Achieving linear scaling in computational cost for a fully polarizable MM/Continuum embedding

Stefano Caprasecca,[†] Sandro Jurinovich,[†] Louis Lagardère,[‡] Benjamin Stamm,[¶,§] and Filippo Lipparini[*,‡,‖,¶,⊥]

Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via Giuseppe Moruzzi 3, I-56124, Pisa, Italy, Sorbonne Universités, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France, Sorbonne Universités, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, CNRS, UMR 7598 and 7616, F-75005, Paris, France, Sorbonne Universités, UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France, and current address: Institut für Physikalische Chemie, Universität Mainz, Duesbergweg 10-14, D-55128 Mainz, Germany

E-mail: flippari@uni-mainz.de

[*]To whom correspondence should be addressed

[†]Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via Giuseppe Moruzzi 3, I-56124, Pisa, Italy

[‡]Sorbonne Universités, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France

[¶]Sorbonne Universités, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

[§]CNRS, UMR 7598 and 7616, F-75005, Paris, France

[‖]Sorbonne Universités, UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France

[⊥]current address: Institut für Physikalische Chemie, Universität Mainz, Duesbergweg 10-14, D-55128 Mainz, Germany

Abstract

In this paper, we present a new, efficient implementation of a fully polarizable QM/MM/Continuum model based on an induced-dipoles polarizable force field and on the Conductor-like Screening Model as a polarizable continuum in combination with a self-consistent field QM method. The paper focuses on the implementation of the MM/Continuum embedding, where the two polarizable methods are fully coupled to take into account their mutual polarization. With respect to previous implementations, we achieve for the first time a linear scaling with respect to both the computational cost and the memory requirements without limitations on the molecular cavity shape. This is achieved thanks to the use of the recently developed ddCOSMO model for the continuum and the Fast Multipole Method for the force field, together with an efficient iterative procedure. Therefore, it becomes possible to include in the classical layer as much as several tens of thousands of atoms with a limited computational effort.

1 Introduction

In the last decades, multiscale strategies have been the object of a wide and fruitful development.[1–10] Such methodologies, including QM/MM methods and QM/Continuum methods, allow to account for the presence of a chemical environment in the description of a molecular system at an affordable computational price, increasing both the possibilities of molecular modeling and the manifold of treatable systems. Continuum models[8,11–13] have originally been developed as solvation models, but can now be used to describe a large and diverse number of environments, ranging from the description of membranes to metal nanoparticles.[8,14–17] These models introduce the effect of the environment by surrounding the molecule with a uniform, infinite, polarizable dielectric: polarization is mutual, in the sense that the field created by the solute polarizes the continuum, which in turn creates an electric field, usually referred to as *reaction field*, that polarizes the solute. Continuum models are therefore capable of dealing with the long-range solute-solvent electrostatic interactions accurately;

however, while they are able to describe the average solvent effect, they fail in reproducing the directional solute-solvent interactions, including, for instance, hydrogen bonding. When specific interactions are relevant, or when the environment is strongly inhomogeneous, as it is the case for biological environments, QM/MM models are more suitable.

In the classical formulation of such models, the MM system is described by employing a set of fixed point charges, and the QM/Classical interaction is that between the quantum-mechanical electronic density and such charges. While QM/MM models can be very effective in including short range interactions, a major shortcoming arises when long range interactions are considered, particularly for large systems extending tens of Ångströms away from the QM system. This is due to the fact that, in order to capture the dynamic effect of the environment, various configurations must be sampled and the QM/MM computation must be repeated for each. This can be computationally very demanding and limits in practice the application of QM/MM models. A possible solution[18–28] would be to combine QM/MM and QM/Continuum models, and employ the former in the description of a not-too-large short-range region around the QM core, and the latter to include long-range effects, limiting therefore the size of the configuration space that has to be sampled.

However, the use of fixed point charges to describe the MM system implies that, while the QM density is polarized by the MM electric field, the opposite is not true. This is a serious limitation, as it has been shown that the explicit response contribution arising from the environment polarization can be crucial, particularly when charged or very polar systems are studied, or when electron excitation processes are considered.[29–36] For this reason, in the last decade various polarizable QM/MM models have been developed, based on induced dipoles,[37–42] fluctuating charges[21,35,36,43,44] or Drude's oscillators,[33,45] to explicitly include polarization effects in the MM force field: such models can be successfully used to describe environmental effects on a large manifold of molecular properties, including electric, magnetic, mixed and vibrational ones. Polarizable QM/MM models are more expensive than their non-polarizable counterparts as they require one to solve a set of linear equations for

each value of the electronic density (and therefore, for instance, at each SCF cycle). Since the computational complexity depends on the number of polarizable MM atoms, a way to limit it would be to reduce the number of atoms, by only selecting a not too large polarizable layer around the QM subsystem. However, this is rarely an acceptable solution, as recent studies have shown that the effect of environment polarization actually extends for longer ranges than previously believed, up to 20 Å in some cases, particularly when studying electronic excitations, and especially in ordered, possibly charged environments, such as proteins.[46]

In the last few years, various fully polarizable QM/MM/Continuum approaches have been proposed and implemented,[33,35–37,43,47,48] where on top of employing a polarizable discrete model, one also includes a further external polarizable continuum layer, coupled to the former. For what concerns the description of complex environments, this solution takes advantage of the strengths of each method, as it exploits the polarizable discrete model at short range and the continuum one for long range polarization. However, the resulting coupled linear systems become easily non treatable with standard linear algebra methods even for small systems, making the use of iterative techniques mandatory.[48,49] For large systems, the continuum solvation part alone can become extremely expensive when standard implementations are used, making the overall computation, which, we remark, has to be repeated for a large number of times due to statistical sampling issues, unfeasible.

Recently, we have introduced a completely new implementation[50–53] of the Conductor-like Screening Model[13,54] (COSMO) which is suitable to treat large and very large systems, as its computational cost and memory requirements vary linearly with respect to the size of the solute. The global procedure is extremely efficient, being hundreds or thousands of times faster than previous implementations[52] and providing smooth potential energy surfaces. This new implementation, which is based on Schwarz's domain decomposition and that we call ddCOSMO, has been recently coupled to dipole-based polarizable force fields in the context of molecular dynamics simulations[55] and has proven to be fast enough to be used for molecular dynamics simulation on small proteins (up to 1000 atoms). Such an implementation,

however, relies on the existing, quadratic-scaling code for the polarization equations[49] and introduces further quadratic steps to compute the coupling terms between the continuum and the force field. This limits the size of treatable systems, as these quadratic-scaling terms dominate the overall computation for large enough systems.

This paper is devoted to overcoming these limitations: we present here a fully polarizable MM/Continuum embedding scheme, which has been coupled to self-consistent field based QM methods, which shows linear scaling properties both in the computational cost and in memory requirements. Such a scheme is based on ddCOSMO for the continuum and on the use of the Fast Multipole Method[56] (FMM) for both the polarizable force field and for the couplings. The FMM parameters have been optimized for the various computations in order to achieve an optimal compromise between accuracy and efficiency and an efficient iterative procedure has been set up. The scheme is based on the MMPol model proposed by Mennucci's group for the force field polarization,[37] but the results presented here can, in principle, be extended to other models. To the best of our knowledge, this is the first linear scaling implementation of a polarizable MM/Continuum embedding scheme with a molecular shaped cavity. This allows to overcome many limitations on size and complexity that affected previous applications, and to explore processes in potentially very large systems previously non treatable at this level of theory. Given these assumptions, and considering the ability of fully polarizable models to accurately describe complex environments, both at short and long distances, the most natural applications seem to be those involving excited state processes in large biological systems, such as proteins.

We remark that this paper focuses on the classical, polarizable MM/Continuum embedding of a QM core in a multiscale computation and on how to make the classical part efficient and achieve linear scaling in computational cost with respect to its size. We will not deal here with the QM computation nor with the coupling between QM methods, polarizable force fields and polarizable continuum models, which have already been extensively discussed in the literature. In particular, for ddCOSMO, an exhaustive presentation of the various quan-

tities that one needs to compute in order to couple it with a QM method has recently been published by some of us[53]

This paper is organized as follows. In section 2, we recap briefly the ddCOSMO method and the MMpol model; in section 3, we discuss the implementation and in particular the use of the FMM for our model; in section 4, we discuss in terms of accuracy and efficiency the effect of varying the FMM parameters and we determine an optimal set thereof; in section 5, we report some numerical tests on the global scaling of the method followed by some conclusions and perspectives in section 6.

2 Theory

In this section, we will briefly recapitulate the ddCOSMO and MMPol equations and derive the coupling. We will focus, in particular, on the coupling between the induced dipoles and the continuum.

2.1 Domain decomposition for COSMO

The domain decomposition strategy[50] we developed for the Conductor-like Screening Model[13] is based on an integral equation reformulation of Schwarz's alternating method applied to the COSMO cavity, which, in this paper, is a Van der Waals cavity, where for each atom a sphere is introduced. Therefore, the Van der Waals cavity is decomposed into elementary subdomains (or shapes) which, of course, consist of spheres. The resulting equations are a set of block-sparse linear equations of the following form:

$$\begin{pmatrix} L_{11} & \dots & L_{1M} \\ \vdots & \ddots & \vdots \\ L_{M1} & \dots & L_{MM} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_M \end{pmatrix} = \begin{pmatrix} g_1 \\ \vdots \\ g_M \end{pmatrix}. \tag{1}$$

The matrix L is composed by M^2 blocks, where M is the number of spheres (which corresponds to the number of atoms for a Van der Waals cavity). Each block is of dimension $(N+1)^2$, where N is the maximum angular momentum used for the spherical harmonics expansion of the local problems. The blocks are defined as specified in the following.[51] The diagonal blocks L_{jj} are themselves diagonal:

$$[L_{jj}]_{ll'}^{mm'} = \frac{4\pi r_j}{2l+1}\delta_{ll'}\delta_{mm'}, \tag{2}$$

where r_j is the radius of the j-th sphere. The off-diagonal blocks L_{jk} are nonzero if and only if the spheres placed on thee j-th and k-th atoms intersect; their expression is

$$[L_{jk}]_{ll'}^{mm'} = -\sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{y}_n) W_n^{jk} \frac{4\pi r_k}{2l'+1} (t_n^{jk})^{l'} Y_{l'}^{m'}(\mathbf{s}_n^{jk}). \tag{3}$$

In eq. 3, the set $\{\mathbf{y}_n, w_n\}_{n=1}^{N_g}$ is composed by the N_g points and weights of a Lebedev grid, W_n^{jk} is a function, evaluated at the n-th Lebedev point, which determines whether the spheres j and k intersect or not and $\mathbf{v}_n^{jk} = t_n^{jk}\mathbf{s}_n^{jk}$ is a vector pointing from the center of sphere j to the point $\mathbf{y}_n$ on the k-th sphere, normalized by the radius of the k-th sphere. All the details and a complete derivation can be found in ref. 50. The sparsity of the ddCOSMO matrix is what determines the scaling properties of the algorithm. We solve the linear system using Jacobi iterations and Pulay's Direct Inversion in the Iterative Subspace[51] to accelerate convergence; at each iteration, a matrix-vector product has to be computed: as the matrix is (very) sparse, only $\mathcal{O}(M)$ floating point operations are needed. The right-hand side to eq. 1 is defined as follows:

$$[g_j]_l^m = \sum_n w_n Y_l^m(\mathbf{y}_n) U_n^j \Phi_n^j, \tag{4}$$

where $U_n^j = 1 - \sum_{k\in\mathcal{N}_j} W_n^{jk}$, $\mathcal{N}_j$ is the list of spheres that intersect the j-th sphere and Φ_n^j is the solute's potential at $\mathbf{y}_n$ on sphere Ω_j.

2.2 The MMPol model and its coupling with ddCOSMO

The MMPol family of models describes the electrostatics of a molecular system by endowing each atom of a point charge and either each atom or a group of atoms of an atomic polarizability. In order to simplify the notation, we will assume that every MM atom is polarizable; a generalization to a mixed treatment is of course straightforward. In the MMPol family, there are various different models which differ in how the close-neighbors and polarization interactions are treated. In this paper, we will focus on one of these models, the one developed by Wang and coworkers. A complete description of the other models can be found in references 57, 58. The charge-charge interaction is the classical Coulomb interaction of a set of point charges, which can be modified by removing or scaling the interactions between 1-2, 1-3 and 1-4 neighbors, according to the following expression:

$$\mathcal{E}_{qq} = \frac{1}{2}\sum_{i=1}^{N_c}\sum_{j\neq i}^{N_c} s_{ij}^{qq}\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{5}$$

where N_c is the number of MM atoms. The factors s_{ij}^{qq} account for the scaling/removal of close neighbors interactions; for the model here considered they are equal to zero for 1-2 and 1-3 neighbors and to one for every other couple. The induced dipoles are determined by minimizing the following energy function:[49]

$$\mathcal{E}[\boldsymbol{\mu}] = \frac{1}{2}\boldsymbol{\mu}^\dagger \mathbf{T}\boldsymbol{\mu} - \mathbf{E}^\dagger\boldsymbol{\mu}. \tag{6}$$

In eq. 6, $\boldsymbol{\mu}$ is a $3N_c$ dimensional vector collecting the N_c induced dipoles, $\mathbf{E}$ a $3N_c$ dimensional vector collecting the electric field produced by the permanent charges at the MM atoms; again, depending on the specific polarization model, the charge-dipole interactions between 1-2, 1-3 and 1-4 neighbors can be either screened or excluded, so the field at the i-th polarizable site will be, in general,

$$\mathbf{E}_i = \sum_{j\neq i}^{N_c} s_{ij}^{qd} q_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}, \tag{7}$$

where the s_{ij}^{qd} scaling factors depend on the model. For the MMPol model, the s_{ij}^{qd} factors are equal to zero for 1-2 and 1-3 neighbors and to one for every other couple. Finally, the matrix $\mathbf{T}$ describes the interaction between the induced dipoles. Such interaction is the Coulomb interaction, which can be screened, damped or both depending on the specific model in order to avoid the possibility of a polarization catastrophe. In any case, the $\mathbf{T}$ matrix is symmetric, positive definite and is made by $N_c \times N_c$ blocks T_{ij}, each one of which is a 3-dimensional rank 2 tensor.[49] The off-diagonal blocks describe the interaction between the two dipoles $\boldsymbol{\mu}_i, \boldsymbol{\mu}_j, i \neq j$, while the diagonal ones are the inverse polarizability tensors α associated with the various polarization sites:

$$\mathbf{T} = \begin{pmatrix} \alpha_1^{-1} & T_{12} & T_{13} & \dots & T_{1M} \\ T_{21} & \alpha_2^{-1} & T_{23} & \dots & T_{2M} \\ T_{31} & T_{32} & \ddots & & \\ \vdots & \vdots & & & \vdots \\ T_{M1} & T_{M2} & & \dots & \alpha_M^{-1} \end{pmatrix}, \tag{8}$$

where

$$T_{ij} = -s_{ij}^{dd} \frac{1}{r_{ij}^3} \lambda_3(u_{ij}) \mathbf{1}_3 + s_{ij}^{dd} \frac{3}{r_{ij}^5} \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix} \lambda_5(u_{ij}). \tag{9}$$

In eq. 9, λ_3 and λ_5 are damping functions which depend on $u_{ij} = r_{ij}/[a(\langle\alpha_i\rangle\langle\alpha_j\rangle)^{1/6}]$, where a is a parameter and $\langle\alpha_i\rangle = 1/3$ tr α_i, and whose expression depends on the model.[57–60] For the original Applequist model,[61] both of these functions are equal to one; for Wang's model they are defined as follows:

$$\lambda_3(u) = \begin{cases} 1 & u \geq 1 \\ 4u^3 - 3u^4 & u < 1; \end{cases} \qquad \lambda_5(u) = \begin{cases} 1 & u \geq 1 \\ u^4 & u < 1. \end{cases} \tag{10}$$

Again, the s_{ij}^{dd} factors account for the exclusion or scaling of close-neighbors interactions between dipoles; for the MMPol model they are equal to zero for 1-2 and 1-3 neighbors and one for every other couple. The minimizer $\boldsymbol{\mu}$ of eq. 6 satisfies the following linear system:

$$\mathbf{T}\boldsymbol{\mu} = \mathbf{E}, \tag{11}$$

which can be solved with standard linear algebra techniques, and in particular, more efficiently, with iterative procedures. We will discuss extensively this possibility in section 3. The coupling between the MMPol polarizable force field and ddCOSMO is easily achieved by minimizing the total polarization energy $\mathcal{G}$ with respect to the dipoles[21,62,63]

$$\mathcal{G} = \frac{1}{2}\boldsymbol{\mu}^\dagger\mathbf{T}\boldsymbol{\mu} - \mathbf{E}^\dagger\boldsymbol{\mu} + \frac{1}{2}f(\varepsilon)\langle\Psi, X\rangle, \tag{12}$$

where we will have $\Psi = \Psi_0 + \Psi(\boldsymbol{\mu})$ and $X = L^{-1}g = L^{-1}(g_0 + g(\boldsymbol{\mu}))$ and where we have separated the contributions of the static charges (g_0 and Ψ_0) from the ones due to the induced dipoles ($g(\boldsymbol{\mu})$ and $\Psi(\boldsymbol{\mu})$). Through some trivial but cumbersome algebra, one gets the coupled polarization equations

$$\begin{pmatrix} \mathbf{T} & \frac{1}{2}f(\varepsilon)A & \frac{1}{2}f(\varepsilon)B \\ -B^* & L & 0 \\ -A^* & 0 & L^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ X \\ S \end{pmatrix} = \begin{pmatrix} \mathbf{E}_0 \\ g_0 \\ \Psi_0 \end{pmatrix}, \tag{13}$$

where the matrices A and B are defined as follows:

$$[(AX)_j] = \frac{4\pi}{3}[X_j]_1^m, \quad [(A^*\boldsymbol{\mu})_j]_l^m = [\Psi(\boldsymbol{\mu})]_l^m = \frac{4\pi}{3}[\mu_j]_l^m, \tag{14}$$

$$(BS)_j = \sum_{k=1}^{N_c}\sum_{lm}\sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{y}_n)U_n^k[S_k]_l^m\frac{\mathbf{v}_n^{kj}}{|\mathbf{v}_n^{kj}|^3}, \tag{15}$$

$$[(B^*\boldsymbol{\mu})_j]_l^m = [g_j(\boldsymbol{\mu})]_l^m = \sum_k \sum_n w_n Y_l^m(\mathbf{y}_n) U_n^j \frac{\mathbf{v}_n^{jk} \cdot \boldsymbol{\mu}_k}{|\mathbf{v}_n^{jk}|^3}. \tag{16}$$

In eq. 13, we have introduced the COSMO scaling function $f(\varepsilon) = (\varepsilon - 1)/\varepsilon$, which is used to account for the dielectric nature of solvents, and the ddCOSMO adjoint equation

$$L^* S = \Psi,$$

where Ψ is a linear function of the solute's density defined in such a way that the ddCOSMO solvation energy reads

$$E_s = \frac{1}{2} f(\varepsilon) \sum_{j=1}^{M} \sum_{lm} [X_j]_l^m [\Psi_j]_l^m.$$

For the MMpol model, Ψ is the sum of a part due to the point charges

$$[\Psi_{0,j}]_l^m = \sqrt{4\pi} q_j \delta_{l0} \delta_{m0}$$

and a part due to the induced dipoles, as in eq. 14. Notice that we use the real, spherical representation of the induced dipoles in equation 14. The matrix A is composed by diagonal blocks and the associated matrix-vector products are straightforward to compute; on the contrary, the matrix B is dense: the matrix-vector products in equations 15 and 16 can be seen, *mutatis mutandis*, as the computation of the electric field due to a distribution of charge or of the electrostatic potential due to a distribution of point dipoles, respectively. We will detail these aspects in section 3. As already mentioned in the introduction, we will not discuss here the coupling of the MMPol and ddCOSMO models with a QM method, as it would be far beyond the aim of this paper; furthermore, the coupling between polarizable models and QM ones has already been extensively discussed elsewhere.[21,63] Let us, however, just sketch the steps needed to couple our model to a self-consistent field QM model.

The QM core bares a density of charge which interacts with the polarizable embedding and, in turn, is polarized by the classical part. QM/Classical interaction is introduced by

adding to the right-hand sides of the coupled equations 13 the corresponding QM quantities - i.e., the QM electric field $\mathbf{E}_{QM}$ at the polarizable sites, the QM potential g_{QM}, scaled and transformed in the spherical harmonics representation as in eq. 4 and the corresponding QM Ψ vector, Ψ_{QM}, which is computed via a numerical integration procedure (see ref. 53 for a complete derivation and discussion). The aforementioned QM quantities are then added to their classical counterparts to form the right-hand side of the coupled polarization equations. Once these equations have been solved, the Classical/QM interaction is introduced through a contribution to the Fock matrix, which is the derivative of the polarization energy with respect to the density matrix. We will assume in the following that such quantities can be assembled efficiently and that they do not represent a bottleneck for the overall computation.

3 Linear scaling implementation of the classical part

In this section, we will discuss how to achieve linear scaling in the computational cost for the coupled MM/ddCOSMO problem. ddCOSMO itself already scales linearly: we will therefore focus here on the polarizable force field and on the coupling terms.

3.1 Linear scaling for the polarizable force field

The MMPol energy is the sum of a (static) charge-charge term plus a polarization term, which requires one to solve for the induced dipoles in order to be computed. In order to solve the polarization equations for the MMpol model, one needs to assemble the electric field produced by the MM charges at the polarizable sites, as in eq. 7, and to solve the linear system in equation 11. This can be conveniently done with an iterative technique: in a recent publication by some of us, two effective strategies have been explored, namely the preconditioned Conjugate Gradient method (PCG) and the combination of Jacobi Iterations with the DIIS extrapolation. Both techniques require to compute $\mathbf{T}\boldsymbol{\mu}$ matrix-vector products: such products can be viewed, *mutatis mutandis*, as the electric field produced by

the induced dipoles at the polarization sites. In this section, we will discuss how to compute both the electric field and the matrix-vector products using the FMM,[56] with a computational cost that scales linearly with respect to the size of the system. We therefore combine two efficient techniques: a fast converging iterative scheme and a fast matrix-vector multiplication method. There are two main difficulties: first, standard FMM libraries usually can only handle point charges, and not dipoles, and second, the interaction between the closest neighbors is screened and damped in the MMPol model, as can be seen from eqs. 5, 7 and 9.

The first point can be solved by using the FMM machinery itself.[56] Without entering in the technicalities [1], in a FMM computation the system is boxified and, at the finest level, a multipolar expansion is computed for each box using the charges inside the box itself. Then, the expansions at the coarser levels are obtained by translating the expansions of the child of a box to the center of the parent box. Let B denote the parent box and b its 8 children and let $(\rho_b, \alpha_b, \beta_b)$ be the spherical coordinates of the center of b with respect to the center of B: the multipolar expansion in B will be

$$[M_B]_l^m = \sum_{b=1}^{8} \sum_{\lambda=0}^{l} \sum_{\mu=-\lambda}^{\lambda} [M_b]_{l-\lambda}^{m-\mu} A_{l\lambda}^{m\mu} \rho_b^{\lambda} \overline{Y}_{\lambda}^{\mu}(\alpha_b, \beta_b), \tag{17}$$

where $A_{l\lambda}^{m\mu}$ is a numerical factor. When higher order multipoles than point charges are considered, including the induced dipoles in the MMPol model or, for more complex force fields, high order static multipoles, it is possible to set up a FMM computation by assembling the multipolar expansions in each at the finest level by translating the atomic multipoles to the center of the box using eq. 17 with a very limited computational cost.

The second point, i.e., dealing with non-Coulombic interactions between the closest neighbors, is more complex. If a scaling that depending on the bond distance only is involved, as in eqs. 5 and 7, it is possible to compute the pure Coulombic quantity and then subtract

[1] An excellent, clear and not-too-technical introduction to the FMM can be found in the notes by Beatson and Greengard "A short course on fast multipole methods" [64]

the unwanted interactions and add the scaled ones:

$$\mathbf{E}_i = \sum_{j \neq i} q_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} + \sum_{j \in N_i} (s_{ij}^{qd} - 1) q_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}, \tag{18}$$

where N_i are the 1-2, 1-3 and 1-4 neighbors of i. The first term in eq. 18 is the standard electric field and can be computed using the FMM without any modification; the second term, which only involves the 1-2, 1-3 and 1-4 neighbors, is a local correction which can be computed subsequently and the cost of which can be estimated with a small constant (for standard organic molecules, roughly 40-50) times the number of atoms, i.e., something linear in the size of the system. The same procedure can be used to compute the charge-charge energy. When a distance dependent damping factor is applied in addition to the scaling of the interactions with close neighbors, as in the $\mathbf{T}\boldsymbol{\mu}$ matrix-vector products, it is more convenient to adopt a mixed strategy. From eq. 10, one can see that there exists a limit distance d_l beyond which the damping factor is always equal to one. If one chooses a box size s for the finest level such that $d_l < s/2$, all the damped interactions will be treated in the near field, i.e., with standard double loops: it is then possible to modify the near field computation to include damping, with a very limited increase in computational cost with respect to the standard FMM procedure and no scaling differences. To summarize, we compute the $\mathbf{T}\boldsymbol{\mu}$ matrix-vector product as follows:

$$\begin{aligned}(\mathbf{T}\boldsymbol{\mu})_i = &\sum_j \left(\frac{\boldsymbol{\mu}_j}{r_{ij}^3} \lambda_3(u_{ij}) + 3 \frac{(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \mathbf{r}_{ij}}{r_{ij}^5} \lambda_5(u_{ij}) \right) + \\ &\sum_{j \in N_i} (s_{ij}^{dd} - 1) \left(\frac{\boldsymbol{\mu}_j}{r_{ij}^3} \lambda_3(u_{ij}) + 3 \frac{(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \mathbf{r}_{ij}}{r_{ij}^5} \lambda_5(u_{ij}) \right).\end{aligned} \tag{19}$$

Again, the first term in eq. 19 is the (damped) electric field produced by the induced dipoles, which can be computed with the FMM provided that all the terms for which the damping is non trivial are computed in the near field and that the library can handle dipoles; the second term is a correction which only requires one a computational effort which is linear in

the size of the system.

3.2 Linear scaling for the complete model

The complete model requires one to compute the solution to the coupled equations in eq. 13: we will here detail how this can be done with linear scaling in the computational cost. First, the right-hand side has to be assembled; again, we will not discuss the computation of the potential, field and Ψ vector due to the QM portion: further details can be found in ref. 53. The computation of the MM electric field has already been discussed; the FMM method can be applied as it is to compute the ddCOSMO right-hand side contribution due to the MM charges at the cavity, as one can compute the potential Φ with the FMM and then assemble the right-hand side g according to equation 4.

The solution of the linear system is the demanding part, and an efficient procedure needs to be set up. We have recently proposed[55] a coupled iterative procedure for the specific task, where an external iterative procedure is performed to solve for the dipoles and, at each iteration, both the direct and adjoint ddCOSMO equations are fully solved to compute the total field. In particular, once the right hand sides have been computed and a guess for the dipoles assembled, the procedure for a generic iteration k can be summarized as follows:

1. Compute the total $g^{[k]}$ and $\Psi^{[k]}$ according to equations 16 and 14;
2. Solve the ddCOSMO direct and adjoint equations $LX^{[k]} = g^{[k]}$ and $L^*S^{[k]} = \Psi^{[k]}$;
3. Compute the ddCOSMO electric field according to equations 14 and 15 and assemble the total field:
$$\mathbf{E}^{[k]} = \mathbf{E}^0 + AX^{[k]} + BS^{[k]};$$
4. Perform a Jacobi step on the dipoles:
$$\tilde{\mu_i}^{[k+1]} = \alpha_i(E_i^{[k]} - \sum_{j \neq i} T^{ij}\mu_j^{[k]});$$

5. Use DIIS to extrapolate the results and get the final $\boldsymbol{\mu}^{[k+1]}$;

6. Check convergence.

Note that there are three steps with quadratic scaling in a straightforward implementation. The first one is the computation of the right-hand side, as in eq. 16, the second the computation of the ddCOSMO field, as in eq. 15 and the third the computation of a $\mathbf{T}\boldsymbol{\mu}$ matrix-vector product. The latter step has already been treated and is an intrinsic problem related to the force-field model: it is possible to compute it with linear scaling in the computation cost as obtained in section 3.1 by adapting the FMM procedure. We will focus here on the first two points.

The dipoles contribution to the ddCOSMO right-hand side can be conveniently assembled by computing the potential produced by the dipoles at the cavity

$$[\Phi_\mu]_n^j = \sum_k \frac{\mathbf{v}_n^{jk} \cdot \boldsymbol{\mu}_k}{|\mathbf{v}_n^{jk}|^3}$$

and then weighting it and computing the numerical integral as in eq. 4. The computation of the potential can be done with a dipole-compatible FMM procedure, as illustrated in section 3.1. The ddCOSMO field, and in particular, the contribution depending on the adjoint variable S, can also be arranged in a suitable way for the use of the FMM machinery. We recall here eq. 15:

$$(BS)_j = \sum_{k=1}^{N_c} \sum_{lm} \sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{y}_n) U_n^k [S_k]_l^m \frac{\mathbf{v}_n^{kj}}{|\mathbf{v}_n^{kj}|^3}.$$

By assembling

$$\xi_n^k = \sum_{lm} w_n Y_l^m(\mathbf{y}_n) U_n^k [S_k]_l^m,$$

the quantity that we need to compute becomes

$$(BS)_j = \sum_{k=1} \sum_{n=1}^{N_g} \xi_n^k \frac{\mathbf{v}_n^{kj}}{|\mathbf{v}_n^{kj}|^3},$$

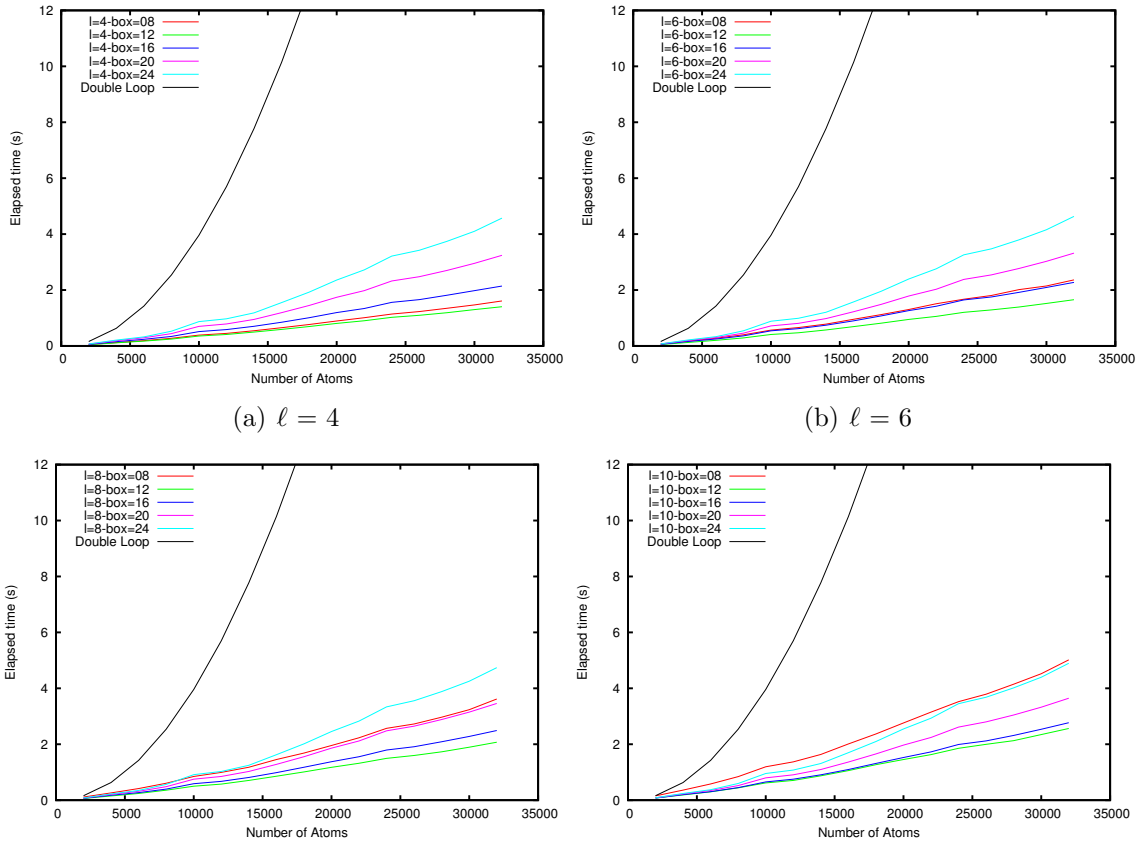
which, in turn, corresponds to the electric field produced by the "charge" distribution ξ at the nuclei and can therefore be computed with the FMM.

4 Calibration

There are two parameters regulating simultaneously the performances and accuracy of a FMM computation: the limit box-size s, i.e., the smallest allowed box size for the finest level of boxification, and the degree of the multipolar expansion ℓ. We recall that the FMM computes the electrostatic properties of the sources contained in a box at points that are either in the box or in a close (touching neighbors) one - the so-called near field - by using a standard double loop and at points in well separated boxes using the multipole machinery - the so-called far field. In this section we will explore the regime for which the FMM becomes competitive with a standard double loop computation and make some remarks about the precision. There are two different cases that need to be separately analyzed, as they require different setups: the computation of MM/MM interactions (i.e., the charge-charge, charge-dipole and dipole-dipole interaction) and the computation of MM/cavity ones (charges or dipoles potential at the cavity, ddCOSMO field at the atoms). The two cases are different as the distribution of the sources/targets varies widely between the two cases: the grid points on the cavity are much more dense, and much more abundant, than the atoms and while the double loop scales as the number of atoms times the number of points, the FMM is linear in the higher dimension. In order to provide two examples of these different behaviors, we report the results of a set of test on the dipole-dipole interaction and on the potential of the MM charges at the cavity. Starting from the structure of a large protein (photoswitchable fluorescent protein Dronpa-C62S, PDB reference: 2GX2), we cut systems of increasing size (from 2000 to 32000 atoms) for which we compute the aforementioned quantities varying both the limit box-size s and the degree of the expansion ℓ. In figure 1, we report the results for the dipole-dipole interactions, with the computational time of the FMM computations

compared to the double loop time for s varying from 8 to 24 Bohrs and ℓ varying from 4 to 10.

Figure 1: Elapsed time (s) as a function of the number of atoms for various FMM box sizes and expansion degrees for the dipole-dipole interaction.



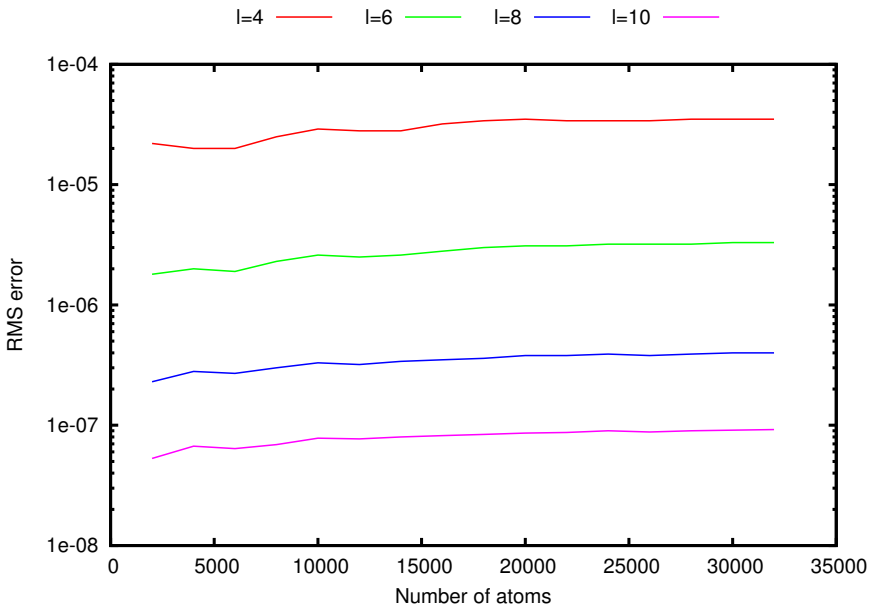
(a) $\ell = 4$ (b) $\ell = 6$

(c) $\ell = 8$ (d) ℓ =10

The size of the FMM boxes at the finest level influences both the near field and the far field computations. For larger boxes, many interactions will be accounted for in the near field regime, and therefore with a double loop, while for smaller boxes the majority of the interactions will be dealt with in the far field regime, but many more operations on the multipolar distributions will be required. This behavior is evident from figure 1, as it is possible to observe a weakly quadratic scaling for the largest box due to the large number of particles treated in the near field, while the scaling is linear for smaller boxes and the differences in cost are due to a higher or smaller number of far-field operations. From this

analysis, a box sizes of 12.0 Bohrs appears as the optimal choice for every tested value of angular momentum of the multipolar distributions. Notice that this choice is compatible with the necessity of treating all the dipole-dipole interactions where the damping functions and scaling factors are not one in the near field. The degree of the multipolar distributions is responsible, given a box size, for the accuracy of the computation with respect to the exact, double loop summation. In figure 2, we report the root-mean-square (RMS) error computed as the RMS of the difference between the FMM and double loop fields as a function of the number of atoms for $s = 12$ Bohrs. A multipolar expansion of degree as low as four already produces acceptably accurate results; a safer choice of a degree of six can be used with very good computational performances and increased accuracy.

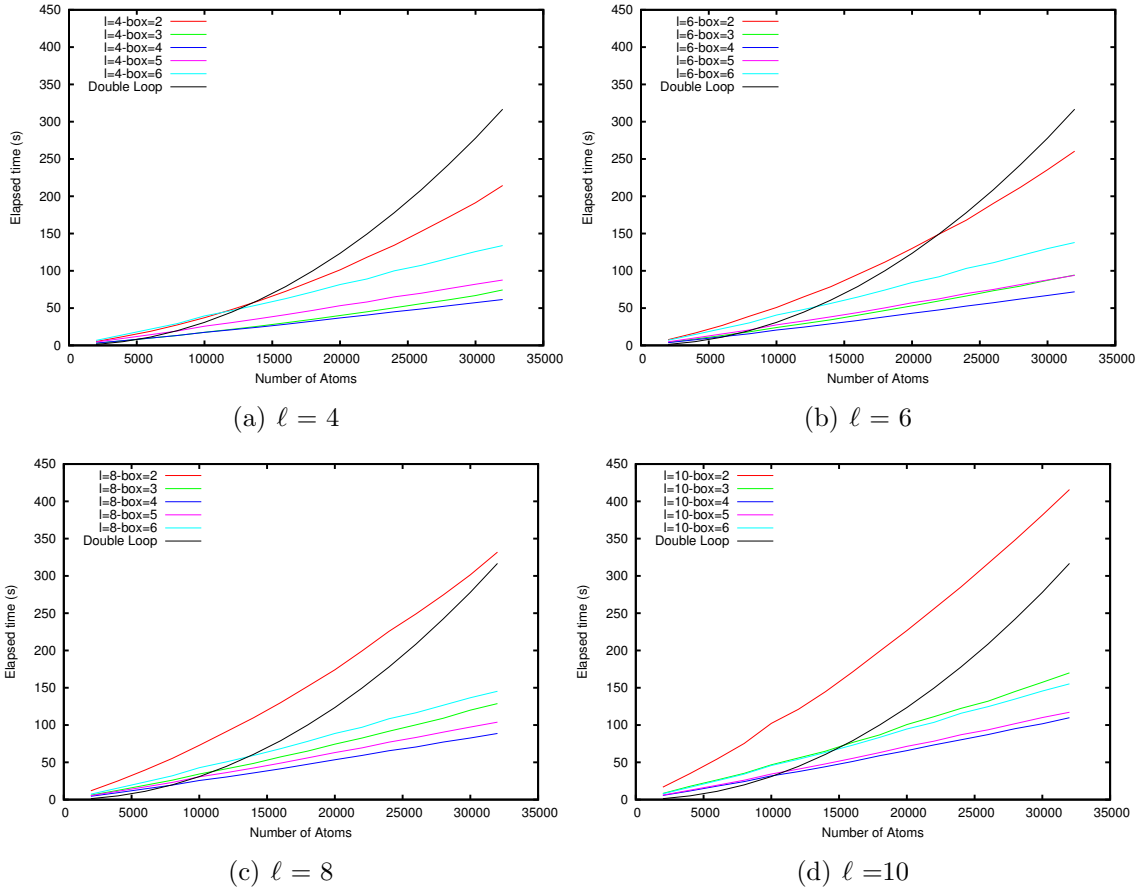
Figure 2: RMS Error on the dipoles field (i.e., the $\mathbf{T}\boldsymbol{\mu}$ product) as a function of the number of atoms for various degrees of the multipolar expansion in the FMM computation, for a given limit box-size of 12 Bohrs.



In figure 3, we report the results for the potential of the MM charges at the ddCOSMO cavity points. As already mentioned, this is a quite different configuration in terms of FMM computations, as the points on the cavity are much denser and numerous than the atoms: the use of a smaller box size will be necessary to achieve linear scaling and we can therefore

expect an increased computational cost for the FMM computation. Figure 3 shows a

Figure 3: Elapsed time as a function of the number of atoms for various FMM box sizes and expansion degrees for the MM charges potential at the ddCOSMO cavity.

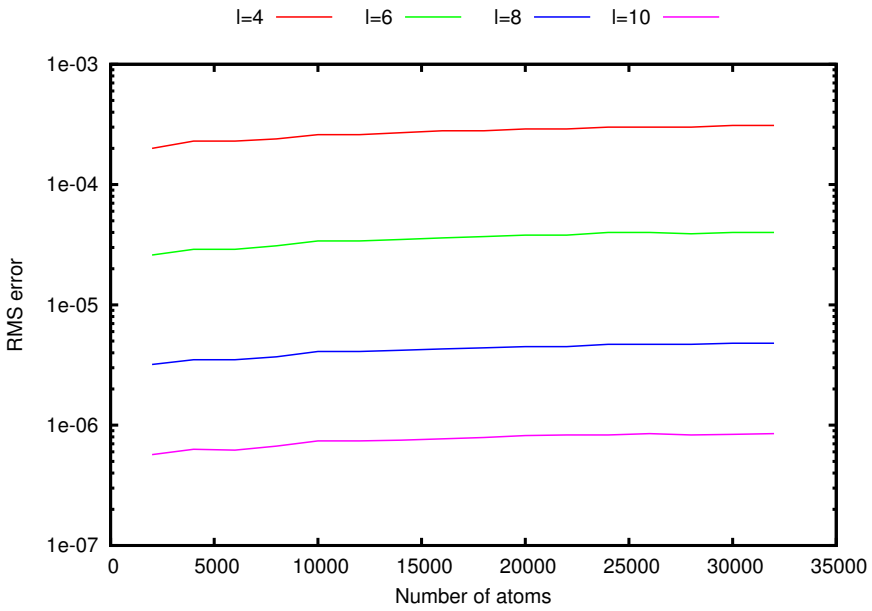


(a) $\ell = 4$ (b) $\ell = 6$

(c) $\ell = 8$ (d) $\ell =10$

very different picture from the one seen with the dipole-dipole interactions: the double loop remains advantageous even for quite large systems and is outpaced by the FMM starting from 5000 atoms for the least accurate approximation. Notice how the FMM curves clearly show that the linear scaling regime has been achieved; however, the crossing-point with the quadratic scaling computation is reached for larger systems. Nevertheless, it is possible to deduct that a box size of 4 Bohrs is optimal from the analysis reported. In figure 4, we report the RMS error on the potential as a function of the number of atoms for various degrees of expansion and for a given box size of 4 Bohrs. Again, the lowest value (4) is sufficient to have an acceptable accuracy; a higher degree can be used to increase the accuracy. In

Figure 4: RMS Error on the MM charges potential at the ddCOSMO cavity as a function of the number of atoms for various degrees of the multipolar expansion in the FMM computation for a given limit box-size of 4 Bohrs.



conclusion, the FMM procedure is very effective for computing the various MM related quantities, i.e., the charge-charge, charge-dipole and dipole-dipole interactions, while the evaluation of the MM/Continuum interaction terms becomes competitive with a standard double loop only for larger systems. This is not surprising, as the latter is a much more difficult case for the FMM; nevertheless the use of the FMM machinery allows for a linear scaling in computational cost with respect to the size of the system, making therefore large and very large systems accessible.

5 Numerical results

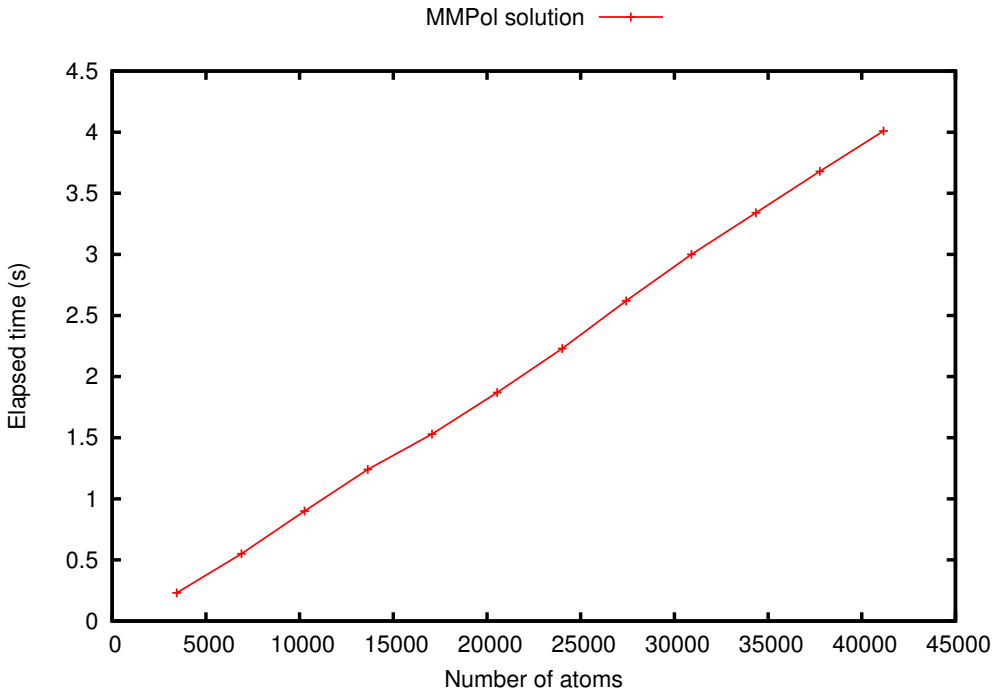
The linear scaling procedure described in this article has been implemented in a local development version of the Gaussian suite of programs,[65] where both ddCOSMO and the MMPol model had already been implemented. All the computation described here were performed on a Dual Xeon X5690 (12 cores, 3.47GHz) cluster node equipped with 24GB of 1,33GHz DDR3 RAM. For the FMM computations, we use the following parameters,

which we determined in order to optimize the computational time obtaining while retaining a good accuracy: $\ell = 8, s = 12$Bohrs for the MM/MM interactions and $\ell = 6, s = 4$Bohrs for the MM/cavity ones. As a model system, we considered the same protein used for the calibration (PDB reference 2G2X). Such a protein is composed by three units, each in turn composed by four chains. We extracted from such a structure one monomer and proceeded by adding one more monomer until the full protein was obtained, in order to get 12 test systems of increasing size. We remark that such systems can be considered as "real life" examples, as they correspond to a real biological molecule which grows in a globular way - it is therefore not a system tailored for the FMM to perform optimally. The size of the test systems varied from 46 QM atoms and 3440 MM atoms for the smallest system up to 552 QM atoms and 41157 MM atoms for the largest. For the QM portion, we used a semiempirical Hamiltonian (in particular, the ZINDO Hamiltonian[66]); while we were not interested in an accurate description of the system, as the main purpose of this paper is to test the scaling and performances of our code, we point out that a DFT computation would have been also feasible on a standard cluster node.

Table 1: Further details for the computations in figure 5. We denote by M the number of MM atoms, by Time the elapsed time in seconds, by Levels the number of hierarchical subdivisions of the space for the FMM computation and by N_{it} the number of iterations required to converge the polarization equations.

M	Time	Levels	N_{it}
3440	0.23	4	11
6895	0.55	5	11
10261	0.90	5	11
13636	1.24	5	11
17067	1.53	5	11
20543	1.87	5	11
24016	2.23	5	11
27426	2.62	5	11
30908	3.00	5	11
34353	3.34	5	11
37765	3.68	5	11
41157	4.01	5	11

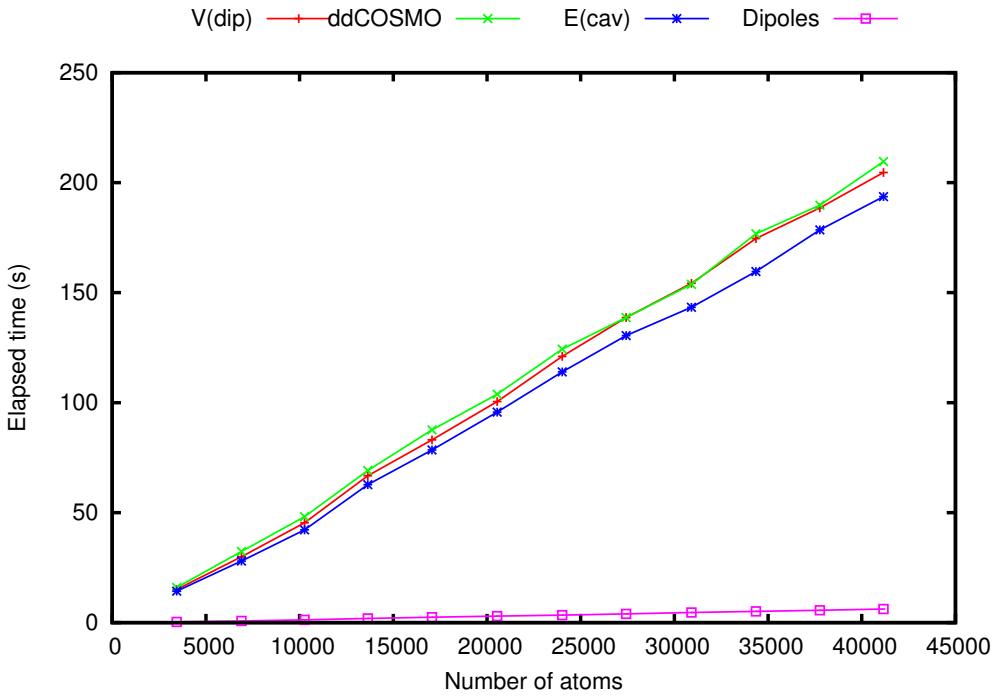
Figure 5: Elapsed time (in seconds) as a function of the system size for the solution of the MMPol equations using the Jacobi/DIIS solver in conjunction with the FMM.



We report in figure 5 the total elapsed time needed to solve the MMPol equations using our Jacobi/DIIS iterative solver in conjunction with the FMM to compute the matrix-vector products, as described in section 3. Convergence was achieved when the Root Mean Square of the dipoles absolute increment was lower than 10^{-7} atomic units and its maximum value was lower than 10^{-6}. Some additional details are reported in table 1 for the same computations. There are two main considerations that can be done: first, the scaling is, as expected, perfectly linear. Notice that, as reported in table 1, the same number of iterations was needed for all systems. Second, the overall computational overhead introduced by the MMPol method is, in our implementation, negligible: for each SCF step, the computational time is increased of only a few seconds. More specifically during, the first SCF cycle, where the MM charges potential and field are also computed, 10 seconds were spent to deal with the MM part, including assembling the QM field and potential and the Fock matrix contributions. If one considers that also the MMPol memory requirements scale linearly with the size of the system, it is possible to conclude that no practical size limitation is introduced by the

MMPol model.

Figure 6: Elapsed time (in seconds) as a function of the system size for the various tasks needed for the solution of the coupled MMPol/ddCOSMO equations using the coupled solver in conjunction with the FMM.

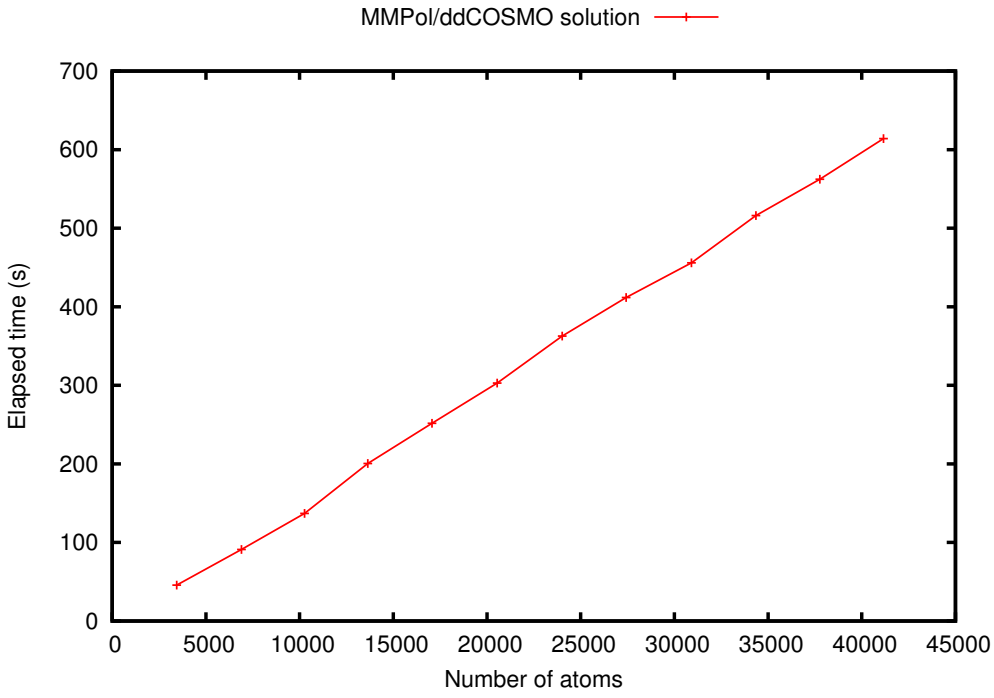


In figure 6, we report the elapsed time for the solution of the coupled MMPol/ddCOSMO equations decomposed in its principal components: computing the dipoles potential at the cavity (V(dip)), solving the ddCOSMO direct and adjoint equations (ddCOSMO), computing the ddCOSMO field at the dipoles (E(cav)) and performing a JI/DIIS step (Dipoles). Notice that each of these actions is performed once per iteration in the iterative procedure to solve eq. 13: it is then apparent that a coupled MMPol/ddCOSMO computation is a much heavier task than a MMPol only one, as for each iteration two full ddCOSMO solutions and two expensive MM/cavity interactions are to be evaluated. Further details are reported in table 2. However, each task exhibits a perfectly linear scaling with respect to the size of the system, as well as the overall computational time, also reported in figure 7. Notice that the small step observed between the third and fourth point in the graph is due to the increased number of iterations (11 to 12) needed to achieve convergence. The computational

Table 2: Further details for the computations in figure 6. We denote by M the number of MM atoms, by Timings the elapsed times in seconds, by Lev(MM) the number of hierarchical subdivisions of the space for the MM/MM FMM computations and by Lev(Cav) the counterpart for the MM/Cavity computations and by N_it the number of iterations required to converge the Polarization equations. V(dip): potential of the dipoles at the cavity, dd: solution of the ddCOSMO direct and adjoint equations, E(Cav): ddCOSMO field at the dipoles, Dipoles: Jacobi/DIIS steps, Total: total time for the solution of the coupled equations.

M	Data			Timings				
	Lev(MM)	Lev(Cav)	N_{it}	V(Dip)	dd	E(Cav)	Dipoles	Total
3440	4	6	11	15.07	15.97	14.31	0.34	45.69
6895	5	6	11	29.96	32.36	27.97	0.77	91.06
10261	5	6	11	45.40	48.16	42.16	1.28	137.00
13636	5	6	12	66.77	69.16	62.75	1.89	200.57
17067	5	7	12	83.11	87.64	78.49	2.43	251.67
20543	5	7	12	100.49	103.80	95.67	2.98	302.94
24016	5	7	12	121.03	124.30	113.97	3.42	362.72
27426	5	7	12	138.66	138.69	130.47	3.99	411.81
30908	5	7	12	154.28	153.72	143.35	4.63	455.98
34353	5	7	12	174.58	176.76	159.59	5.10	516.03
37765	5	7	12	188.54	189.72	178.51	5.61	562.38
41157	5	7	12	204.62	209.59	193.64	6.20	614.05

Figure 7: Total elapsed time (in seconds) for the solution of the coupled MMPol/ddCOSMO equations as a function of the system size.



effort required by a MMPol/ddCOSMO computation is, as expected, much larger than for a MMPol computation: up to roughly ten minutes per SCF step are required to solve the coupled equations for the largest system. Nevertheless, the overhead is not overwhelming and computations as large as the one here reported remain feasible on a standard cluster node, again, also thanks to the linear scaling of the memory requirements. It is interesting to notice how the MM part represents a negligible portion of the total MMPol/ddCOSMO cost: this is expected as the MM/cavity terms involve a much larger number of points at which one needs to evaluate the electrostatic properties and, because of the much smaller box size needed, requires a much deeper level of boxification with up to seven hierarchical subdivisions of the space, to be compared with up to 5 for the MM only terms. However, the major improvement introduced by our implementation is the use of ddCOSMO to solve the continuum solvation equations: as reported in two recent papers, ddCOSMO is as much as two to three order of magnitude faster than other existing implementations. Even by using a state-of-art, FMM based traditional implementation of COSMO,[67] the computational cost

of the coupled MMPol/continuum computation would be largely dominated by the cost of solving the COSMO equations which, for the largest system here reported, would take several hours.[52]

6 Conclusions

In this paper we have presented a new implementation of the MMPol model based on the Fast Multipole Method which achieves linear scaling in computational cost. Furthermore, we have coupled this new implementation with the recently developed ddCOSMO algorithm for continuum solvation and used the FMM to compute the MM/Continuum interaction terms in order to obtain a globally linear-scaling polarizable QM/MM/Continuum code. The FMM parameters have been calibrated in order to get an optimal compromise between efficiency and accuracy and the linear scaling regions have been explored and compared with the standard double-loop implementations. For the MMPol code, the FMM code is faster than the double-loop one even when medium-large systems (roughly 1000 atoms) are considered; moreover, the code is in absolute very fast, adding a very limited computational overhead to a QM/MM computation even for very large systems. Coupled MMPol/ddCOSMO are much more computationally intensive: the FMM code for the MM/cavity interaction terms becomes faster than the standard, double-loop code only for considerably large systems, roughly beyond 6000-8000 atoms, and the absolute timings are much larger than the ones implied in a MMPol only computation; this is amply justified by the much larger complexity of the numerical problem and by the difficult FMM setup. The key ingredient of the coupled code is, however, the ddCOSMO algorithm itself, which makes very large computations feasible in reasonable time. While expensive, a three-layered, fully polarizable QM/MMPol/Continuum computation remains feasible and, for large system and accurate levels of theory for the QM core, we can expect that the classical portion of the computation will have a limited impact on the overall computation even for very large MM regions: for the largest system tested in

this paper, which was composed of more than 40000 MM atoms, roughly 10 minutes were needed to solve the coupled MM/ddCOSMO equations at the first SCF cycle. We point out that, by using the dipoles and ddCOSMO quantities computed at a previous SCF cycle as a guess, the number of iterations needed to achieve convergence is strongly reduced: along the SCF, the cost of the classical treatment will therefore become smaller and smaller. Nevertheless, the coupled MMPol/ddCOSMO computation remains in absolute expensive, and further investigation is needed in order to improve its computational efficiency. One possible way to reduce the cost of both ddCOSMO and the MMPol/ddCOSMO coupling would be to introduce some coarse-graining in the cavity, for instance by including the hydrogen atoms in the sphere of the heavy atom they are bounded to: this would simplify the cavity topology and eliminate a large quantity of spheres without affecting much the precision of the computation.

To the best of our knowledge, this is the first general, linear scaling implementation of a fully polarizable MM/Continuum embedding scheme to be used in a multiscale approach: our code, by making large and very large QM/MM(/Continuum) simulations accessible even with standard computers, extends the range of application of such approaches to large biological systems and opens the door to future applications that were not possible with existing implementations.

Acknowledgments

The authors are thankful to Dr. Giovanni Scalmani and Dr. Michael J. Frisch for many useful discussions. We would also like to thank Prof. Benedetta Mennucci for carefully reading the paper. This work was supported in part by the ANR Manif and the French state funds managed by CALSIMLAB and the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02. F.L., S.C. and S.J. gratefully acknowledge Gaussian, Inc. for funding. F.L. also acknowledges financial support from the Alexander

von Humboldt Foundation.

References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227 – 249.

(2) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.

(3) Gao, J.; Xia, X. *Science* **1992**, *258*, 631–635.

(4) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(5) Lin, H.; Truhlar, D. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(6) Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(7) Barone, V.; Biczysko, M.; Brancato, G. In *Combining Quantum Mechanics and Molecular Mechanics. Some Recent Progresses in QM/MM Methods*; Sabin, J. R., Canuto, S., Eds.; Advances in Quantum Chemistry; Academic Press, 2010; Vol. 59; pp 17–57.

(8) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

(9) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.

(10) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4226.

(11) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(12) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

(13) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(14) Mennucci, B. *J. Phys. Chem. Lett.* **2010**, *1*, 1666–1674, and references therein.

(15) Mennucci, B. *WIREs Comput. Mol. Sci.* **2012**, *2*, 386–404.

(16) Klamt, A. *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709.

(17) Mennucci, B., Cammi, R., Eds. *Continuum Solvation Models in Chemical Physics*; Wiley, New York, 2007.

(18) Pedone, A.; Biczysko, M.; Barone, V. *ChemPhysChem* **2010**, *11*, 1812–1832.

(19) Rega, N.; Cossi, M.; Barone, V. *J. Am. Chem. Soc.* **1998**, *120*, 5723–5732.

(20) Brancato, G.; Rega, N.; Barone, V. *J. Chem. Phys* **2006**, *124*, 214505.

(21) Lipparini, F.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 3711–3724.

(22) Vreven, T.; Mennucci, B.; da Silva, C.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62–72.

(23) Cui, Q. *J. Chem. Phys.* **2002**, *117*, 4720.

(24) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 238–249.

(25) Bandyopadhyay, P.; Gordon, M. S. *J. Chem. Phys.* **2000**, *113*, 1104.

(26) Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023.

(27) Li, H. *J. Chem. Phys.* **2009**, *131*, 184103.

(28) Li, H.; Pomelli, C. S.; Jensen, J. H. *Theor. Chem. Acc.* **2003**, *109*, 71–84.

(29) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(30) List, N. H.; Olsen, J. M. H.; Jensen, H. J. A.; Steindal, A. H.; Kongsted, J. *J. Phys. Chem. Lett.* **2012**, *3*, 3513–3521.

(31) Curutchet, C.; Scholes, G. D.; Mennucci, B.; Cammi, R. *J. Phys. Chem. B* **2007**, *111*, 13253–65.

(32) Scholes, G. D.; Curutchet, C.; Mennucci, B.; Cammi, R.; Tomasi, J. *J. Phys. Chem. B* **2007**, *111*, 6978–82.

(33) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 4527–4538.

(34) Hsu, C.-P.; Fleming, G. R.; Head-Gordon, M.; Head-Gordon, T. *J. Chem. Phys.* **2001**, *114*, 3065–3072.

(35) Lipparini, F.; Cappelli, C.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.

(36) Lipparini, F.; Cappelli, C.; Barone, V. *J. Chem. Phys.* **2013**, *138*, 234108.

(37) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(38) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(39) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

(40) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. *J. Chem. Phys.* **2007**, *126*, 154112.

(41) Steindal, A. H.; Olsen, J. M. H.; Ruud, K.; Frediani, L.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5440–5451.

(42) Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.

(43) Lipparini, F.; Cappelli, C.; Scalmani, G.; De Mitri, N.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 4270–4278.

(44) Lipparini, F.; Egidi, F.; Cappelli, C.; Barone, V. *J. Chem. Theory Comput.* **2013**, *9*, 1880–1884.

(45) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2014**, *10*, 1795–1809.

(46) Beerepoot, M. T.; Steindal, A. H.; Ruud, K.; Olsen, J. M. H.; Kongsted, J. *Comp. Theor. Chem.* **2014**, *1040-1041*, 304–311.

(47) Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027–3037.

(48) Caprasecca, S.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2012**, *8*, 4462–4473.

(49) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2014**, *10*, 1638–1651.

(50) Cancès, E.; Maday, Y.; Stamm, B. *J. Chem. Phys.* **2013**, *139*, 054111.

(51) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. *J. Chem. Theory Comput.* **2013**, *9*, 3637–3648.

(52) Lipparini, F.; Lagardère, L.; Scalmani, G.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. *J. Phys. Chem. Lett.* **2014**, *5*, 953–958.

(53) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. *J. Chem. Phys.* **2014**, *141*, 184108.

(54) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.

(55) Lipparini, F.; Lagardère, L.; Raynaud, C.; Stamm, B.; Cancès, E.; Mennucci, B.; Schnieders, M.; Ren, P.; ; Maday, Y.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2015**, Just Accepted, DOI: 10.1021/ct500998q.

(56) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325–348.

(57) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. *J. Phys. Chem. B* **2011**, *115*, 3100–3111.

(58) Sala, J.; Guàrdia, E.; Masia, M. *J. Chem. Phys.* **2010**, *133*, 234101.

(59) Thole, B. *Chem. Phys.* **1981**, *59*, 341–350.

(60) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(61) Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.

(62) Lipparini, F.; Scalmani, G.; Mennucci, B.; Cancès, E.; Caricato, M.; Frisch, M. J. *J. Chem. Phys.* **2010**, *133*, 014106.

(63) Lipparini, F.; Scalmani, G.; Mennucci, B.; Frisch, M. J. *J. Chem. Theory Comput.* **2011**, *7*, 610–617.

(64) Beatson, R.; Greengard, L. A short course on fast multipole methods. Wavelets, Multilevel Methods and Elliptic PDEs. 1997; pp 1–37.

(65) 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.; Jr.,; 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.; Fores-

man, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian Development Version, Revision H.36. Gaussian Inc. Wallingford CT 2010.

(66) Zerner, M. C. *Rev. Comput. Chem.*; John Wiley & Sons, Inc., 2007; pp 313–365.

(67) Scalmani, G.; Barone, V.; Kudin, K.; Pomelli, C.; Scuseria, G.; Frisch, M. *Theor. Chem. Acc.* **2004**, *111*, 90–100.

Figure 8: Graphical TOC