Density-dependent formulation of dispersion-repulsion interactions in hybrid QM/MM models

Carles Curutchet,[1]* Lorenzo Cupellini,[2] Jacob Kongsted,[3] Stefano Corni,[4,5] Luca Frediani,[6] Arnfinn Hykkerud Steindal,[6] Ciro A. Guido,[7] Giovanni Scalmani[8] and Benedetta Mennucci[2]

[1]*Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain*

[2]*Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via G. Moruzzi 13, I-56124 Pisa, Italy*

[3]*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark*

[4]*Dipartimento di Scienze Chimiche, Università di Padova, v. F. Marzolo 1, 35131 Padova, Italy*

[5]*CNR-NANO Istituto Nanoscienze, Modena, Italy*

[6]*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, N-9037, Tromsø, Norway*

[7]*Laboratoire CEISAM - UMR CNRS 6230, Université de Nantes, 2 Rue de la Houssinière, 44322 Nantes Cedex 3, France*

[8]*Gaussian, Inc., 340 Quinnipiac Street, Building 40, Wallingford, Connecticut 06492, United States*

*Corresponding author: CC (carles.curutchet@ub.edu)

Abstract

Mixed QM/MM models are widely used to explore the structure, reactivity and electronic properties of complex chemical systems. Whereas such models typically include electrostatics, and potentially polarization in so-called electrostatic and polarizable embedding approaches, respectively, non-electrostatic dispersion and repulsion interactions are instead commonly described through classical potentials

despite their quantum mechanical origin. Here we present an extension of the Tkatchenko-Scheffler semiempirical van der Waals (vdW^{TS}) scheme aimed at describing dispersion and repulsion interactions between quantum and classical regions within a QM/MM polarizable embedding framework. Starting from the vdW^{TS} expression, we define a dispersion and a repulsion term, both of them density-dependent and consistently based on a Lennard-Jones-like potential. We explore transferable atom type-based parametrization strategies for the MM parameters, based on either vdW^{TS} calculations performed on isolated fragments or on a direct estimation of the parameters from atomic polarizabilities taken from a polarizable force field. We investigate the performance of the implementation by computing self-consistent interaction energies for the S22 benchmark set, designed to represent typical non-covalent interactions in biological systems, in both equilibrium and out-of-equilibrium geometries. Overall, our results suggest that the present implementation is a promising strategy to include dispersion and repulsion in multiscale QM/MM models incorporating their explicit dependence on the electronic density.

1. Introduction

Multiscale quantum/molecular mechanics (QM/MM) methods have become a uniquely powerful tool in quantum chemistry for determining energies and properties of (supra)molecular systems in either solutions, biomatrices, or in composite systems.[1,2] Currently, the largest part of QM/MM implementations only account for the electrostatic part of the QM/classical interactions (this scheme is generally indicated as electrostatic embedding), but more recently, polarizable embedding, where mutual induction effects between the QM and the MM subsystems are also included, have appeared in the literature.[3–10]

Non-electrostatic interactions, namely dispersion and Pauli repulsion, are instead commonly described through standard atom-based parametrization of the London forces,[11] through a Lennard-Jones type of potential,[12] in analogy to what is done for fully classical simulations.[13] Although this parametrization yields a correction to the total energy of the system, it does not affect the QM Hamiltonian. Some examples have been presented so far which go beyond this approximation. A parametrization at the Hartree-Fock level which includes dispersion was presented in 1996 by Van Duijnen and De Vries.[14] An alternative which has gained momentum in recent years is provided by the Effective Fragment Potential (EFP) method,[15] which computes interaction energies based on extensive quantum mechanical parametrization of the MM fragments. The first model (EFP1)[16] was available for water molecules and thus of limited general applicability but the second model (EFP2) is generic and can in principle be employed for any kind of fragments. The EFP method shows very good performance,[17] but the parametrization of the non-electrostatic contribution is however not affecting the electronic degrees of freedom directly.[18]

One of the strengths of electrostatic and polarizable embeddings is instead their formulation within a general quantum mechanical response theory framework meaning that this method is capable not only of predicting energies but also general response properties. In order for dispersion and Pauli repulsion to enter into the computational method on the same footing as electrostatics and polarization, a fully self-consistent formulation is required where the appropriate QM operators are formulated and included into the expression for the effective Hamiltonian. An example in this direction, the charge-dependent exchange and dispersion (QXD) model reported by Kuechler, Giese and York,[19,20] addressed this issue through a scaled overlap model for exchange and dispersion interactions that is a function of atomic charge. Very recently, another

approach has been presented[21] by combining a polarizable QM/fluctuating charge (FQ) approach[8] with the formalism developed by Tkatchenko and Scheffler (vdW^{TS})[22] for the dispersion and introducing an auxiliary density on the MM portion for calculating the repulsion by a density-overlap formula. The approach is applicable to water as solvent.

In this work, we present an extension and a generalization of a polarizable QM/MM approach to specifically include also dispersion and repulsion QM operators. The specific QM/MM model used, known as QM/MMPol,[6] is based on the combination of fixed point charges and induced dipoles for describing the MM atoms, For the dispersion contribution, the Tkatchenko and Scheffler formalism (vdW^{TS})[22] is used while the repulsion term is automatically obtained by exploiting a Lennard-Jones-like relation with the dispersion. In such a way, both nonelectrostatic contributions will be obtained in a single step through the assignment of the dispersion coefficients for the MM sites. The method can treat any atom-type and it thus can be applied not only to a solvent but also to more complex environments such as proteins and other biomatrices. We note that recently, the vdW^{TS} approach has also been extended to plane waves.[23]

We especially focus on the generality and portability of the calculated parameters, in order to remove the need of an *ad hoc* procedure for each MM fragment, in line with the transferability assumption that characterizes classical force fields. Our results indicate that vdW^{TS}-derived parameters are indeed largely transferable among elements with different hybridization states or located in different molecular environments, thus allowing an accurate description of dispersion and repulsion interactions through the QM/MM boundary.

We test our method for dispersion and repulsion terms on interaction energies for non-covalent complexes contained in the S22 benchmark set, designed to represent typical interactions in biological systems. By adopting the parameters defined in the original Tkatchenko-Scheffler formulation, we find that our approach is able to accurately describe the interaction energies of complexes dominated by dispersion. Not surprisingly, we find it necessary to adjust the free radius of polar hydrogens bound to heteroatoms in hydrogen-bonded and mixed complexes. After this minimal calibration, the overall polarizable potential with dispersion-repulsion terms is able to accurately describe interaction energies for the whole S22 benchmark set, with mean absolute errors ~1-2 kcal/mol compared to full QM reference calculations. In addition, we show that this performance is not degraded in out-of-equilibrium structures. Overall, we show

that the present implementation is a promising strategy to incorporate dispersion and repulsion effects in QM/MM models including their explicit dependence on the QM wave function.

2. Theory

The presentation of the new implementation of dispersion-repulsion terms within a QM/MM framework is organized as follows. First we introduce, separately, the expressions used for dispersion and repulsion energies. Then we present their common extension to a self-consistent approach.

2.1 **Dispersion Energy term**

The starting point of our dispersion model is the van der-Waals (vdW) density functional scheme originally proposed by Tkatchenko and Scheffler[22] (vdW^{TS}). The original derivation of this vdW^{TS} functional builds on the usual London expression for the pairwise short-term attraction between two atoms, which goes with the inverse sixth power of the distance:[24]

$$E_{vdW} = -\frac{1}{2}\sum_{AB} f_{AB}[n(r)]C_{6,AB}[n(r)]R_{AB}^{-6} \quad (1)$$

where the sum goes over all pairs of atoms AB in the system. R_{AB} is the interatomic distance between atoms A and B, $C_{6,AB}$ is the *effective* dipole vdW coefficient and f_{AB} is a damping function to correct for the divergence at short distances. Both the dipole vdW coefficient and the damping function are density-dependent. The expression for the effective vdW dipole coefficient is

$$C_{6,AB} = \gamma_A[n(r)]\gamma_B[n(r)]C_{6,AB}^{free} \quad (2)$$

where the atomic volume ratio γ_A is defined as

$$\gamma_A[n(r)] = \frac{V_A[n(r)]}{V_A^{free}} = \frac{\int r^3\omega_A(r)n(r)dr^3}{\int r^3 n_A^{free}(r)dr^3} \quad (3)$$

In this expression ω_A is the Hirshfeld weight.[25] In the practical implementation of the Hirshfeld partitioning, the integrals in Eq. 3 are evaluated numerically using the standard approach used in DFT calculations,[26] i.e.

$$\int r^3\omega(r)n(r)dr^3 \simeq \sum_i w_i\Omega_{i,A}n(r_i) \quad (4)$$

where w_i are the quadrature weights, and we adopted an extended definition of the Hirshfeld weights,

$$\Omega_{i,A} = \varphi_i \frac{n_{i,A}^{free}}{\sum_B n_{i,B}^{free}} + (1-\varphi_i)\delta_{(i),A} = \varphi_i \frac{n_{i,A}^{free}}{n_{i,T}^{free}} + (1-\varphi_i)\delta_{(i),A} \quad (5)$$

In the above expression, the quantity $\delta_{(i),A}$ is equal to one only if the integration grid point i belongs to the integration grid centered on atom A, while φ_i is a continuous switching function (with continuous derivatives) whose argument is $\ln n_{i,T}/\ln 10$. In this work we used the switching function introduced in Ref. [27]. This extended definition of the Hirshfeld weight is more robust and is important for the self-consistent implementation of the method (see below). In this work, we use atomic free densities $n_{i,A}^{free}$ that were generated as described in Ref. [28].

The free C_6 coefficient is defined as

$$C_{6,AB}^{free} = \frac{2C_{6,AA}C_{6,BB}}{\frac{\alpha_B^0}{\alpha_A^0}C_{6,AA} + \frac{\alpha_A^0}{\alpha_B^0}C_{6,BB}} \quad (6)$$

and the damping factor is defined as

$$f_{AB} = \left(1 + e^{\left\{-d\left(\frac{R_{AB}}{s_r R_{AB}^0} - 1\right)\right\}}\right)^{-1} \quad (7)$$

The density dependence of the damping function enters through R_{AB}^0, which accordingly is defined as

$$R_{AB}^0 = \gamma_A [n(r)]^{1/3} R_{A,free}^0 + \gamma_B [n(r)]^{1/3} R_{B,free}^0 \quad (8)$$

In order to adapt the QM formalism described above to the description of dispersion interactions between a solute treated quantum mechanically and a classical environment described by an embedding potential, we begin by partitioning the energy expression in Eq. 1, dividing the atoms between molecules treated at the QM level, m, and the classically treated environment s

$$E_{vdW} = E_{vdW}^{mm} + E_{vdW}^{ms} + E_{vdW}^{ss} \quad (9)$$

In a similar manner, the electronic density is likewise partitioned into a molecular part $n_m(r)$ and an environment part $n_s(r)$. The latter is not present in the calculation because the environment molecules are treated classically. However, in order to derive the dispersion parameters for the molecules making up the environment, we will assume that the parameters for the environment atoms can be obtained by a corresponding calculation on each molecule in the environment separately.

The first term in Eq. 9 is identical to Eq. 1, but restricted to the atoms of the molecule treated using QM, the second term is the molecule-environment dispersion energy and the last contribution is the dispersion energy of the environment, which is calculated by

classical means, e.g. through a parametrized function, depending on the force field of the MM part. We will consequently not consider this contribution further. In order to evaluate E_{vdW}^{ms} we start from Eq. 1, where now atoms A belong to the molecule and atoms B to the environment:

$$E_{vdW}^{ms} = -\sum_{AB} f_{AB}[n_m(r)] C_{6,AB}[n_m(r)] R_{AB}^{-6} \quad (10)$$

The expression is slightly modified: the factor $1/2$ is removed because the indices run on two distinct sets and the density is the molecular density n_m, which is now computed in the presence of the environment, i.e. incorporating now effects of environment electrostatics and polarization as described previously. The expression for the effective vdW coefficient in case of a molecule-environment pair of atoms becomes:

$$C_{6,AB} = \gamma_A[n_m(r)] \gamma_B C_{6,AB}^{free} \quad (11)$$

where γ_A is computed in the same way as for an isolated molecule, whereas γ_B is pre-computed for each atom in a generic molecule making up the environment. γ_B is therefore a constant, not bearing any density dependence. Furthermore, the expression for the damping function is unchanged, whereas R_{AB}^0 is now computed as

$$R_{AB}^0 = \gamma_A[n_m(r)]^{1/3} R_{A,free}^0 + {\gamma_B}^{1/3} R_{B,free}^0 \quad (12)$$

reflecting again the lack of explicit density dependence of γ_B. Thus, based on tabulated values for the free-atom parameters as well as pre-calculated atomic γ values for the atoms making up the environment enables a prediction of the solute-environment dispersion energy taking explicitly into account only the solute density dependence.

In the following section we will pursue such atomic environment γ calculations and investigate the possibility to divide these into specific atom-types following the conventional strategy adopted in the construction of molecular mechanical force fields.

2.2 Repulsion Energy term

The Pauli repulsion energy is formally defined as the sum of penetration and exchange contribution

$$\mathrm{E_{Rep}^{ms}} = \frac{1}{2} \int \mathrm{dr_1 dr_2} \frac{\eta_A(\mathrm{r_1},\mathrm{r_2})\eta_B(\mathrm{r_1},\mathrm{r_2})}{\mathrm{r_{12}}} \quad (13)$$

This expression can be extended to a QM/MM partition working out a suitable approximate form for the density matrix of the classical region, for instance, using an extension of the Amovilli-Mennucci approach[29] in the framework of the polarizable continuum model.[21,30] However, a computationally inexpensive alternative way to

include these effects is to assume an effective Lennard-Jones (LJ) behavior of the repulsion-dispersion terms:

$$E_{LJ,rep}^{ms} = \sum_{AB} C_{12,AB} R_{AB}^{-12} \tag{14}$$

If the equilibrium position of the A–B potential is defined as R_{AB}^{0}(eq. 6), one obtains

$$C_{12,AB} = \frac{1}{2} C_{6,AB} (R_{AB}^{0})^{6} \tag{15}$$

and therefore:

$$E_{LJ,rep}^{ms} = \sum_{AB} \frac{1}{2} C_{6,AB} (R_{AB}^{0})^{6} R_{AB}^{-12} \tag{16}$$

By using this expression, a single step will be necessary to simultaneously obtain dispersion and repulsion parameters. This is indeed a quite cost-effective strategy that allows an easy extension of the model to different kind of classical environments.

2.3 Self Consistent Field implementation

The self consistent field implementation of the approach presented here is obtained by expanding the QM electron density $\mathrm{n_m(r)}$ in a finite basis set and taking the derivative with respect to the density matrix $\mathrm{P_{\mu\nu}}$, in order to define the contribution to the Fock (or Kohn-Sham) matrix. All density-dependent terms in the QM/MM dispersion and repulsion expression are actually function of the atomic volume ratio $\gamma_A[\mathrm{n_m(r)}]$, therefore one obtains:

$$F_{vdW}^{ms} = \frac{\partial E_{vdW}^{ms}}{\partial \gamma_A[\mathrm{n_m(r)}]} \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} =$$

$$-\sum_{AB} \left[\frac{\partial f_{AB}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} C_{6,AB}[n_m(r)] + \frac{\partial C_{6,AB}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} f_{AB}[n_m(r)] \right] \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} R_{AB}^{-6} \tag{17}$$

$$F_{rep}^{ms} = \frac{\partial E_{LJ,rep}^{ms}}{\partial \gamma_A[\mathrm{n_m(r)}]} \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} =$$

$$\frac{1}{2}\sum_{AB} \left[\frac{\partial C_{6,AB}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} (R_{AB}^{0})^{6} + 6(R_{AB}^{0})^{5} \frac{\partial R_{AB}^{0}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} C_{6,AB}[n_m(r)] \right] \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} R_{AB}^{-12} \tag{18}$$

where the partial derivatives with respect to the atomic volume ratio γ_A are

$$\frac{\partial f_{AB}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} = \frac{\partial f_{AB}[n_m(r)]}{\partial R_{AB}^{0}[\mathrm{n_m(r)}]} \frac{\partial R_{AB}^{0}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} =$$

$$-\left(1 + e^{\left\{-d\left(\frac{R_{AB}}{s_r R_{AB}^{0}} - 1\right)\right\}}\right)^{-2} \frac{dR_{AB}}{s_r (R_{AB}^{0})^{2}} e^{\left\{-d\left(\frac{R_{AB}}{s_r R_{AB}^{0}} - 1\right)\right\}} \frac{\partial R_{AB}^{0}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} \tag{19}$$

$$\frac{\partial R_{AB}^{0}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} = \frac{R_{A,free}^{0}}{3\gamma_A[\mathrm{n_m(r)}]^{2/3}} \quad (20)$$

$$\frac{\partial C_{6,AB}[n_m(r)]}{\partial \gamma_A[\mathrm{n_m(r)}]} = \gamma_B C_{6,AB}^{free} \quad (21)$$

By introducing the expressions (19)-(21) in the "vdW" and "rep" terms of the Fock operator, we get:

$$F_{vdW}^{ms} = \sum_{AB} \left[f_{AB}[n_m(r)] \frac{dR_{AB}}{s_r (R_{AB}^{0})^2} e^{\left\{-d\left(\frac{R_{AB}}{s_r R_{AB}^{0}}-1\right)\right\}} \frac{R_{A,free}^{0}}{3\gamma_A[\mathrm{n_m(r)}]^{2/3}} C_{6,AB}[n_m(r)] - \right.$$

$$\left. \gamma_B C_{6,AB}^{free} f_{AB}[n_m(r)] \right] f_{AB}[n_m(r)] \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} R_{AB}^{-6} \quad (22)$$

$$F_{rep}^{ms} = \sum_{AB} \left[\gamma_B C_{6,AB}^{free} \frac{R_{AB}^{0}}{2} + \frac{R_{A,free}^{0}}{\gamma_A[\mathrm{n_m(r)}]^{2/3}} C_{6,AB}[n_m(r)] \right] (R_{AB}^{0})^5 \frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} R_{AB}^{-12} \quad (23)$$

where, from eq. (3), the derivative of the atomic volume ratio takes the simple form:

$$\frac{\partial \gamma_A[\mathrm{n_m(r)}]}{\partial P_{\mu\nu}} = \frac{2}{V_A^{free}} \int r^3 \omega_A(r) \chi_\mu(r) \chi_\nu(r) dr^3 \quad (24)$$

3. Computational Details

The self-consistent dispersion-repulsion model is here applied to ground state interaction energies computed for the S22 benchmark set, which is designed to represent typical noncovalent interactions in biological systems, including complexes with predominant hydrogen bond and dispersion contributions, as well as a mixture of these two.[31] Calculations are based on the PBE functional[32] combined with the 6-31G(d), 6-31+G(d) or 6-311++G(d,p) basis sets. Single point calculations were performed for the complexes and the individual molecules at the S22 benchmark geometries, and interaction energies were corrected for basis-set superposition errors (BSSE) using the counterpoise correction (CP).[33]

The performance of the QM/MM dispersion-repulsion implementation was then tested using different strategies, which differ in (i) the way the γ_i values representing the atomic volume ratios of the MM atoms were parametrized and (ii) if the calculations are based on densities modified due to electrostatics and polarization or only electrostatics, as described using the polarizable or electrostatic embedding potentials.

Electrostatic and polarization terms are described through the MMPol polarizable embedding scheme in which the embedding potential consists of atom-centered charges and isotropic polarizabilities, in addition to the γ_i dispersion parameters.[6] In particular,

we adopted the AL set of isotropic polarizabilities derived by Wang and co-workers,[34] and partial charges derived using the Polchat tool[35] in order to account for self-polarization effects.[36] In electrostatic embedding calculations no polarizabilities were assigned to the MM atoms and the charges were derived following the RESP approach[37] using the Antechamber module of the Amber 14 software.[38] Finally, electrostatic and polarizable embedding calculations were also performed adopting the LJ classical empirical potentials from the Amber ff14SB[39] and ff12pol[40] force fields, respectively.

Different sets of γ_i parameters were considered. First, specific sets of γ_i values were parametrized for each atom in the individual molecules making up the S22 set. These calculations were based on the PBE functional using the standard QM formulation of the vdW^{TS} dispersion term (set A). In order to investigate the degree of transferability of the dispersion parameters and avoid the need for *ad hoc* parametrizations, we then computed and applied the average γ_i values for each element (set B) present in the S22 set of molecules (avoiding repeated molecules), as well as a more flexible set in which separate values were averaged for polar and apolar Hs and for each hybridization state of the C, N and O atoms (set C). Finally, we explored an alternative scheme in which the γ_i values were directly approximated by the ratio $\sim\alpha_A^{MM}/\alpha_A^0$, where α_A^{MM} is the isotropic polarizability used to model the polarization term in the expression for the embedding potential and α_A^0 denotes the free-atom reference value (set D).

In all cases, the parametrization of the MM γ_i values and point charges was performed at the PBE/6-31G(d), PBE/6-31+G(d) or PBE/6-311++G(d,p) levels of theory. In order to investigate the dependence of the γ_i values on the functional adopted, however, additional calculations were performed at the B3LYP/6-311++G(d,p) level of theory.

We note that, for each complex in the S22 benchmark set, QM/MM interaction energies were computed twice, either considering one or the other interacting molecules at the corresponding QM or MM level of description.

Finally, QM and QM/MM interaction energy profiles were computed for selected complexes by increasing/decreasing the intermolecular distance along the vector connecting the geometric centers of the molecules while keeping the internal geometries of the fragments frozen.

All calculations were performed in a locally modified version of the Gaussian development version,[41] in which we implemented the QM vdW^{TS} and QM/MM dispersion+repulsion scheme. We used the free atom reference values (α_A^0 , $C_{6,AA}$ and

$R^0_{A,free}$) and damping factor parameters ($d = 20$ and $s_r = 0.94$) adopted in the original implementation of the vdW^{TS} scheme by Tkatchenko and Scheffler[22], with α^0_A and $C_{6,AA}$ values taken from the database reported by Chu and Dalgarno and free-atom vdW radii $R^0_{A,free}$ equal to 1.64, 1.90, 1.77 and 1.66 Å for H, C, N and O atoms, respectively.[42] For polar hydrogens bound to heteroatoms, the latter was adjusted to values 0.7 and 1.3 Å for hydrogens participating or not in hydrogen bonds, respectively.

4. Results and discussion

In order to validate the implementation, we computed interaction energies for the 22 non-covalent complexes contained in the popular S22 benchmark set.[31,43] In Table 1 we provide a summary of the mean absolute errors (MAE) obtained in each case, as compared to the CCSD(T)/CBS-CP benchmark data (CBS – complete basis set). All the CP corrected interaction energies calculated for each complex at the PBE/6-31G(d), PBE/6-31+G(d) or PBE/6-311++G(d,p) levels of theory are reported in Tables S1, S2 and S3, respectively.

The errors reported in Table 1 clearly illustrate the drastic improvement of the results upon the addition of the vdW^{TS} term to the energies. When the vdW^{TS} contribution is not included, the performance of the PBE functional with the basis sets adopted is rather poor, with MAEs close to 3 kcal/mol in all cases. The addition of the dispersion term, however, lowers the errors to ~0.4-0.6 kcal/mol, depending on the basis set adopted. Interestingly, the further flexibility of the 6-311++G(d,p) basis set provides results essentially with the same quality as 6-31+G(d) basis set. Because dispersion is accounted for through the vdW^{TS} term, the importance of the good description of polarizability through a large basis set is in this case less important.

Table 1. Mean absolute errors (kcal/mol) in interaction energies computed at different levels of theory for the S22 benchmark set as compared to reference CCSD(T)/CBS-CP data. The set includes 7 hydrogen-bonded (HB) complexes, 8 complexes dominated by dispersion interactions (DD), and 7 complexes of mixed character (Mix).

	6-31G(d)		6-31+G(d)		6-311++G(d,p)	
	$\Delta E_{PBE\text{-}CP}$	$\Delta E_{PBE+TS\text{-}CP}$	$\Delta E_{PBE\text{-}CP}$	$\Delta E_{PBE+TS\text{-}CP}$	$\Delta E_{PBE\text{-}CP}$	$\Delta E_{PBE+TS\text{-}CP}$
HB	1.05	0.90	1.64	0.66	1.24	0.57
DD	5.40	0.68	4.85	0.34	4.77	0.40
Mix	1.82	0.25	1.92	0.25	1.92	0.25
All	2.88	0.61	2.90	0.41	2.74	0.41

If we focus on our best estimates calculated at the PBE/6-311++G(d,p)-CP level of theory, shown in Fig. 1, the incorporation of the vdW^{TS} term greatly reduces the errors from 2.74 to 0.41 kcal/mol considering all the noncovalent complexes, leading to an excellent correlation between calculated and reference CCSD(T) data (slope and squared correlation coefficients of 1.02 and 0.99). As expected the impact of the vdW^{TS} term is stronger for dispersion-dominated complexes, but the errors in all three classes of complexes considered is lowered. In particular, the MAEs for hydrogen-bonded, dispersion-dominated and mixed character complexes are 0.57, 0.40 and 0.25 kcal/mol when employing the largest basis set. Such errors compare well with the reported errors of 0.46, 0.30 and 0.14 kcal/mol for PBE in the original implementation of the vdW^{TS} scheme.[22] The minor differences obtained can be attributed to the different basis sets employed, and thereby overall demonstrate the correctness of our implementation.

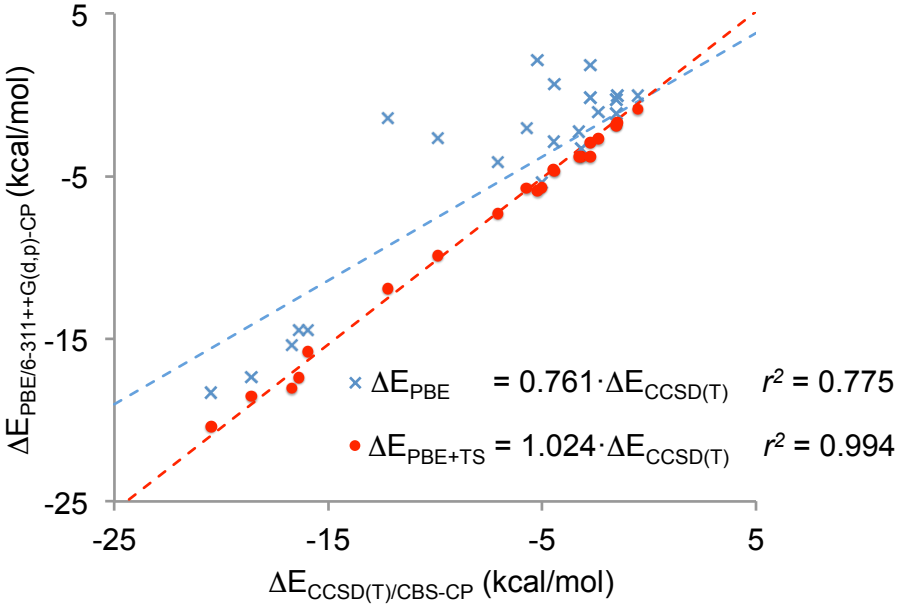


Figure 1. Interaction energies computed for the S22 benchmark set at the PBE/6-311++G(d,p)-CP and PBE+TS/6-311++G(d,p)-CP levels of theory compared to CCSD(T)/CBS-CP reference data.

We then explored the performance of the QM/MM dispersion-repulsion formulation, with special attention to the transferability of the atomic volume ratio parameters γ, which are defined as a fixed parameter for each atom in the MM region. Thus, first we

parametrized the γ parameters from PBE+TS and B3LYP+TS calculations performed on the isolated molecules (set A). In this scheme, the γ parameters are computed *ad hoc* for each molecule, in a strategy reminiscent of the derivation of individual point charge models in classical force fields. We then averaged the resulting γ parameters over elements, as obtained for all unique molecules included in the S22 set, and we recomputed the interaction energies based on such averaged parameters (set B). We also explored a more flexible definition of the atom types, distinguishing between polar and apolar hydrogen atoms, and among the hybridization states of the heavy atoms (set C), given the known relation among polarizability and hybridization. Finally, we explored a simpler approach, where the γ parameters were directly estimated from the Amber AL isotropic polarizabilities used to model polarization effects in the environment region (set D). In Table 2, we report the γ parameters adopted in the B, C and D sets, as well as the corresponding standard deviations. Because only one atom of N_{sp} and N_{sp3} type was contained in the S22 set, the values for this atom types in set C correspond to a single atom.

If we focus on the transferability of the γ's, we immediately observe that the values for a given element in set B present rather small standard deviations, ranging from 0.02 – 0.03 for C and H, and somewhat larger values for N and O (around 0.05) at the PBE+TS/6-31+G(d) and PBE+TS/6-311++G(d,p) level of theory, whereas the adoption of the smaller 6-31G(d) basis set leads to slightly smaller deviations. On the other hand, the introduction of additional atom types in the set C shows smaller deviations, illustrating the fact that the γ's are more transferable among atoms with the same hybridization state. For C, γ slightly decreases when passing from sp to sp^2 and sp^3 hybridization states, as one would expect. For N, γ again decreases when moving from N_{sp} to N_{sp2}, but then it increases again when passing to N_{sp3}. As noted before, however, the N_{sp} and N_{sp3} values here are only based on a single atom in the S22 set, so its value is not averaged over different environments as it is, for example, for N_{sp2}, averaged over 16 cases.

Table 2. Atomic volume ratio parameters γ and corresponding standard deviations adopted for the molecules in the S22 benchmark set.

Set B: γ parameters averaged per element					
			γ_{PBE}		γ_{B3LYP}
	n_{data}	6-31G(d)	6-31+G(d)	6-311++G(d,p)	6-311++G(d,p)
H	70	0.57 ± 0.03	0.66 ± 0.03	0.65 ± 0.03	0.65 ± 0.03
C	56	0.79 ± 0.02	0.83 ± 0.02	0.83 ± 0.02	0.83 ± 0.02
N	18	0.88 ± 0.04	0.90 ± 0.05	0.89 ± 0.05	0.89 ± 0.05
O	10	0.89 ± 0.03	0.98 ± 0.05	0.97 ± 0.05	0.97 ± 0.05
Set C: γ parameters averaged per atom type					
	n_{data}	6-31G(d)	6-31+G(d)	6-311++G(d,p)	6-311++G(d,p)
H_{apolar}	50	0.59 ± 0.02	0.66 ± 0.02	0.65 ± 0.02	0.65 ± 0.02
H_{polar}	20	0.54 ± 0.02	0.66 ± 0.04	0.66 ± 0.04	0.65 ± 0.04
C_{sp}	3	0.76 ± 0.03	0.84 ± 0.02	0.84 ± 0.02	0.84 ± 0.02
C_{sp2}	51	0.80 ± 0.02	0.84 ± 0.02	0.83 ± 0.02	0.83 ± 0.02
C_{sp3}	2	0.75 ± 0.03	0.77 ± 0.02	0.77 ± 0.03	0.77 ± 0.03
N_{sp}	1	0.91	0.97	0.97	0.97
N_{sp2}	16	0.87 ± 0.04	0.89 ± 0.05	0.89 ± 0.05	0.89 ± 0.05
N_{sp3}	1	0.88	0.93	0.92	0.92
O_{sp2}	7	0.91 ± 0.01	1.00 ± 0.02	1.00 ± 0.02	1.00 ± 0.02
O_{sp3}	3	0.84 ± 0.01	0.92 ± 0.03	0.91 ± 0.03	0.91 ± 0.03
Set D: γ parameters estimated from AL Amber polarizabilities[a]					
		γ_{Amber}			
H		0.64			
C1 (sp)		0.78			
C2 (sp^2)		0.73			
C3 (sp^3)		0.53			
N (non-nitro N)		0.88			
O2 (sp^2)		0.76			
O3 (sp^3)		0.77			

[a] γ values approximated by the ratio $\sim\alpha_A^{MM}/\alpha_A^0$ with α_A^{MM} values from Ref [34].

Interestingly, the γ values seem to be quite robust both with respect to the choice of basis set or DFT functional. The exception is the small 6-31G(d) basis set, where significant differences appear in the derived γ's, as the inclusion of diffuse functions is rather important to correctly describe the shape of the electron densities. On the contrary, both 6-31+G(d) and 6-311++G(d,p) basis sets give virtually the same results. Moreover, the adoption of either PBE or B3LYP functionals also leads to essentially equivalent parameters. Therefore, γ values parametrized for a given functional seem to be widely transferable to be used with other functionals.

We applied the different sets of γ parameters to compute the QM/MM interaction energies, which were then compared to full QM calculations. For Set A, both electrostatic and polarizable embedding calculations were performed in order to explore the impact of explicit MM polarization on the results. For the sake of comparison, we also performed electrostatic and polarizable embedding calculations based on the classical Amber ff14SB and ff12pol LJ potentials instead of the density-dependent dispersion-repulsion term. For each AB complex in the S22 benchmark set, we performed two different calculations QM(A)/MM(B) and QM(B)/MM(A), depending on which molecule was described classically.

As a first test, PBE/6-311++G(d,p) interaction energies were computed with a polarizable embedding (Set A of γ parameters) using the free atom reference values originally developed for the vdW^{TS} scheme.[22] Quite strikingly, the QM/MM model, which includes electrostatics, polarization, dispersion and repulsion terms, is able to accurately describe non-covalent interactions in all complexes lacking hydrogen bonds with no additional tuning of the dispersion-repulsion parameters needed. For such complexes, we obtain a mean absolute error (MAE) of 1.2 kcal/mol, and only two complexes present errors slightly larger than 2 kcal/mol (adenine·thymine stacked and benzene·indole T-shaped complexes). In complexes were the interaction involves hydrogen bonds, however, repulsion is strongly overestimated leading to highly repulsive interactions in most cases (except for the ammonia and phenol dimers), which results in a MAE of 30 kcal/mol. This indicates that the density-dependent atomic volume ratio γ cannot capture the reduction of the van der Waals radius of QM hydrogens when they participate in hydrogen bonds with the MM region.[44] A more consistent treatment for both the QM and MM region thus relies on tuning the free-atom

vdW radii $R^0_{A,free}$ for hydrogen. Our results indicated that using a value of 0.7 Å for polar hydrogens allows a satisfactory description of all complexes considered, leading to a MAE of 1.4 kcal/mol. The largest error was found for the formic acid dimer, which displays the shortest hydrogen-bond distance in the set, in which the interaction is 7.8 kcal/mol higher than the reference. Further reduction of the polar hydrogen radius to improve the description of the formic acid dimer, in turn, led to a worsening of the shape of interaction energy profiles computed for other hydrogen-bonded systems, so we choose to keep the 0.7 Å value. We note that different values for van der Waals radii in aliphatic and polar hydrogens are often applied in LJ potentials of classical force fields, for example in Amber ff14SB, where the radius of aliphatic H (HC atom type) equal to 1.487 Å is reduced to 0.6 Å for the polar case (H atom type). Once we adjusted the value of the radius for polar hydrogens, we computed the interaction energies for the S22 set adopting the different schemes for γ parameters shown in Table 2.

In Table 3 we report the mean absolute errors (MAE) in QM/MM interaction energies obtained adopting the 4 sets of γ parameters as well as the LJ Amber potentials with respect to full QM PBE+TS calculations (see Tables S4-S10 for the complete list of interaction energies). Most of these QM/MM calculations are based on densities obtained by including effects of environment electrostatics and polarization. In order to investigate how environment polarization impact our results, however, we also computed the set A and the Amber LJ energies based on densities obtained using electrostatic embedding. The comparison between the Set A and Set A with electrostatic embedding, indeed shows a clear improvement when polarization is included, with the MAE decreasing from 2.8 to 1.4 kcal/mol. This improvement is due to the increased overall electrostatics in hydrogen-bonded complexes when MM polarization is accounted for. However, the results also indicate that the impact of environment electrostatics and polarization only induces very small changes in the γ parameters of the QM region, and dispersion-repulsion terms in both cases are almost identical with differences below 0.2 kcal/mol. The relative insensitivity of the γ to the impact of MM polarization supports their transferability among different environments.

Regarding the performance of the model with respect to basis set, it is somewhat unexpected that the 6-31G(d) basis set provides errors ~0.5 kcal/mol lower than those obtained with 6-31+G(d) and 6-311++G(d,p) ones for hydrogen-bonded systems. This

can be explained by two different effects. On one side, the electrostatic interactions are attenuated with 6-31G(d), but this effect is counteracted by a considerably smaller repulsion due to the systematic underestimation of atomic volumes using this basis. This effect, for instance, lowers the γ value of polar hydrogens from 0.65 to 0.54, as shown in Table 2. The changes in electrostatics and dispersion-repulsion nearly cancel each other in most systems analyzed, leading to overall differences less than 0.5 kcal/mol with the exception of the formic acid dimer, where the close hydrogen-bond distance leads to a more drastic decrease on repulsion leading to change of 3 kcal/mol, which explains the changes in MAEs obtained using different basis sets.

The impact of assuming the transferability of the γ's among different atoms can be assessed by comparing the results obtained using sets A, B and C. In this case, we observe that the MAEs are only marginally increased. For example, at the PBE+TS/6-311++G(d,p) level of theory, adopting transferable γ's for each element or atom type only increases the MAEs from 1.4 to 1.5 kcal/mol, compared to the adoption of *ad hoc* sets of γ's derived for a given molecule. We note, however, that whereas the adoption of transferable parameters has a minor impact on the energies of dispersion-dominated complexes, for hydrogen-bonded systems the MAE increases from 2.1 to 2.4 and 2.5 kcal/mol. These results are in line with the D2 and D3 dispersion corrections introduced by Grimme and co-workers for DFT, where transferability is also assumed by adopting dispersion coefficients specific per element type or hybridization state, respectively, the latter defined in terms of fractional coordination numbers.[45]

Alternatively, if we roughly estimate the γ's from the polarizabilities taken from the Amber AL force field (set D), we find very similar results compared to sets B and C, also based on transferable parameters. Indeed, the systematic underestimation of γ's for set D apparent in Table 2 leads to a similar effect to the adoption of the 6-31G(d) basis set discussed above. Thus, the results for set D with large basis sets are improved, whereas its performance when compared to sets A, B and C based on 6-31G(d) calculations show a very similar performance.

Table 3. Mean absolute errors (kcal/mol) in the interaction energies of the S22 benchmark set as estimated from full QM+TS and QM/MM calculations at the PBE level of theory using different descriptions for the MM region. Values reported in parentheses refer to an electrostatic (instead of polarizable) embedding.[a]

	6-31G(d)	6-31+G(d)	6-311++G(d,p)
		Set A	
HB	1.6 (5.6)	2.2 (5.8)	2.1 (6.2)
DD	0.4 (0.5)	0.9 (1.1)	1.0 (1.2)
Mix	0.9 (1.0)	1.1 (1.2)	1.1 (1.2)
All	0.9 (2.3)	1.4 (2.6)	1.4 (2.8)
		Set B	
HB	1.7	2.5	2.4
DD	0.4	0.9	0.9
Mix	1.0	1.2	1.1
All	1.0	1.5	1.5
		Set C	
HB	1.6	2.6	2.5
DD	0.4	0.9	0.9
Mix	0.9	1.2	1.1
All	1.0	1.5	1.5
		Set D	
HB	1.6	2.4	2.2
DD	0.4	0.7	0.8
Mix	1.0	1.1	1.1
All	1.0	1.4	1.3
		Amber ff12pol (Amber ff14SB)[a]	
HB	3.0 (1.9)	6.7 (1.4)	6.2 (1.1)
DD	0.3 (0.4)	0.6 (0.7)	0.7 (0.8)
Mix	0.7 (0.8)	0.8 (0.6)	0.8 (0.6)
All	1.3 (1.0)	2.6 (0.9)	2.5 (0.8)

[a] QM/MM calculations based on densities for the QM fragment obtained with electrostatic embedding.

We have also computed QM/MM interaction energies with electrostatic and polarizable embeddings by simply adding an empirical LJ dispersion-repulsion term taken from the Amber ff14SB and ff12pol classical force fields, as is typically done in many QM/MM applications. Amber electrostatic embedding calculations display MAEs of ~1 kcal/mol, similar to the best results obtained using the 6-31G(d) basis set for our dispersion-repulsion model. The considerable increase in the flexibility given by Amber atom types, however, seems to improve some situations, like the formic acid dimer. Such improvement however could be the result of a cancellation of errors, as the adoption of the polarizable Amber ff12pol force field leads to larger errors, although the latter is expected to provide a better description of electrostatics. The interaction energies obtained for all complexes adopting different MM descriptions are reported in Tables S4-S10. Overall, these results support the transferability of the γ's among similar atoms in different molecules, and suggest that excellent results can be obtained without needing specific parametrizations for different molecules.

Finally, we have also explored the performance of the QM/MM interaction energies along different distance profiles, in which the center-to-center separation between the molecules was increased/decreased while keeping the internal geometries of the molecules frozen at their S22 benchmark geometries, in order to assess the behavior of the model at non-equilibrium interfragment separations. The results for a selection of hydrogen bonded, dispersion-dominated and mixed complexes are shown in Fig. 2 (see Figs. S1-S3 for all complexes).

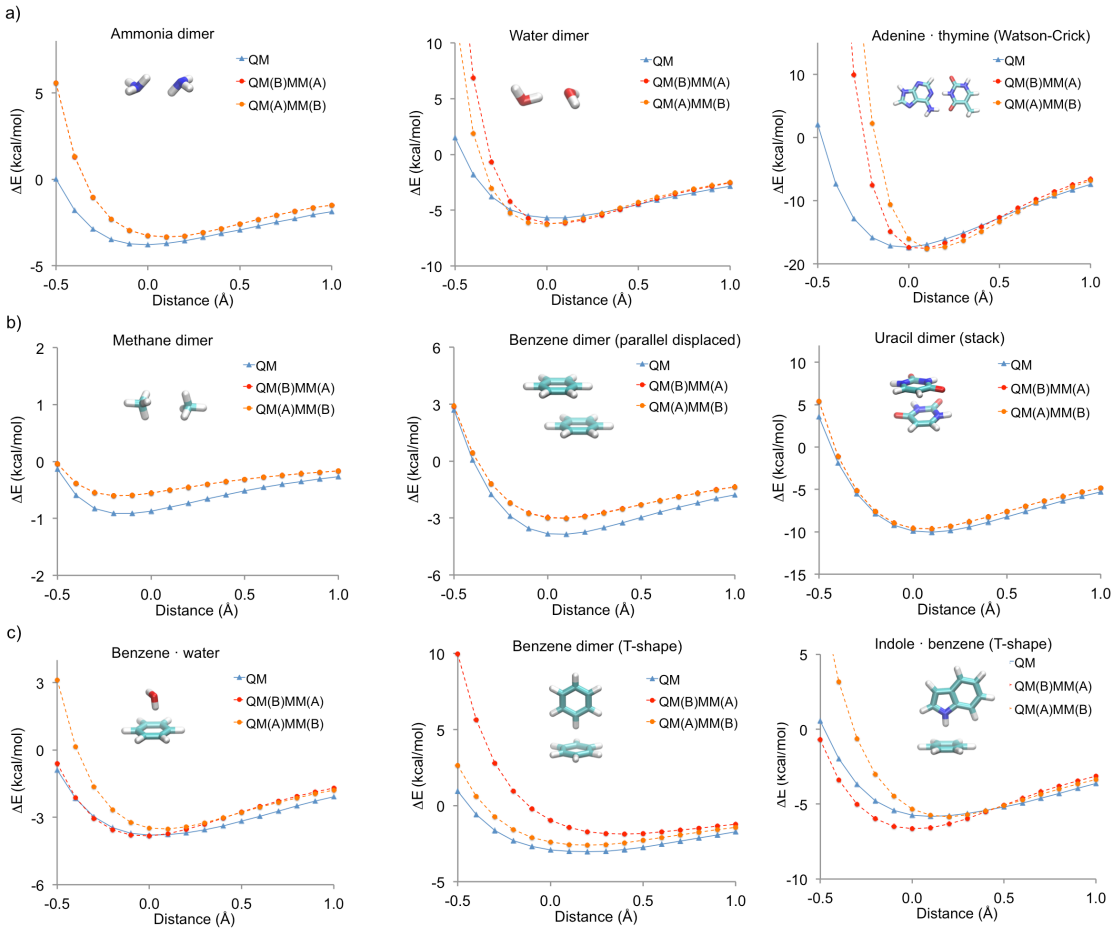


Figure 2. Comparison of full QM+TS and QM/MM (Set A – polarizable embedding) interaction energy profiles computed for selected complexes of the S22 benchmark set at the PBE/6-311++G(d,p) level of theory: a) hydrogen-bonded (HB) complexes, b) dispersion-dominated (DD) complexes and c) mixed complexes. Distances relative to the minimum energy geometry are built along the vector connecting the geometric centers.

The overall profiles are well described in terms of both the location of the minima and the well depth. Some more critical cases, however, are identifiable.

For hydrogen-bonded systems, the curve at close distances is too steep, probably due to the difficulty of describing the variation of the hydrogen vdW radius in such regime. For complexes formed by the interaction between a polar hydrogen and an aromatic ring (benzene · indole T-shape, benzene · water and benzene · ammonia), instead the optimal distances and the well depths obtained assuming a 0.7 Å free-atom radius for polar hydrogens were not well reproduced, as shown in Fig. S4. A similar behavior was also found in the weak ammonia dimer, which is characterized by a longer H···N

optimal distance with respect to the other hydrogen-bonded complexes analyzed. Thus, the adoption of the 0.7 Å free-atom radius for polar hydrogens seems to be well-suited for strong hydrogen-bonded systems, whereas for weaker complexes that radius should be somewhat enlarged. Test calculations indicated that indeed the adoption of a radius of 1.3 Å for polar hydrogens leads to a much better agreement for these complexes, and these is our final choice reflected in the profiles of Fig. 2.

Overall, our results thus show that through a minimal parametrization of the free-atom radius of polar hydrogens involved or not in strong hydrogen bonds, the final QM/MM model allows to describe quite accurately the interactions in a QM/MM boundary when structures fluctuate around equilibrium geometries due to thermal effects, for example along a molecular dynamics simulation.

Conclusions

We presented an extension of the Tkatchenko-Scheffler semiempirical van der Waals scheme vdW^{TS} aimed at describing dispersion interactions between quantum and classical regions in a QM/MM framework based on a polarizable embedding. In addition, we coupled the dispersion term with an analogous density-dependent repulsion term obtained by exploiting a LJ-like relation between the two, so that both contributions are obtained in a single step through the assignment of the dispersion coefficients for the MM sites.

While keeping the density dependence of the dispersion parameters for the QM region, we explored different ways to define the atomic volume ratios γ for the MM part of the system. First, we investigated a strategy resembling the derivation of point charge models in classical force fields, by which γ values are parametrized *ad hoc* from DFT calculations of isolated MM fragments. We then explored the transferability of these parameters by deriving element-specific, or hybridization state-specific, γ values averaged for atomic sites in different molecular environments, in analogy with the transferability assumption widely used in Lennard-Jones parameters in MM force fields. As an even simpler scheme, we tested the possibility to estimate γ values by the ratio $\sim\alpha_A^{MM}/\alpha_A^0$, where α_A^{MM} is the same isotropic polarizability used in the polarizable embedding and α_A^0 denotes the free-atom reference value.

We assessed the performance of the implementation by computing total interaction energies for the S22 benchmark set, designed to represent typical non-covalent interactions in biological systems, which includes complexes with predominant

hydrogen bond and dispersion contributions, as well as systems of mixed character. Our results indicate a good performance of the QM/MM implementation, with mean absolute errors ~1-2 kcal/mol in interaction energies compared to reference full QM calculations computed at the S22 benchmark geometries. To obtain such an agreement, we find it necessary to calibrate only the free-atom radius of the hydrogen atoms.

Interestingly, we find similar performance either when dispersion parameters are parametrized on the basis of vdW^{TS} calculations on separate fragments or when they are estimated from the polarizabilities taken from the polarizable force field. Moreover, our results show that γ values are widely transferable, and either element-specific or more flexible hybridization state-specific parameters can be used to model dispersion-repulsion energies without a significant deterioration of the accuracy of the model. In addition, we find a weak dependence on basis set and virtually insensitivy to the choice of DFT functional used in their derivation. This suggests that γ values derived at a given QM level of theory are also widely transferable to calculations based on other QM methods.

Finally, we computed interaction energy profiles by increasing/decreasing the center-to-center separation between fragments in the complexes of the S22 set while keeping the internal geometries frozen, in order to assess the performance of the model in out-of-equilibrium inter-fragment separations. These results show a good agreement with reference full QM calculations for complexes dominated by hydrogen bonds, dispersion, or in mixed-character complexes. This indicates that the present implementation can be extended to describe dispersion-repulsion interactions through a QM/MM boundary when structures fluctuate around equilibrium geometries, such as along a molecular dynamics simulation.

Overall, the novel implementation has shown to be a promising strategy to include dispersion-repulsion effects in multiscale QM/MM models by incorporating an explicit dependence on the QM electronic density.

Supporting Information

The Supporting Information is available free of charge via the Internet at http://pubs.acs.org

Tables of interaction energies and figures of interaction energy profiles computed at QM and QM/MM levels of theory.

Acknowledgements

C.C. is a Serra Húnter Fellow (Generalitat de Catalunya) and acknowledges financial support from the Spanish Ministerio de Economía y Competitividad (MINECO; grants RYC2011-08918 and CTQ2012-36195). J.K. thanks the Danish Council for Independent Research (the Sapere Aude program) and the Villum Foundation. L.F. and A.H.S. acknowledge support from the Research Council of Norway through its Centres of Excellence scheme (project number 262695).

References

(1) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chemie-International Ed.* **2009**, *48*, 1198–1229.

(2) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.

(3) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An Effective Fragment Method for Modeling Solvent Effects in Quantum Mechanical Calculations. *J. Chem. Phys.* **1996**, *105*, 1968–1986.

(4) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. A Discrete Solvent Reaction Field Model within Density Functional Theory. *J. Chem. Phys.* **2003**, *118*, 514–521.

(5) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. Protein Influence on Electronic Spectra Modeled by Multipoles and Polarizabilities. *J. Chem. Theory Comput.* **2009**, *5*, 649–658.

(6) Curutchet, C.; Munoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(7) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(8) Lipparini, F.; Cappelli, C.; Barone, V. Linear Response Theory and Electronic Transition Energies for a Fully Polarizable QM/Classical Hamiltonian. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.

(9) Severo Pereira Gomes, A. A.; Jacob, C. R. Quantum-Chemical Embedding Methods for Treating Local Electronic Excitations in Complex Chemical Systems. *Annu. Rep. Prog. Chem., Sect. C Phys. Chem.* **2012**, *108*, 222–277.

(10) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited States in Large Molecular Systems through Polarizable Embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

(11) London, F. The General Theory of Molecular Forces. *Trans. Faraday Soc.* **1937**, *33*, 8–26.

(12) Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1924**, *106*, 463–477.

(13) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *28*, 671–690.

(14) Van Duijnen, P. T.; de Vries, A. H. Direct Reaction Field Force Field: A Consistent Way to Connect and Combine Quantum-Chemical and Classical Descriptions of Molecules. *Int. J. Quantum Chem.* **1996**, *60*, 1111–1132.

(15) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. Chapter 10 The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions. *Annual Reports in Computational Chemistry*. **2007**, *3*, 177–193.

(16) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An Effective Fragment Method for Modeling Solvent Effects in Quantum Mechanical Calculations. *J. Chem. Phys.* **1996**, *105*, 1968–1986.

(17) Bertoni, C.; Slipchenko, L. V.; Misquitta, A. J.; Gordon, M. S. Multipole Moments in the Effective Fragment Potential Method. *J. Phys. Chem. A* **2017**, *121*, 2056–2067.

(18) Slipchenko, L. V.; Gordon, M. S.; Ruedenberg, K. Dispersion Interactions in QM/EFP. *J. Phys. Chem. A* **2017**, *121*, 9495–9507.

(19) Giese, T. J.; York, D. M. Charge-Dependent Model for Many-Body Polarization, Exchange, and Dispersion Interactions in Hybrid Quantum Mechanical⁄molecular Mechanical Calculations. *J. Chem. Phys.* **2007**, *127*, 194101.

(20) Kuechler, E. R.; Giese, T. J.; York, D. M. Charge-Dependent Many-Body Exchange and Dispersion Interactions in Combined QM/MM Simulations. *J. Chem. Phys.* **2015**, *143*, 234111.

(21) Giovannini, T.; Lafiosca, P.; Cappelli, C. A General Route to Include Pauli

Repulsion and Quantum Dispersion Effects in QM/MM Approaches. *J. Chem. Theory Comput.* **2017**, *13*, 4854–4870.

(22) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 73005.

(23) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Tkatchenko-Scheffler van Der Waals Correction Method with and without Self-Consistent Screening Applied to Solids. *Phys. Rev. B* **2013**, *87*, 64110.

(24) Ferri, N.; DiStasio, R. a.; Ambrosetti, A.; Car, R.; Tkatchenko, A. Electronic Properties of Molecules and Surfaces with a Self-Consistent Interatomic van Der Waals Density Functional. *Phys. Rev. Lett.* **2015**, *114*, 1–5.

(25) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(26) Becke, A. D. A Multicenter Numerical Integration Scheme for Polyatomic Molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(27) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling in Exchange-Correlation Density Functional Quadratures. *Chem. Phys. Lett.* **1996**, *257*, 213–223.

(28) Keith, T. A.; Frisch, M. J. Subshell Fitting of Relativistic Atomic Core Electron Densities for Use in QTAIM Analyses of ECP-Based Wave Functions. *J. Phys. Chem. A* **2011**, *115*, 12879–12894.

(29) Amovilli, C.; Mennucci, B. Self-Consistent-Field Calculation of Pauli Repulsion and Dispersion Contributions to the Solvation Free Energy in the Polarizable Continuum Model. *J. Phys. Chem. B* **1997**, *101*, 1051–1057.

(30) McWeeny, R. *Methods of Molecular Quantum Mechanics*; Academic Press, 1992.

(31) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(32) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(33) Boys, S.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors.

Mol. Phys. **1970**, *19*, 553–566.

(34) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.

(35) Caprasecca, S.; Curutchet, C.; Mennucci, B.; Jurinovich, S. *PolChat: A Polarisation-Consistent Charge Fitting Tool, Molecolab Tools*; University of Pisa: 2017. molecolab.dcci.unipi.it/tools/polchat (accessed July 31, 2017).

(36) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **2011**, *115*, 3100–3111.

(37) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(38) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*. University of California: San Francisco, 2014.

(39) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(40) Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M. J.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations. 4. van Der Waals Parametrization. *J. Phys. Chem. B* **2012**, *116*, 7088–7101.

(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Janesko, B. G.; Lipparini, F.; Zheng, G.; Sonnenberg, J. L.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery J. A., J.; Peralta, J. E.;

Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, T.; Keith, T.; Kobayashi, R.; Normand, J. Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Parandekar, P. V.; Mayhall, N. J.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian Development Version, Revision H.36*; Gaussian, Inc.: Wallingford CT, 2010.

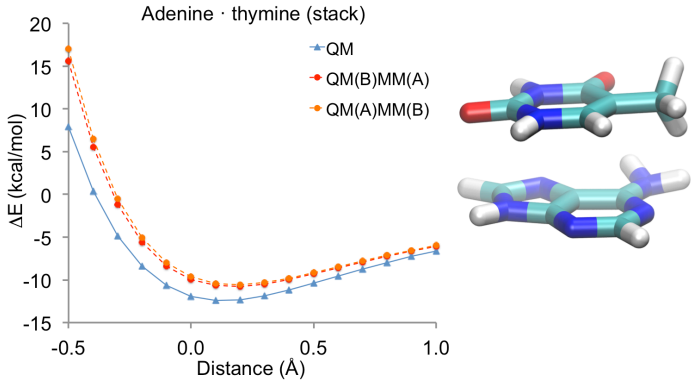
(42) Chu, X.; Dalgarno, A. Linear Response Time-Dependent Density Functional Theory for van Der Waals Coefficients. *J. Chem. Phys.* **2004**, *121*, 4083–4088.

(43) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(44) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J. Defining the Hydrogen Bond: An Account (IUPAC Technical Report). *Pure Appl. Chem.* **2011**, *83*, 1619–1636.

(45) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.

TOC Graphic


Adenine · thymine (stack)
QM
QM(B)MM(A)
QM(A)MM(B)
ΔE (kcal/mol)
20
15
10
5
0
-5
-10
-15
-0.5
0.0
0.5
1.0
Distance (Å)