurred when the BLYP QC method and the Medium/C QC/MM density representation were being used. In these instances, the electron density only needed to be fit once per SCF cycle, as opposed to twice in the general case, and the simplification given in equation 32 for the coordinate derivatives was applicable.

Implementation Details

All calculations in this paper were performed with the pDynamo3 QC/MM molecular modeling package. This is an updated version of the pDynamo program[18] that has been ported to the Python3 programming language. The principal developments required for the work reported here concerned the introduction of a module for determining the QC/MM interactions with a fitted QC density, and the evaluation of the extra Gaussian integrals needed for property calculations with a one-center density expansion and for the fitting procedures with anti-Coulomb and overlap operators. The latter consisted of the two-, three- and four-center integrals of equations 11,12 and 5, respectively, together with the derivatives of the two- and three-center integrals that are necessary for geometry optimizations. Although the four-center integrals are not needed in the fitting, they are required to determine the value of the fitting function, $\mathcal{F}$, in equation 3, which is useful for providing some measure of the precision of the fitting procedure.

Practically, the additional overlap integrals are straightforward to determine, whereas the anti-Coulomb operator integrals can be evaluated efficiently using an extension of the Rys polynomial algorithm[19–21] that pDynamo employs for determining Gaussian basis functions over the Coulomb operator. As a basic and brief description of this extension consider the

following two-electron Coulomb primitive integral:

$$\mathcal{I}^{\mathrm{C}}_{nm} = \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 \; (x_1 - x_i)^{n_x}(y_1 - y_i)^{n_y}(z_1 - z_i)^{n_z} \exp\left(-p(\boldsymbol{r}_1 - \boldsymbol{r}_P)^2\right) \frac{1}{r_{12}} \\ (x_2 - x_k)^{m_x}(y_2 - y_k)^{m_y}(z_2 - z_k)^{m_z} \exp\left(-q(\boldsymbol{r}_2 - \boldsymbol{r}_Q)^2\right) \tag{38}$$

In this equation, the integral is between a charge distribution of a product of basis functions centered at positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, and another with positions at $\boldsymbol{r}_k$ and $\boldsymbol{r}_l$. By the Gaussian Product Theorem, the centers and exponents of each product are $\boldsymbol{r}_P$ and p, and $\boldsymbol{r}_Q$ and q, respectively.

The core of the Rys polynomial algorithm is to take the Gaussian transform of the Coulomb operator:

$$\frac{1}{r_{12}} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} du \; \exp\left(-u^2 r_{12}^2\right) \tag{39}$$

and to substitute it into equation 38. This gives:

$$\mathcal{I}^{\mathrm{C}}_{nm} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} du \; G_{n_x m_x}(u) \; G_{n_y m_y}(u) \; G_{n_z m_z}(u) \tag{40}$$

where $G_{n_x m_x}$ takes the form:

$$G_{n_x m_x}(u) = \int dx_1 \int dx_2 \; (x_1 - x_i)^{n_x}(x_2 - x_k)^{m_x} \exp\left(-p(x_1 - x_P)^2 - q(x_2 - x_Q)^2 - u^2(x_1 - x_2)^2\right) \tag{41}$$

and similarly for $G_{n_y m_y}$ and $G_{n_z m_z}$.

Dupuis, King and Rys showed that the integral over u in equation 40 could be evaluated efficiently via numerical quadrature using Rys polynomials[19–21] and derived a series of recurrence relations for determining the G_{nm} functions starting with the function G_{00}.[21] Likewise, they derived so-called transfer relations that linked integrals of the form of equation 38 that only involve polynomial factors of $\boldsymbol{r}_i$ and $\boldsymbol{r}_k$ to those involving factors of $\boldsymbol{r}_j$ and $\boldsymbol{r}_l$ as well.[21]

For the anti-Coulomb integral, $\mathcal{I}^{\mathrm{A}}_{nm}$, the appropriate modification of the above procedure is to note that the Gaussian transform of the anti-Coulomb operator is simply $-r_{12}^2$ times

the values in equation 39. Following the same steps as before gives:

$$\mathcal{I}^{\mathrm{A}}_{nm} = -\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} du \; [H_{n_x m_x}(u)\; G_{n_y m_y}(u)\; G_{n_z m_z}(u) + G_{n_x m_x}(u)\; H_{n_y m_y}(u)\; G_{n_z m_z}(u) + G_{n_x m_x}(u)\; G_{n_y m_y}(u)\; H_{n_z m_z}(u)\;] \tag{42}$$

where $H_{n_x m_x}$ takes the form:

$$H_{n_x m_x}(u) = \int dx_1 \int dx_2 \, (x_1 - x_2)^2 (x_1 - x_i)^{n_x} (x_2 - x_k)^{m_x} \exp\left(-p(x_1 - x_P)^2 - q(x_2 - x_Q)^2 - u^2(x_1 - x_2)^2\right) \tag{43}$$

The H_{nm} can be evaluated by noting that $(x_1 - x_2) = (x_1 - x_i) - (x_2 - x_k) + (x_i - x_k)$. Substituting in equation 43 gives an equation for the H_{nm} in terms of the G_{nm} functions:

$$H_{n,m} = G_{n,m} x_{ik}^2 + 2\left(G_{n+1,m} - G_{n,m+1}\right) x_{ik} + G_{n+2,m} + G_{n,m+2} - 2G_{n+1,m+1} \tag{44}$$

where $x_{ik} = x_i - x_k$.

This immediately suggests the following algorithm for evaluating the anti-Coulomb integrals. For each Rys root (related to the u parameter) required for the numerical quadrature:

- Determine the G_{nm} functions for the integral using the standard recurrence relations but noting that H_{nm} requires n and m values two higher than those for Coulomb integrals.
- Determine the H_{nm} functions using equation 44.
- Apply the transfer relations[21] to both G_{nm} and H_{nm} to generate the full set of integrals of polynomial powers over the centers i, j, k and l. Note that this step can be omitted or simplified when integrals involving only two or three basis functions are being calculated.
- Assemble the integrands in equation 42 using the G_{nm} and H_{nm} values and add them

into the final integrals for the centers i, j, k and l with the appropriate quadrature weight factors.

This algorithm was implemented in pDynamo3 for evaluating the two-, three- and four-function anti-Coulomb integrals and the two- and three-function integral derivatives. The integral implementation was checked by hand against the analytic values for s- and p-functions and by comparing with the previously published derivations of Preiskorn and Żurawski[22].

Results and Discussion

The complete set of results from the monomer and dimer calculations is very extensive, but the main conclusions to be drawn from it are clear and so only a selection of pertinent results will be presented here. Complementary results to the ones presented below may be found in the Supporting Information.

There were 111 monomer property calculations in total consisting of 37 different monomers with three different QC methods. All three QC methods gave very similar results for a given molecule and so only those for the B3LYP method are discussed here. Concerning the monomers themselves, the data can be divided into two broad groups, denoted I and II, the first of which comprises monomers made up solely of atoms from the first and second rows of the periodic table, and the second which has atoms from the third and higher rows of the table as well. Two sets of results which are representative of these groups are given in tables 1 and 2 for the methanol and dibromomethane monomers, respectively. Tables of the B3LYP results for all monomers are given in the Supporting Information.

First consider the percentage errors in the density fits, which are estimated using the expression $100 \times \sqrt{\mathcal{F}/\langle\rho|\hat{\mathcal{O}}_{12}|\rho\rangle}$. For all basis sets these errors are smallest for the anti-Coulomb fit operator and largest for the overlap operator with values, in general, of at most a few percent. The exceptions are for the Coulomb and overlap operators with the Medium and Large basis sets for group II monomers whose values can reach as high as 70%!

Comparison of these basis sets with the Small basis set shows that the latter has s-type functions with large exponents for third and higher row elements, whereas they are absent in the Medium and Large sets. That this is indeed the cause of the difference can be proved by repeating the property calculations with artificial Medium and Large basis sets augmented with the appropriate high-exponent s-functions from the Small basis set and showing that the fit errors decrease equivalently. Overall this behavior of the fit operators can be understood by noting that the anti-Coulomb operator emphasizes longer range interactions compared to the shorter range overlap operator with the Coulomb operator somewhere between the two.

Turning now to the property values, it is clear that the atomic charges calculated by partitioning the fitted densities are not reliable and bear little resemblance to the Mulliken charges obtained from the full density or even to those that might be expected from chemical intuition. By contrast, the remaining properties, namely the ESP charges and the dipole and quadrupole moments, which depend globally on the density rather than its partition, might be expected to be better reproduced by the fitting process. Inspection of the tables shows that this is indeed the case. The trends are similar to those reported above for the fit errors as the anti-Coulomb fit operator generally gives the most precise results, followed closely by the Coulomb operator and then the overlap operator which often gives poor results, especially when using the Medium and Large basis sets for group II monomers. If the overlap operator is excluded, then these results appear quite promising for the use of fitted densities in determining the QC/MM electrostatic interactions, particularly as there is good agreement for the ESP charges which is a popular approach for deriving charges in MM force fields.

The first set of dimer tests consisted of forty three interaction curves which were calculated for each of the 101 dimers. Four representative sets of curves are given in figures 1 to 4 for the water-water, water-methanol, fluoromethane-water and acetate anion-water dimers, respectively. Equivalent plots for all the dimers in the database are given in the Supporting Information. In these plots, and in light of the monomer results discussed above, the curves

determined using the overlap fit operator are omitted as they can deviate considerably from the others, except in a few cases, such as for those dimers dominated by non-polar interactions. Likewise, only curves obtained with the HF QC method are shown as the BLYP method behaves very similarly, although its interaction energies are typically more attractive than the HF ones. A few other general remarks about the curves can be made, namely: (i) the reference curves from the database are often much less repulsive at short range than the calculated ones; (ii) the MM Lennard-Jones terms typically dominate the MM, MM/QC and QC/MM interaction energies, especially at short-range due to the overly repulsive r^{12} term; (iii) the dimer interaction energies are more sensitive to how the dimers are partitioned, either MM/QC or QC/MM, than to which interaction method is used; (iv) there is relatively little difference between the energies with the different basis sets, although the Medium basis set often deviates the most from the Small and Large basis set results; and (v) the curves calculated with the anti-Coulomb operator agree more closely with those determined using the Full interaction method than do those obtained with the Coulomb operator.

The second set of dimer tests were the unconstrained geometry optimizations of the 33 dimers from the database which contained at least one water monomer with 16 optimizations being carried out for each dimer. All but twelve of the 528 optimizations converged. Those that failed were the seven optimizations of the hydrogen phosphate-water dimer (s101x7 database number 73) with the MM/QC partitioning and five of the seven optimizations of the dihydrogen phosphate-water dimer (s101x7 database number 74) with the same MM/QC partitioning (the two exceptions were those with the Small basis set). The reason for these failures is due to the absence of Lennard-Jones interactions on water hydrogens in the standard OPLS force field, which means that in anionic systems it is possible for these hydrogens, when treated in the QC region, to collapse onto negatively charged MM atoms, as is observed here. This problem can be prevented by including suitable non-zero values for the relevant Lennard-Jones parameters, a solution which is adopted, for example, in CHARMM's modified OPLS mTIP3P water model.[23]

Histogram plots of the RMS coordinate differences between sets of optimized structures with two different energy models are displayed in figure 5, and a statistical analyses of these values in table 3. From the plots it can be seen that the: (i) the Full model structures in general resemble more closely those optimized with the MM model than the HF model; (ii) the use of the Medium and Large basis sets is to be preferred over that of the Small basis set; and (iii) fitting with the anti-Coulomb operator gives fewer outlying structures compared to the Full model than does fitting with the Coulomb operator.

The outlying structures in the histogram plots, those with the largest RMS coordinate differences, are from only a small subset of dimers, namely those that contain peptide, DMSO, phosphoric acid, benzylchloride and phenol together with water (s101x7 database numbers 4, 71, 75, 87 and 99, respectively). In the peptide, DMSO, benzylchloride and phenol dimers the hydrogen bonds are between the water hydrogen and the carboxyl oxygen, the sulfoxide oxygen, the chloride and the phenol oxygen, respectively. These bonding patterns are maintained in the optimized structures with the variations being due, depending on the model, to shifts of the water position (for example, by a 180^{o} rotation about one of the water bonds), and they can occur when the water is in either the QC or MM regions. By contrast, the differences for the phosphoric acid dimer are due to a distored phosphoric acid structure when it is treated with the MM model, although this distortion is not observed after MM optimization of the isolated phosphoric acid monomer.

As can be seen from table 3, removal of these dimers from the statistical analysis greatly improves the agreement between the Full and fitted density energy models. The remaining high value in the table, 0.682 for the Small/C model, is due to the QC/MM structure of the methylsulfide-water dimer (s101x7 database number 67) which adopts a sulfur hydrogen-water oxygen hydrogen bond instead of the water hydrogen-sulfur bond found in the other optimized structures. If this structure is also removed, the maximum RMS coordinate difference between the Small/C and Full model structures becomes 0.355, similar to that of the other models.

As a final note, it was observed that the average times required for a combined energy and gradient evaluation during the dimer geometry optimizations with the various density-fitting methods were often very close to and, in the case of the Medium/A approximation sometimes faster, than those for the Full model with an un-fitted density. This provides support for the potential promise of these approaches for the treatment of QC/MM interactions in much larger simulations given the relatively small systems tested here, and the expectation that the advantage of density fitting methods should grow as the number of MM atoms increases.

Conclusions

This paper has investigated the use of density fitting approaches for the calculation of electrostatic interactions between QC and MM atoms in hybrid potential calculations. It was shown that these approximations can reproduce to good precision the interactions determined using the full, un-fitted, densities even with fit basis sets that have been optimized for the calculation of electron repulsion energies. Interestingly, employing an anti-Coulomb operator to perform the density fit gives accuracies that are as good or better than those obtained with the more usual Coulomb operator. Overall these results are promising and suggest that density fitting methods could prove useful for helping to speed up QC/MM simulations in large systems if sufficiently precise and sufficiently small fit basis sets specific to these applications can be parametrized. Such developments are left for future work.

Supporting Information Available

A pdf file with: (i) tables of calculated property values for the 37 monomers from the s101x7 database using the B3LYP QC method and ten different representations of the QC electron density; and (ii) figures of energy interaction curves calculated for the 101 dimers from the s101x7 database using the HF QC method, MM/QC and QC/MM partitionings and seven different representations of the QC electron density.

Acknowledgement

The author acknowledges the support of the CoMX group in the Chemistry and Biology of Metals Laboratory at the CEA, and of the Theory Group at the Institut Laue-Langevin during the preparation of this work.

References

(1) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations I. The Computational Procedure. *Chem. Phys.* **1973**, *2*, 41–51.

(2) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.

(3) Jafri, J. A.; Whitten, J. L. Electron Repulsion Integral Approximations and Error Bounds: Molecular Applications. *J. Chem. Phys.* **1974**, *61*, 2116–2121.

(4) Sambe, H.; Felton, R. H. Calculation of the Ionization Potentials of Ozone and Ammonia by a LCAO-Xα Method. *J. Chem. Phys.* **1974**, *61*, 3862–3863.

(5) Sambe, H.; Felton, R. H. A New Computational Approach to Slater's SCF-Xα Equation. *J. Chem. Phys.* **1975**, *62*, 1122–1126.

(6) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On Some Approximations in Applications of Xα Theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.

(7) Köster, A. M.; del Campo, J. M.; Janetzko, F.; Zuniga-Gutierrez, B. A MinMax Self-Consistent-Field Approach for Auxiliary Density Functional Theory. *J. Chem. Phys.* **2009**, *130*, 114106–1–114106–8.

(8) Liu, M.; Wang, Y.; Chen, Y.; Field, M. J.; Gao, J. QM/MM through the 1990s: the First Twenty Years of Method Development and Applications. *Israel. J. Chem.* **2014**, *54*, 1250–1263.

(9) Ferré, N.; Ángyán, J. G. Approximate Electrostatic Interaction Operator for QM/MM Calculations. *Chem. Phys. Lett.* **2002**, *356*, 331–339.

(10) Schwinn, K.; Ferré, N.; Huix-Rotllant, M. Analytic QM/MM Atomic Charge Derivatives Avoiding the Scaling of Coupled Perturbed Equations with the MM Subsystem Size. *J. Chem. Phys.* **2019**, *151*, 041102–1–041102–4.

(11) Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Taylor, S. W. Modeling the Potential of a Charge Distribution. *J. Chem. Phys.* **1992**, *96*, 7178–7179.

(12) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Auxiliary Basis Expansions for Large-Scale Electronic Structure Calculations. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6692–6697.

(13) Reine, S.; Tellgren, E.; Krapp, A.; Kjærgaard, T.; Helgaker, T.; Jansik, B.; Høst, S.; Salek, P. Variational and Robust Density Fitting of Four-Center Two-Electron Integrals in Local Metrics. *J. Chem. Phys.* **2008**, *129*, 104101–1–104101–12.

(14) Renison, C. A.; Fernandes, K. D.; Naidoo, K. J. Quantum Supercharger Library: Hyperparallel Integral Derivatives Algorithms for ab initio QM/MM Dynamics. *J. Comput. Chem.* **2015**, *36*, 1410–1419.

(15) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theor. Comp.* **2015**, *11*, 2609–2618.

(16) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Mod.* **2019**, *59*, 4814–4820.

(17) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: an Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucl. Acids Res.* **2017**, *45*, W331–W336.

(18) Field, M. J. The pDynamo Program for Molecular Simulations using Hybrid Quantum Chemical and Molecular Mechanical Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1151–1161.

(19) Dupuis, M.; Rys, J.; King, H. F. Evaluation of Molecular Integrals over Gaussian Basis Functions. *J. Chem. Phys.* **1976**, *65*, 111–116.

(20) King, H. F.; Dupuis, M. Numerical Integration Using Rys Polynomials. *J. Comput. Phys.* **1976**, *21*, 144–165.

(21) Rys, J.; Dupuis, M.; King, H. F. Computation of Electron Repulsion Integrals Using the Rys Quadrature Method. *J. Comput. Chem.* **1983**, *4*, 154–157.

(22) Preiskorn, A.; Żurawski, B. Many-Electron, Multicenter Integrals in Superposition of Correlated Configurations Method. I. One- and Two-Electron Integrals. *Int J. Quantum Chem.* **1985**, *27*, 641–651.

(23) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

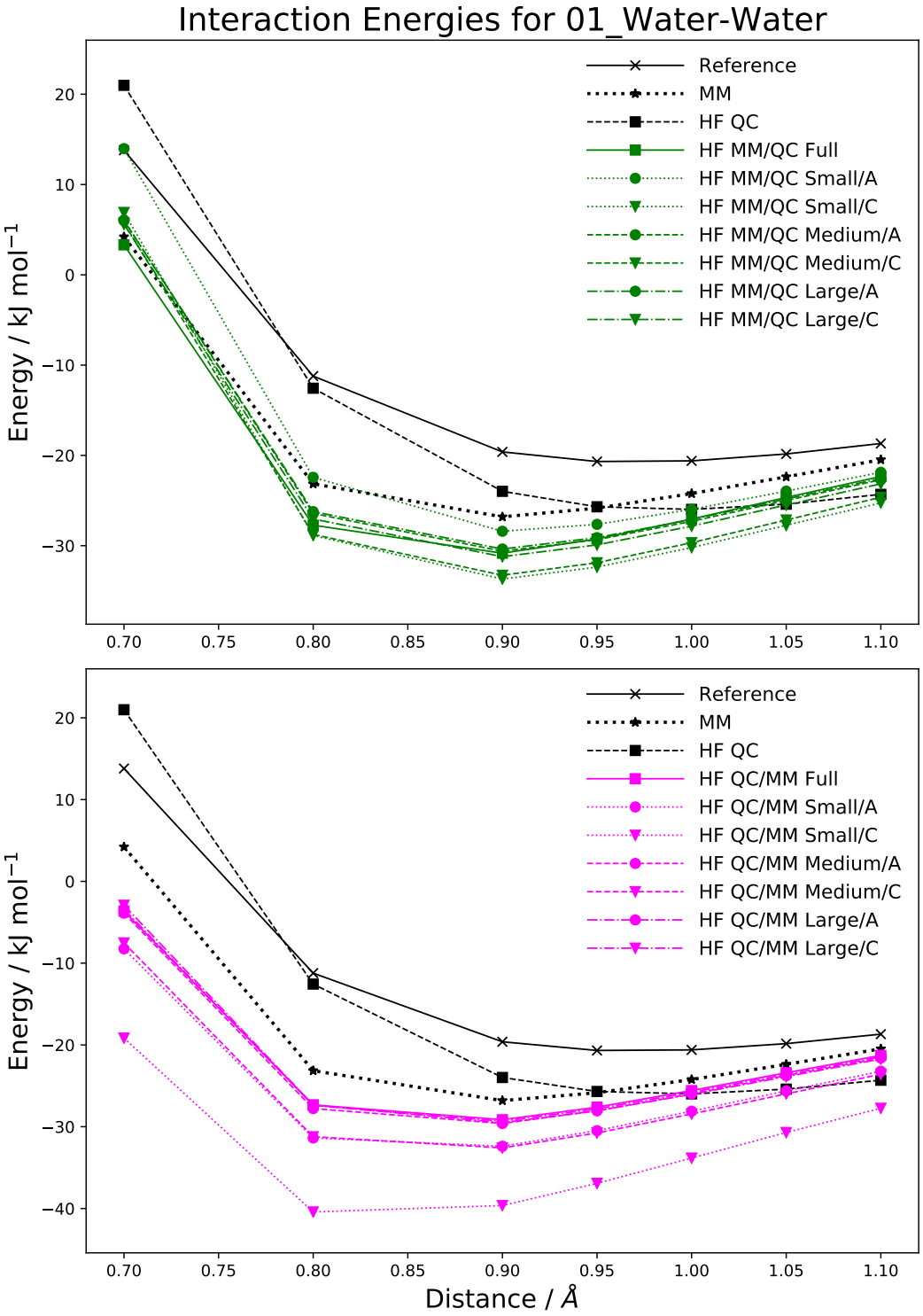

Figure 1: Selected interaction curves for the water-water dimer (s101x7 database number 1).

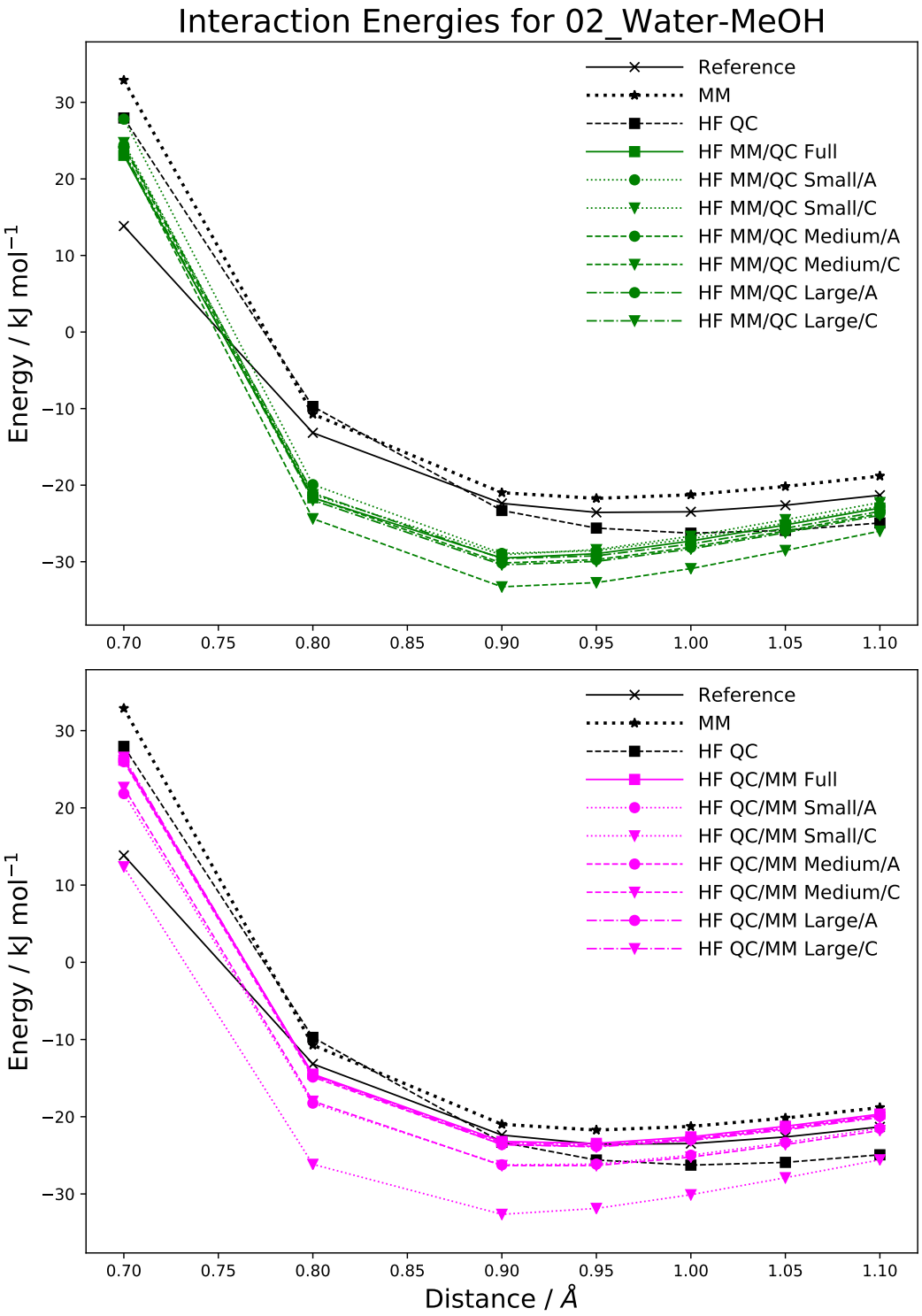

Figure 2: Selected interaction curves for the water-methanol dimer (s101x7 database number 2).

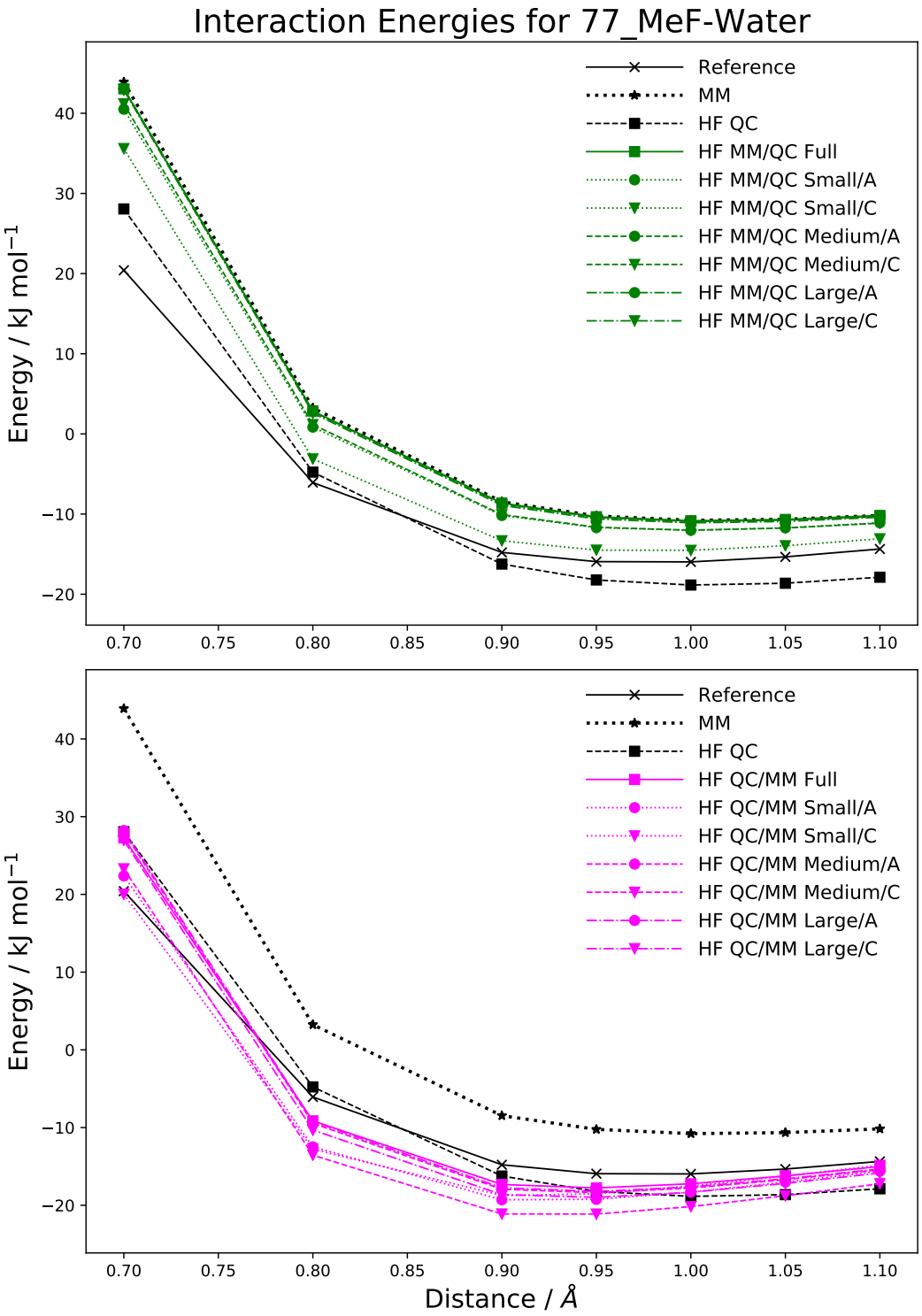

Figure 3: Selected interaction curves for the fluromethane-water dimer (s101x7 database number 77).

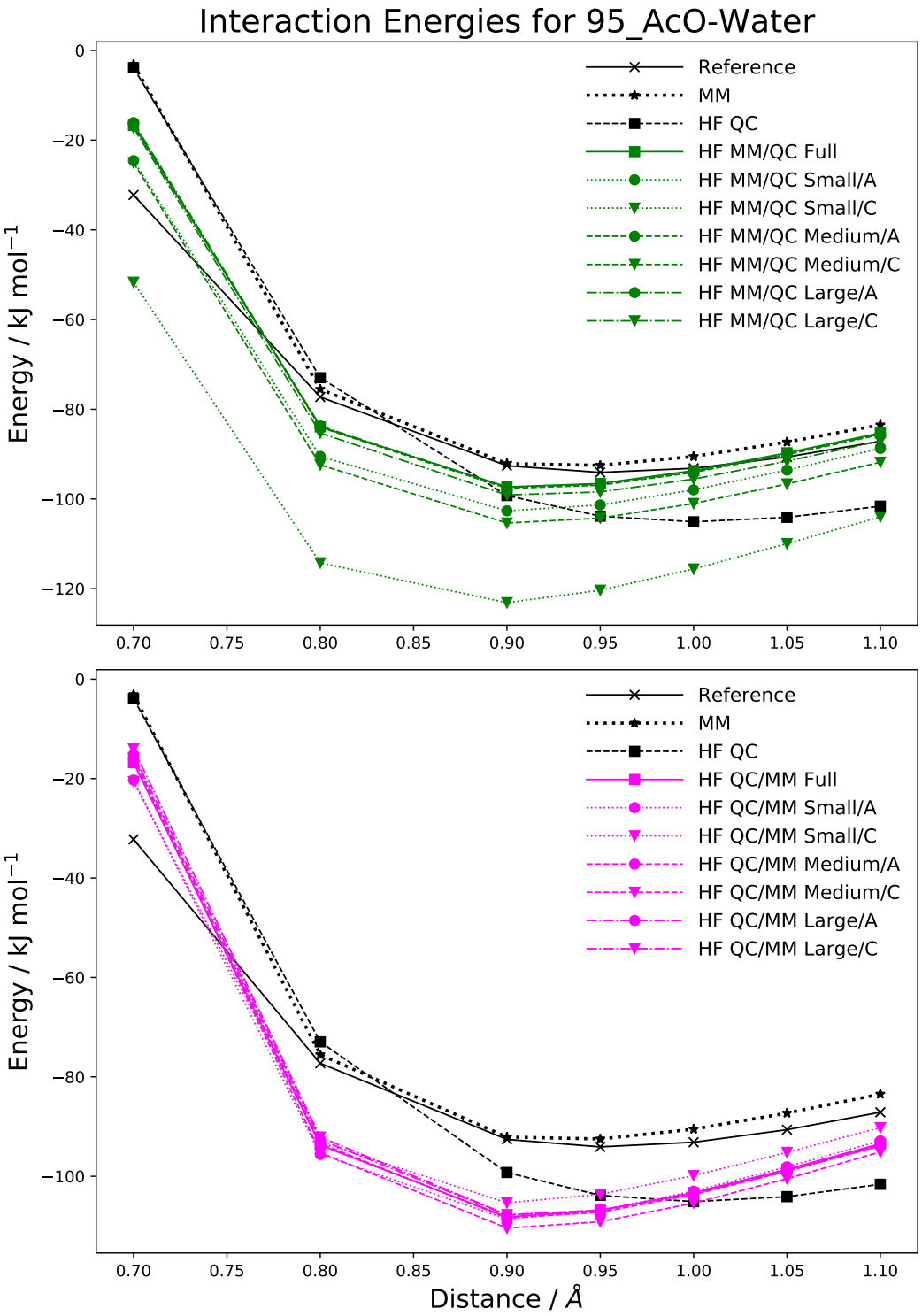

Figure 4: Selected interaction curves for the acetate anion-water dimer (s101x7 database number 95).

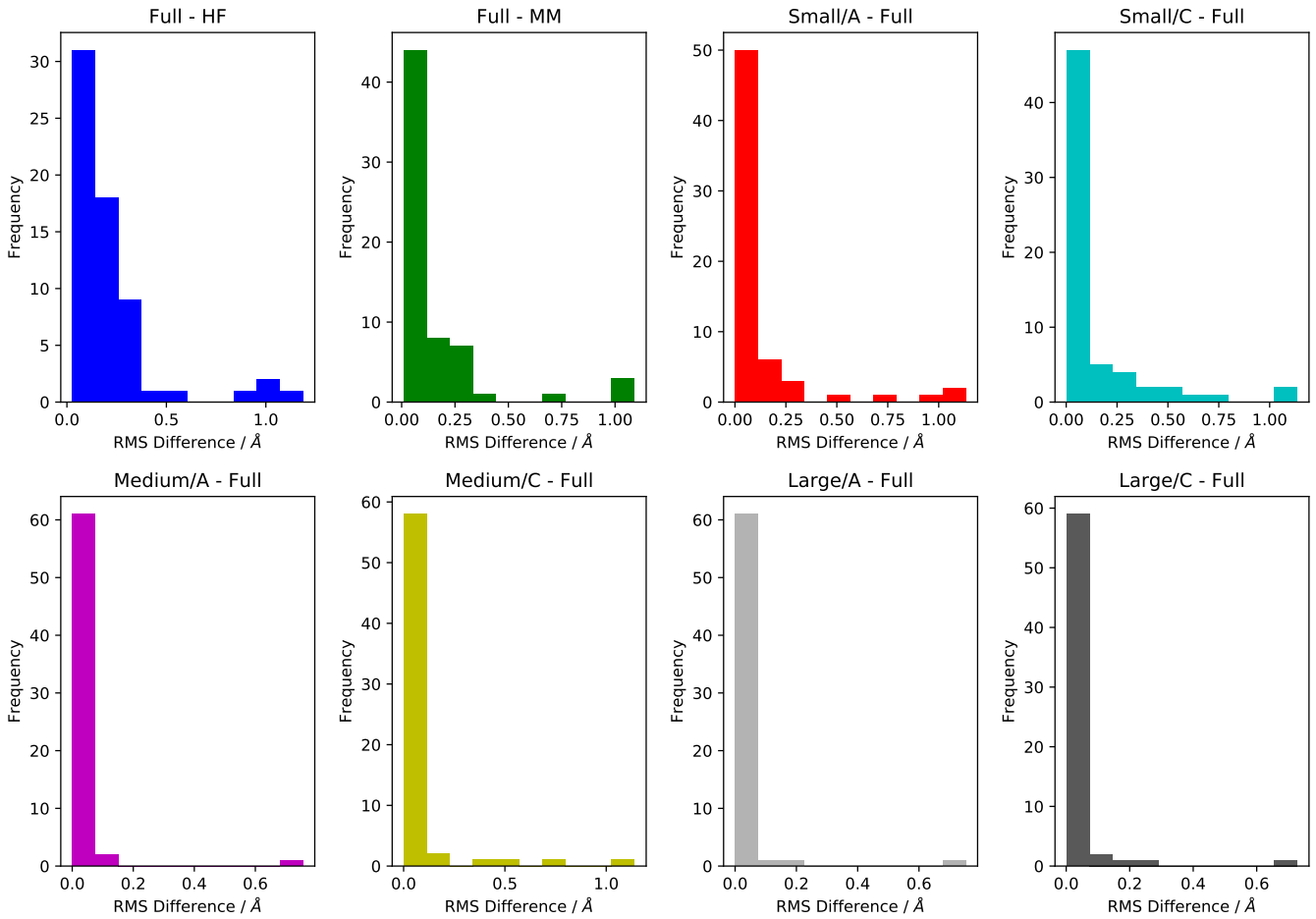

Figure 5: Histogram plots of the RMS coordinate differences between optimized dimer structures determined with different energy models. Each plot contains the results for 64 dimer optimizations made up of 33 with QC/MM partitioning but only 31 with MM/QC partitioning due to the failure of the Full hydrogen phosphate-water and dihydrogen phosphate-water cases (s101x7 database numbers 73 and 74, respectively). Note that all differences were determined after superposition of the relevant structures and that, where pertinent, structures with the same partitioning were employed (i.e. MM/QC with MM/QC and QC/MM with QC/MM).

Table 1: B3LYP calculated electronic properties for Methanol.

Property		*Analysis Method*									
		Small/A	Small/C	Small/O	Medium/A	Medium/C	Medium/O	Large/A	Large/C	Large/O	Full
Charge / e	O0	-1.617	-2.154	-2.431	0.461	0.279	-3.057	-0.026	0.096	-0.172	-0.474
	H1	0.453	0.471	0.530	-0.355	-0.380	0.383	-0.208	-0.231	-0.170	0.296
	C2	0.400	0.373	0.051	0.819	1.152	2.547	0.509	0.341	0.898	-0.064
	H3	0.250	0.459	0.689	-0.265	-0.326	-0.066	0.017	0.049	-0.041	0.097
	H4	0.257	0.426	0.580	-0.330	-0.363	0.097	-0.146	-0.128	-0.257	0.073
	H5	0.256	0.426	0.580	-0.329	-0.363	0.097	-0.146	-0.128	-0.257	0.073
ESP Charge / e	O0	-0.609	-0.615	-0.593	-0.638	-0.702	-1.569	-0.634	-0.646	-0.633	-0.628
	H1	0.426	0.463	0.539	0.411	0.440	0.894	0.410	0.408	0.374	0.409
	C2	0.011	-0.089	-0.265	0.222	0.286	0.802	0.213	0.255	0.299	0.189
	H3	0.102	0.132	0.166	0.042	0.040	0.040	0.044	0.035	0.030	0.049
	H4	0.034	0.055	0.076	-0.019	-0.032	-0.084	-0.016	-0.026	-0.035	-0.010
	H5	0.034	0.055	0.076	-0.019	-0.032	-0.084	-0.016	-0.026	-0.035	-0.010
Dipole / D	x	1.698	1.737	1.855	1.698	1.832	4.304	1.697	1.711	1.668	1.693
	y	0.675	0.537	-0.099	0.693	0.818	2.387	0.693	0.746	0.916	0.688
	z	1.558	1.652	1.852	1.550	1.640	3.581	1.549	1.540	1.394	1.547
Quadrupole / B	xx	-12.733	-12.765	-12.535	-12.609	-12.576	-12.725	-12.613	-12.530	-12.372	-12.634
	yy	-12.148	-12.148	-12.263	-12.111	-11.914	-9.450	-12.111	-12.072	-12.020	-12.125
	zz	-13.624	-13.743	-14.056	-13.523	-13.465	-12.423	-13.520	-13.482	-13.335	-13.517
	xy	-2.288	-2.482	-2.864	-2.237	-2.440	-5.118	-2.235	-2.239	-2.120	-2.220
	xz	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	yz	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fit Error / %		0.216	0.984	2.005	0.057	0.888	5.050	0.024	0.137	1.989	

Table 2: B3LYP calculated electronic properties for Dibromomethane.

Property		*Analysis Method*									
		Small/A	Small/C	Small/O	Medium/A	Medium/C	Medium/O	Large/A	Large/C	Large/O	Full
Charge / e	Br0	-1.396	-3.758	-14.374	0.038	0.703	-4.020	0.681	5.473	46.934	-0.067
	Br1	-1.396	-3.758	-14.374	0.038	0.703	-4.020	0.681	5.473	46.934	-0.067
	C2	2.070	6.042	22.964	0.398	0.768	27.912	-1.420	-8.423	-41.209	-0.214
	H3	0.361	0.737	2.892	-0.237	-1.087	-9.936	0.030	-1.262	-26.330	0.174
	H4	0.361	0.737	2.892	-0.237	-1.087	-9.936	0.030	-1.262	-26.330	0.174
ESP Charge / e	Br0	0.105	-0.102	-1.798	0.029	-0.028	0.178	0.028	-0.118	-3.483	0.031
	Br1	0.105	-0.102	-1.798	0.029	-0.028	0.178	0.028	-0.118	-3.483	0.031
	C2	-1.174	-0.500	6.254	-0.702	-0.296	5.870	-0.697	0.138	21.959	-0.710
	H3	0.482	0.352	-1.329	0.323	0.176	-3.113	0.320	0.049	-7.497	0.324
	H4	0.482	0.352	-1.329	0.323	0.176	-3.113	0.320	0.049	-7.497	0.324
Dipole / D	x	1.523	2.926	11.190	1.423	1.184	19.972	1.428	1.580	1.819	1.427
	y	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	z	1.523	2.926	11.190	1.423	1.184	-19.972	1.428	1.580	-1.819	1.427
Quadrupole / B	xx	-45.336	-53.857	-97.799	-43.674	-40.245	160.342	-43.771	-43.015	74.945	-43.761
	yy	-40.944	-41.442	-41.287	-40.312	-37.558	99.558	-40.340	-38.531	65.270	-40.395
	zz	-43.140	-43.180	-35.761	-42.671	-38.865	137.422	-42.638	-38.689	133.253	-42.701
Fit Error / %		0.199	1.604	2.203	0.123	8.421	68.698	0.123	8.420	68.697	

Table 3: Statistics of the RMS coordinate differences between structures resulting from the geometry optimizations with two different energy models. The first statistics in each column are those when all converged optimized structures are included in the calculation (64 values), whereas the second are those when the dimers containing peptide, DMSO, phosphoric acid, benzylchloride and phenol together with water (s101x7 database numbers 4, 71, 75, 87 and 99, respectively) are excluded (54 values). Units are in Å.

Model 1	Model 2	Mean ± Standard Deviation	Maximum
Full	HF	0.224 ± 0.227 ; 0.163 ± 0.086	1.188 ; 0.363
Full	MM	0.150 ± 0.232 ; 0.093 ± 0.090	1.092 ; 0.372
Small/A	Full	0.115 ± 0.240 ; 0.053 ± 0.076	1.134 ; 0.339
Small/C	Full	0.136 ± 0.240 ; 0.080 ± 0.128	1.137 ; 0.682
Medium/A	Full	0.025 ± 0.094 ; 0.013 ± 0.018	0.757 ; 0.087
Medium/C	Full	0.070 ± 0.177 ; 0.034 ± 0.057	1.140 ; 0.364
Large/A	Full	0.025 ± 0.095 ; 0.010 ± 0.012	0.754 ; 0.082
Large/C	Full	0.031 ± 0.096 ; 0.017 ± 0.034	0.728 ; 0.235

Graphical TOC Entry