Contribution to the Development of Advanced Approaches for Electron and Molecular Dynamics Simulations in Extended Biomolecules

Xiaojing Wu

NNT : 2018SACLS252

Contribution to the Development of Advanced Approaches for Electron and Molecular Dynamics Simulations in Extended Biomolecules

Thèse de doctorat de l'Université Paris-Saclay
préparée à Université Paris-Sud

École doctorale n°571 : sciences chimiques : molécules, matériaux, instrumentation et biosystèmes (2MIB)
Spécialité de doctorat: Chimie

Thèse présentée et soutenue à Orsay, le 11 septembre 2018, par

Mme Xiaojing WU

Composition du Jury :

M. Dennis R. SALAHUB Professeur, Université de Calgary (Canada)	Président
Mme Tzonka MINEVA Directrice de Recherche, CNRS-Institut Charles Gerhardt (UMR 5253)	Rapportrice
Mme Isabelle NAVIZET Professeur, CNRS-Université Paris-Est (UMR 8208)	Rapportrice
M. Gilles FRISON Chargé de Recherche, CNRS-Ecole Polytechnique (UMR 9168)	Examinateur
Mme Valerie BRENNER Directrice de Recherche, CEA (UMR 9222)	Examinatrice
M. Yi ZHAO Professeur, Université de Xiamen (China)	Examinateur
M. Daniel BORGIS Directeur de Recherche, CNRS-Ecole Normale Supérieure (UMR 8640)	Examinateur
M. Aurélien de la LANDE Chargé de Recherche, CNRS-Université Paris-Sud (UMR 8000)	Directeur de thèse

致我最亲爱的家

“On ne voit bien qu’avec le coeur.

L’essentiel est invisible pour les yeux.”

Antoine de Saint-Exupéry

“Dans la vie, rien n’est à craindre,

tout est à compredre.”

Marie Skłodowska-Curie

Acknowledgments

I am not very good at expressing myself. Words of gratitude are too thin to express my sincere thanks for your kindness help and support. Your name carved in my heart. I will never forget. But I will use my action to express.

Top name list in Xiaojing's heart:

Aurélien de La Lande, Carine Clavaguéra, Fabien Cailliez,

Jean-Marie Teuler, Dennis R. Salahub, Rodolphe Vuilleumier, Laura Baciou, David Lauvergnat, Isabelle Demachy, Federica Agostini, Ridard Jacqueline, Bernard Lévy, Pascal Pernot, Karim Hasnaoui, Louis Lagardère, Jean-Philip Piquemal, Bernard Rousseau, Michèle Desouter-Lecomte, Tzonka Mineva, Isabelle Navizet, Daniel Borgis, Valerie Brenner, Gilles Frison.

Aurelio Alvarez-Ibarra, Florent Amiot, Angela Parise, Rémi Fauve, Pierre Lehujeur, Danielle Molina, Cecia Cioletta,

吴文义, 吴淑芳, 吴礼宣, 陈乌尼, 吴有理, 张玉美 ,吴博雅, 吴鸿博,
邱金红, 原世豪, 张露, 王芙蓉, 张向阳, 赵仪,梁万珍, 邹海峰, 梁畅

大恩不言谢

Résumé

Les transferts d'électrons sont des processus physico-chimiques fondamentaux qui ont lieu au cœur des biosystèmes (photosynthèse, respiration cellulaire, catalyse enzymatique, mécanismes de protections et de réparations de dommages photo-induits ou radio-induits). Comprendre les mécanismes par lesquels les systèmes naturels parviennent à générer des transferts efficaces dans les protéines permettrait le développement de catalyseurs biomimétiques. Ces processus impliquent des déplacements d'électrons sur des échelles de temps très courtes (de 10^{-5} au 10^{-18} s). La simulation numérique est un moyen puissant d'étudier ces mécanismes au niveau microscopique. Mon travail de doctorat vise à augmenter la précision des méthodes de simulation moléculaire pour décrire les transferts d'électrons dans les systèmes biologiques. Il s'articule autour de deux projets.

Dans la premièr projet nous cherchons à comprendre les transferts d'électrons dans les proteines qui sont dominés par le mouvement nucléaire. De nombreux travaux expérimentaux et théoriques ont cherché à élucider les propriétés d'oxydo-réduction associées à ces transferts d'électrons. La simulation en biologie est un domaine en plein développement. Mais il reste encore beaucoup de progrès à faire notamment pour améliorer les précisions afin de prédire les mécanismes enzymatiques. La complexité des protéines nécessite des méthodologies de calcul avancées. La simulation existante est insuffisamment précise par rapport aux résultats des expériences. En conséquence, ce travail vise à mettre en place une nouvelle approche qui améliore significativement la précision pour simuler des propriétés redox des protéines, et plus particulièrement des hémoprotéines.

Une étape importante a été de construire de champ de force reposant sur une description multipolaire des interactions électrostatiques (AMOEBA) pour estimer les potentiels redox des hémoprotéines. Nous avons dérivé des paramètres pour AMOEBA afin de décrire précisément les interactions électrostatiques avec l'hème. Une amélioration très encourageante est obtenue par rapport aux champs de forces standards. Maintenant, ils ont utilisés pour le calcul des potentiels d'oxydo-réduction sur une série d'hémoprotéines pour lesquelles des données expérimentales sont disponibles. En raison de l'omniprésence des protéines contenant un hème en biologie, ces travaux ouvriraient alors vers de très larges applications. Par exemple, simuler les mécanismes de transport d'oxygène et réactions enzymatiques au sein du métabolisme.

Le second projet de cette thèse consiste à explorer le domaine fascinant de la dynamique des électrons à l'échelle de l'attoseconde (10^{-18} s) dans des molécules complexes. Du côté numérique, beaucoup d'efforts ont été consacrés à la conception d'algorithmes de simulation. Mais la plupart des implémentations ont été conçues pour étudier des systèmes moléculaires en phase gazeuse ou dans un environnement homogène. J'ai développé des nouvelles méthodes pour étudier la dynamique des électrons dans des biomolécules à l'échelle attoseconde en incluant les effets d'environnement hétèrogène. Nous avons conçu un couplage original entre la théorie de la fonctionnelle de la en temps réel densité dépendant du temps (RT-TDDFT) et un modèle de mécanique moléculaire polarisable (MMpol). Une implémentation efficace et robuste de cette méthode a été réalisée dans le logiciel deMon2k. L'utilisation de techniques d'ajustements de densités électroniques auxiliaires permet de réduire drastiquement le coût de calcul des propagations RT-TDDFT/MMpol. La méthode est appliquée à l'analyse de la dissipation d'énergie dans l'environement d'un peptide excité par une impulsion laser. Les simulations ont montré que cette méthode est suffisamment efficace pour envisager de simuler la dynamique des électrons dans de grands systèmes moléculaires. De plus, ces développements ont permis d'ouvrir une nouvelle ligne de recherche sur les effets des rayonnements ionisants sur les biomolécules.

Contents

GENERAL INTRODUCTION

Electron transfers (ET) are basic chemical processes, with prominent importance in chemistry, physics, biology, life science, materials and microelectronics disciplines. In biology, ET are involved in numerous processes like light harvesting, cellular respiratory chains, enzymatic reactions or defense against oxidative stress. Unravealing the mechanisms by which natural systems manage to control the directionality, speed and efficiency of ET within proteins is commonly expected to feed reflection to design innovative catalytic processes for industrial applications, or to inspire innovative medicinal projects, for example, for the development of selective enzymatic inhibitors or of new antibiotics. In this context, numerical approaches offer powerful means to understand ET mechanisms at the microscopic level. Simulations can even provide mechanistic insights that are not accessible by experiments.

A fascinating characteristic of biological ET is the temporal scales associated with these phenomena. They cover several orders of magnitude, ranging from a few microseconds in long-range ET down to tens of attoseconds in charge migration. For theoretical physical-chemists, understanding how such a wide range of rates is possible requires advanced computational methodologies, both for molecular dynamics (MD) and for electron dynamics (ED) simulations. The objective of this PhD thesis is to contribute to the development of advanced approaches for simulating these phenomena. This manuscript contains two distinct, though connected parts.

In Part I, we are interested in redox potentials of heme cofactors when inserted within proteins. Redox potentials are a central quantity of redox theories as they are related to the free energies of redox reactions, hence their thermodynamical feasibilities, and also to the reaction rates, for example in the Marcus Theory. The accuracy of numerical simulations to predict redox potentials of cofactors encapsulated in protein matrices is far from reaching the experimental ones. My PhD project aims at improving the reliability of computational approaches dedicated to the simulation of redox potentials.

The computational machineries for the evaluation of redox potentials based on the linear response approximation are now well established. They often involve combinations of quantum mechanical calculations and classical molecular dynamics simulations. The computed redox potentials depend not only on the structure of the heme cofactors but also on the interactions between hemes and their environments (protein, solvent, counter-ions). The accuracy of simulations to evaluate the latter is highly dependent on the functional form of mechanical Force Fields (FF). It has been shown by several research groups that electrostatic induction (polarization) is essential and must be taken into account. On the other hand

electrostatic interactions among permanent charge densities are also important. Our objective is to make significant progress in the evaluation of redox potential by using advanced FF. In Chapter 1 we will describe the computational methodology for redox potentials evaluation in proteins based on the Linear Response Approximation. In Chapter 2 we will report an important step toward equivalent simulations with the sophisticated AMOEBA force field, namely the development of dedicated parameters describing multipolar electrostatic interactions of heme in ferric and ferrous forms.

In Part II, we are interested in ultrafast electron transfers taking place at the sub-femtosecond time scales. These ET are driven by electronic correlation. They take place when tunneling is involved between donor and acceptor, or more generally when a molecule is subjected to a perturbation (electric fields, components of electromagnetic waves or collisions with charged particles).

Real-Time Time-Dependent Density Functional Theory (RT-TDDFT) is a powerful method to simulate ED with excellent computational cost/accuracy ratio. With the development of efficient algorithms, a current challenge is to simulate ED in complex biological systems, where environment effects are possibly important. The developments described in Part II are intended to developing an original methodology coupling RT-TDDFT and polarizable force fields. Toward this end we report in Chapter 3 a new implementation of RT-TDDFT in the software deMon2k. The implementation relies on variationally fitted electron densities to improve computational performance. In Chapter 4 the coupling scheme between the two methods is justified and we report analyses of energy dissipation from the out-of-equilibrium electron density of a peptide toward its polarizable environment.

PART I

Toward a More Accurate Evaluation of Redox Potentials of Hemoproteins

Toward a More Accurate Evaluation of Redox Potentials of Hemoproteins

Redox potentials are reference values to characterize the propensity of molecules to accept or release electrons in a medium. In the Marcus theory of electron transfer, the redox potentials enter into the determination of the rates via the so-called driving force ($-\Delta G°$). In biology common redox cofactors found in oxidoreductases are either of organic nature (flavins, nicotinamides or quinones) or inorganic complexes (iron, nickel, copper, vanadium or manganese). In the latter case we talk about metalloproteins. More than one third of all currently known proteins involve redox reactions which serve as significant catalysts for numerous biological processes. Figure 1 illustrates the remarkable diversity of redox potentials of these cofactors when they are embedded inside proteins.

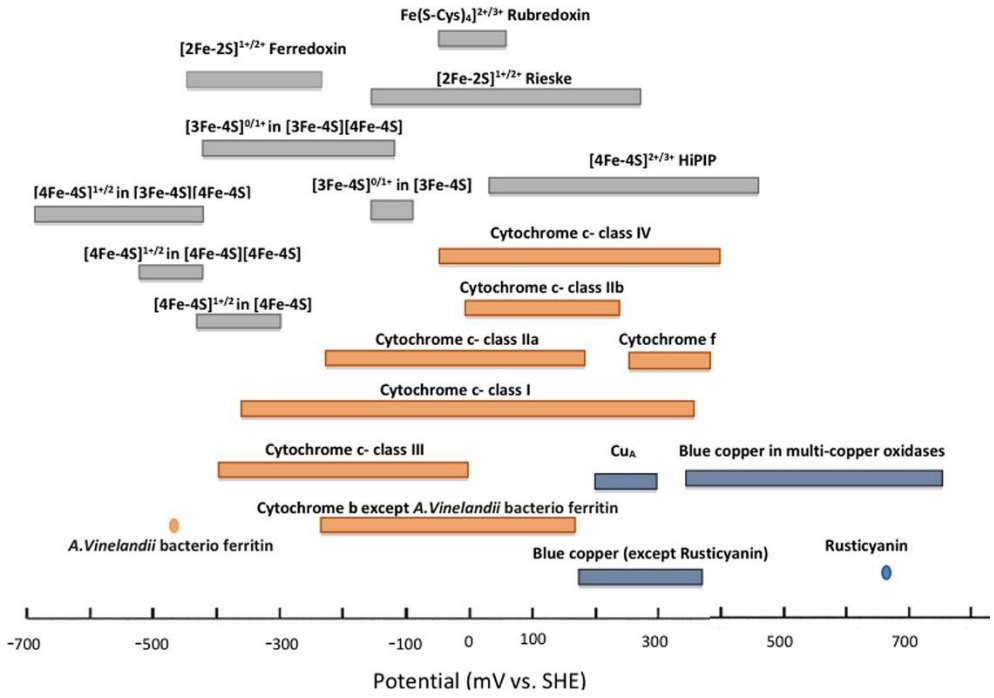


Figure 1. Reduction potential range of redox centers in electron transfer processes (Figure 1 in ref 1)

We are interested in heme proteins. They are abundant in biomolecules, and exist in the ferrous Fe^{II} or ferric Fe^{III} oxidation states. They participate in several biological functions: electron-transfer reactions, oxygen transport and storage, oxygen reduction to hydrogen peroxide or water, oxygenation of organic substrates, and the reduction of peroxides. This versatility in functions is made possible by a combination of differences in both the heme cofactor and the protein matrix of the various hemoproteins.

heme cofactor consists of a prosthetic group and different ligands. The prosthetic group contains an iron cation chelated by the four nitrogen atoms of a porphyrin ligand. According to the different structures of porphyrin rings, prosthetic groups are classified as different types of heme shown in Figure 2.

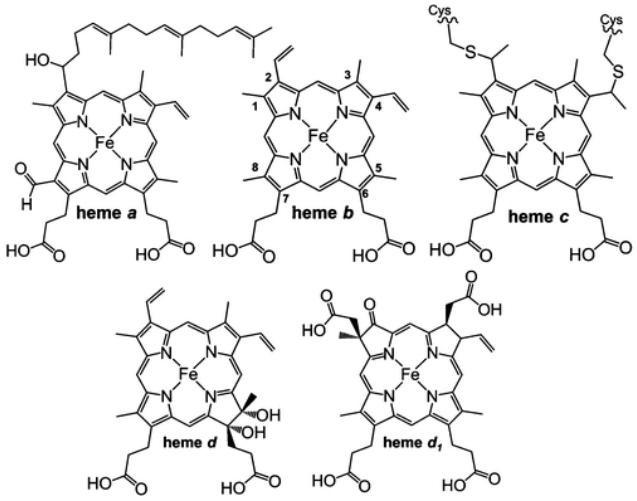


Figure 2. Different type of heme found in cytochromes. (Figure 3 from ref 1)

The iron coordination sphere also incorporates other ligands in axial positions, i.e. below and above the porphyrin ring. In Figure 3, some common axial ligands found in cytochromes are shown. Histidine or methionine side chains commonly act as axial ligands but other types of ligands can be found in some enzymes (the structure C on the Figure 3, for instance).

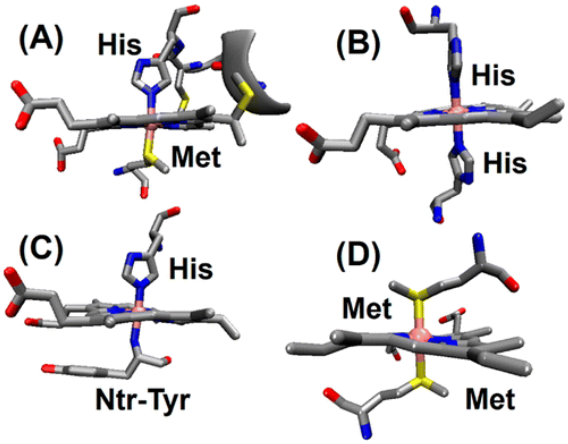


Figure 3. Commonly found heme axial ligands in various cytochromes. (A) Class I cyts c. (B) Cyts b. (C) cyt f. (D) c-type cytochromes. (Figure 4 from ref 1.)

Theoretical chemistry presents an alternative to experimental methods to evaluate redox potentials of hemes incorporated into protein matrices. The Marcus theory of electron transfer initially developed for electron transfer in solutions and at electrodes was found by various groups to be adaptable for protein redox properties. Research is still on-going worldwide to identify situations where the underlying hypotheses of the theory fail. This includes for example ultrafast processes for which the characteristic time scales of the redox process are much shorter than the relaxation times of protein matrices. This also includes processes involving large conformational rearrangements of the solvation states of the cofactors, or strong polarization of the redox cofactors. Another road to be explored is the improvement of computational approaches for more accurate predictions of calculated redox potentials and electron transfer rates. This is the guiding line of Part I of this PhD thesis.

Toward this aim, a series of small heme proteins with known experimental redox potentials are considered. The theory and the computational set up are described in Chapter 1 with the simulation results of non-polarizable force fields. The second Chapter is to achieve MD simulations with an advanced force field, namely AMOEBA. This force field describes electrostatic interactions among molecules beyond the monopolar approximation, and for that reason we expect significant improvement of computed redox potentials. A first essential step is to derive force fields parameters. This is the topic of Chapter 2.

1. Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y., Metalloproteins Containing Cytochrome, Iron–Sulfur, or Copper Redox Centers. *Chem. Rev.* **2014,** *114* (8), 4366.

CHAPTER 1

Numerical Simulations of Redox Potentials under the Linear Response Approximation

Numerical Simulations of Redox Potentials under the Linear Response Approximation

The modern description of the Marcus theory from first principles considerations emerged in the 1980's from the works of Warshel[1], Tachiya[2] and Hynes[3], to name but a few key researchers of this research field. These authors developed the conceptual framework to relate the phenomenological theory developed by Marcus in the 1950's to microscopic description of matter. Importantly they also introduced algorithms for testing the theory by numerical simulations. These algorithms include classical continuum electrostatics[4], Multi-Conformation Continuum Electrostatics (MCCE)[5], Protein-Dipole Langevin-Dipoles (PDLD)[6], semi-microscopic PDLD analysis[7], free energy perturbation and microscopic Linear Response Approximation (LRA)[8], Molecular Dynamics (MD)[9], QM and QM/MM (Quantum Mechanics/Molecular Mechanics) methods[10].

As explained in the Introduction the objective of this PhD thesis is to contribute to the improvement of redox potentials of proteins, and more particularly of heme proteins. Toward this end our approach will be to use advanced force fields (FF). This will be the object of Chapters 2.

Heme proteins, at least the simplest ones, can be understood with concepts from the Marcus Theory, which is equivalent for microscopic simulation in relying on the Linear Response Approximation. In this Chapter we therefore introduce the conceptual framework of the LRA, showing how to estimate redox free energies and the important reorganization energy. In the second section of the Chapter we illustrate how one uses this formalism taking a series of six small heme proteins as example with simple force fields.

I COMPUTATIONAL APPROACHES FOR REDOX PROPERTIES

In Section I we introduce the theoretical framework commonly used in the literature to rationalize the redox properties of proteins from numerical simulations based on the Linear Response Approximation. We start by describing how thermodynamic properties can be computed from microscopic simulations based on the LRA. Formulation within the context of hybrid QM/MM methods is given and we discuss the different FF one can use for carrying out this type of simulation. This allows us to highlight the current methodological limitations.

I.1 The Linear Response Approximation

I.1.a The Marcus theory at the microscopic level

We are interested in Part I in the free energies associated with the reduction of a redox cofactor (C) encapsulated within a protein (P). The protein itself may be solvated in water (globular proteins) or inserted in a membrane:

C@P + 1e- → C^-@P

The starting point is to consider the existence of two electronic states corresponding to the reactants and products. The initial and final states have potential energies E_1 and E_2 respectively. A redox reaction is different from more standard chemical reactions involving the transfer of atoms, or groups of atoms, from two molecular fragments, in the sense that the reduction of C introduces a change of electric charge on C that in turn, induces adjustment of the internal structure of C (inner-sphere contribution) and a different polarization of the environment (outer-sphere contribution). For biological systems the environment encompasses protein residues and solvent (water). By polarization we mean new orientations and translations of atoms or molecules and polarization of the electron cloud (electrostatic induction). The more polar the environment the larger the reorganization to stabilize the new redox form of the cofactor. This specificity of redox processes means that hundreds of thousands of atoms participate in the reaction coordinate.

I.1.b The vertical energy gap as reaction coordinate

A breakthrough came from the work of Warshel[1] who proposed to consider the vertical energy gap $\Delta E_{12} = E_2 - E_1$ as the reaction coordinate. The choice ΔE_{12} as reaction coordinate is justified by the fact that ΔE_{12} collects all the nuclear motions contributing to the progress of the reaction due to their distinct polarization in the two redox states. The free energy for state $x(= 1 \text{ or } 2)$ is expressed by the Landau formula:

$$G_x(\varepsilon) = -\beta ln\big(p_x(\varepsilon)\big) + G_x^0 \qquad (1)$$

with $\beta = 1/k_B T$ and $p_i(\varepsilon)$ the probability of having $\varepsilon = \Delta E_{12}$ energy gap when the system is on electronic state x. G_x^0 is the "full" free energy of the state x. It is given by:

$$G_x^0 = -k_B T ln\left[\int \exp(-\beta E_x)\mathrm{d}\Gamma\right] \qquad (2)$$

where the integration is done over the entire phase space Γ accessible to the system in state x. The probability function $p_x(\varepsilon)$ is given by Eq. 3.

$$p_x(\varepsilon) = \frac{\int \exp(-\beta E_x)\, \delta(\Delta E - \varepsilon) d\Gamma}{\int \exp(-\beta E_x) d\Gamma} = \frac{\int \exp(-\beta E_x)\, \delta(\Delta E - \varepsilon) d\Gamma}{Z_x} \quad (3)$$

From the definition of $p_x(\varepsilon)$ in Eq. 3, if the system is ergodic we have:

$$G_2(\varepsilon) - G_1(\varepsilon) = -\beta ln(p_2(\varepsilon)/p_1(\varepsilon)) = \varepsilon \quad (4)$$

If we assume that $p_i(\varepsilon)$ follows a Gaussian statistics, we can write:

$$p_x(\varepsilon) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{(\varepsilon - \langle \Delta E \rangle_x)^2}{2\sigma_x^2}\right) \quad (5)$$

where $\langle \Delta E \rangle_x$ is the average vertical energy gap when the system is in state x and σ_x is the standard deviation of the distribution. By inserting Eq. 5 into Eq. 1 we have,

$$G_x(\varepsilon) = G_x^0 + \frac{(\varepsilon - \langle \Delta E \rangle_x)^2}{4\lambda_x^{var}} + \frac{k_B T}{2} \ln(4\pi k_B T \lambda_x^{var}) \quad (6)$$

In this equation we have introduced a so-called reorganization energ $\lambda_x^{var} = \beta\sigma_x^2/2 = \beta\langle \delta\Delta E . \delta\Delta E \rangle_x / 2$, with $\delta\Delta E = \Delta E - \langle \Delta E \rangle_x$. As it is defined from the variance of the energy gap we will refer to it as the "variance reorganization energy".

To fulfill Eq. 4 for all values of ε the distributions p_1 and p_2 should have the same standard deviations, therefore leading to the same variance reorganization energies: $\lambda_1^{var} = \lambda_2^{var}$. Applying Eq. 6 for $\varepsilon = \langle \Delta E \rangle_1$ and $\varepsilon = \langle \Delta E \rangle_2$ further leads to the two following relations:

$$\Delta G^0 = G_2^0 - G_1^0 = \frac{\langle \Delta E \rangle_1 + \langle \Delta E \rangle_2}{2} \quad (7)$$

$$\lambda = \frac{\langle \Delta E \rangle_1 - \langle \Delta E \rangle_2}{2} = \lambda^{St} \quad (8)$$

which give an alternative definition of the reorganization energy often referred to as the Stokes reorganization energy λ^{St}. The “Marcus” reorganization free energies λ_x^r are also defined by the reversible work necessary to bring the system from its optimum configuration in state x (x being 1 or 2) to the optimum configuration of the other state involved in the reduction of the cofactor (resp. 2 or 1). Using this definition, one gets:

$$\lambda_1^r = G_1(\langle \Delta E \rangle_2) - G_1(\langle \Delta E \rangle_1) = \frac{(\langle \Delta E \rangle_2 - \langle \Delta E \rangle_1)^2}{4\lambda} = \lambda \quad (9)$$

and a similar result for λ_2^r. As a consequence, the LRA leads to the equality between all the possible definitions of the reorganization energy: $\lambda = \lambda_1^{var} = \lambda_2^{var} = \lambda_1^r = \lambda_2^r = \lambda^{St}$.

Figure 1 depicts a visual representation of the Marcus free energy parabolas.

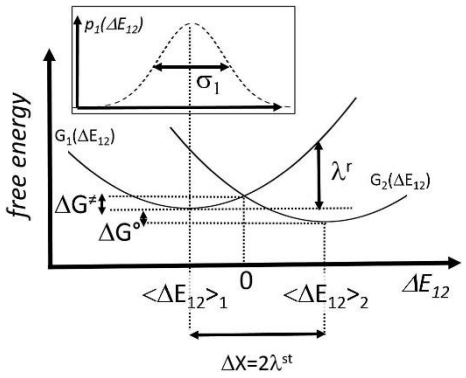


Figure 1. Free energy profile in the Marcus Theory where the collective reaction coordinate is taken to be the diabatic energy gap (Figure extracted from ref [11])

I.2 Potential energy functions

I.2.a Hybrid QM/MM methods

In I.1 we saw the central role of the vertical energy gap in the calculation of the free energies of reduction by numerical simulations. In practice one needs a methodology for calculating the potential energies in the two redox states. Various reviews have been published in the literature. We focus our attention on methodologies based on hybrid QM/MM approaches[12].

i) Electrostatic embedding

A QM electronic structure theory is mandatory to account for the intrinsic electron affinity of the cofactor. Effectively the added electron on the redox cofactor interacts with the bath of electrons of the cofactor, creating geometrical relaxation of the nuclei. The longer-range interactions between the electrons of the cofactor and the atoms belonging to the protein and to the solvent can probably be captured by molecular mechanics FF. The QM/MM partition is straightforward. The QM region describes the cofactor and the MM region encompasses all other atoms. We will use the terms QM atoms and MM atoms to refer to the atoms belonging to the QM and MM regions respectively. We assume that Density Functional Theory (DFT)[13], in its Kohn-Sham formulation[14], is used for the QM region. For the MM region we use the same FF parameters in the two redox states. Two flavors of DFT/MM are possible depending on the treatment of the interaction between the QM and MM atoms, namely the mechanical and electrostatic embedding schemes. The latter is the most satisfactory. In that case the potential energy of state x reads:

$$E_x^{DFT/MMee}[\rho_x] = T_s[\rho_x] + V_{nuc}[\rho_x] + J[\rho_x] + E_{xc}[\rho_x] + V_{embed}[\rho_x] + \sum_{A\epsilon QM}\sum_{K\in MM} Z_A v_{embed} + E_x^{MM} \quad (10)$$

$$E_x^{DFT/MMee}[\rho_x] = E_x^{DFTee}[\rho_x] + E_x^{embed}[\rho_x] + E_x^{MM} \quad (11)$$

ρ_x is the electron density of the QM region in state x. The first four terms on the r.h.s. of Eq. 10 are the kinetic energy of the electrons of the reference Kohn-Sham system (T_s), the interaction energy of the electron density with the nuclei of the QM region (V_{en}), the classical Coulomb interaction among electrons (J) and the exchange-correlation energy (E_{xc}). The sum of these terms defines E_x^{DFT} in Eq. 11. V_{embed} is the embedding energy, that is, the electrostatic interaction energies between the electron density and the MM atoms. With the interaction energy of the QM nuclei and the MM atoms, these terms define E_x^{embed}. Finally E_x^{MM} is the MM energy. In the electrostatic embedding scheme, the electron density is optimized so as to minimize the energy functional $E_x^{DFT/MMee}$ energy. The electron cloud of the QM region is polarized by the MM atoms thanks to the inclusion of the embedding Hamiltonian. The latter is defined as the derivative of V_{embed} with respect to the density ρ_x.

After solving the Kohn-Sham equations for a given configuration of the atom nuclei the two redox states the vertical energy gap is obtained as:

$$\Delta E_{12}^{DFT/MMee} = \Delta E_{12}^{DFTee} + \Delta E_{12}^{embed} + \Delta E_{12}^{MM} \quad (12)$$

After sampling of $\Delta E_{12}^{DFT/MMee}$ for the two potential energy surfaces the free energy of the reaction can be calculated by Eq. 7. The variance reorganization energy

$$\Delta G^{DFT/MMee} = \Delta G_{12}^{DFTee} + \Delta G_{12}^{embed} + \Delta G_{12}^{MM} \quad (13)$$

ii) Non-polarizable - monopolar force fields

The specific form of the last two terms of Eq. 13 depends on the FF used for the MM part. Since we are interested in calculating vertical energy gaps, *i.e*. at constant nuclear/atomic positions, and that the same FF parameters are used for the two redox states, ΔE_{12}^{MM} will always involve only non-bonded terms (electrostatic and van der Waals). The bonded terms (bonds, angles, dihedrals, torsions…) cancel out when taking the difference between the two redox states. The main point is to evaluate the electrostatic interactions involving QM and/or MM atoms. A first distinction has to be made depending on whether the FF incorporates electronic induction (polarizable force fields MMpol) or not. A second

distinction has to be made regarding the treatment of electrostatics arising from MM atoms, using only permanent charges or higher electrostatic multipoles.

For the sake of illustration we assume a simple FF which ignores electronic induction and relies on point charges q_K to describe the electrostatic potential generated by the MM atoms. In that case ΔE_{12}^{MM} vanishes to zero and Eq. 13 reduces to:

$$\Delta E_{12}^{DFT/MMee} = \Delta E_{12}^{DFTee} + \sum_{K \in MM} \int \frac{(\rho_2(r) - \rho_1(r))q_K}{|r - R_K|} dr \quad (14)$$

R_K is the position of MM atom K. The application of the LRA (Eq. 7) leads to:

$$\Delta G^0 = \Delta G^0_{inner-sphere} + \Delta G^0_{outer-sphere} \quad (15)$$

$$\Delta G^0_{inner-sphere} = \frac{\langle \Delta E_{12}^{DFTee} \rangle_1 + \langle \Delta E_{12}^{DFTee} \rangle_2}{2} \quad (16)$$

$$\Delta G^0_{outer-sphere} = \frac{\langle \Delta E_{12}^{embed} \rangle_1 + \langle \Delta E_{12}^{embed} \rangle_2}{2} \quad (17)$$

When using a non-polarizable FF, the free energy of the reaction (and also the Stokes reorganization energy) can thus be written as a sum of an inner-sphere and an outer-sphere contribution. The former being computed at the hybrid DFT/MM level with electrostatic embedding includes polarization of the redox cofactor by the environment. For the modelling of the redox properties of highly polarizable redox cofactors like the special pair within the photo-reactive center, it is important to retain this feature of the QM/MM methodology[15].

iii) Energy evaluation accuracy vs. conformational sampling

To apply Eq. 7-8 from the Marcus theory/LRA, the QM/MM energy gap (Eq.12 or 13) must be sampled for the entire phase space accessible to the systems in the two redox states (see Eq. 2). This is usually done by MD simulations. In general MD has to be run for several tens or hundreds of nanoseconds to reach proper conformational sampling. These timescales are currently inaccessible for plain DFT/MM MD simulations. One possible strategy is to rely on classical MD simulations to conduct molecular sampling in the two redox states, and then post-process the trajectories to evaluate the energy gap at the DFT/MM level[16]. This strategy is however delicate to follow. First because FF parameters must be available for the cofactor in the two redox states. Second because the ensemble of configurations sampled from the FF PES will not match exactly the ensemble of configurations that would have been obtained from exploration of the DFT/MM PES. This can introduce artefacts in the evaluation of Eq. 14-16. It should be also remarked that the DFT/MM PES can become extremely computationally consuming. In general the vertical energy gap fluctuates on the sub-picosecond time scale. Therefore

a post-treatment of a 100 ns classical MD simulation every picosecond leads to 10^5 DFT/MM evaluations! In summary we see the nature of the central dilemma faced with the numerical simulation of redox properties of proteins: *finding an optimum balance between extensiveness of conformational sampling and accuracy of the potential energy functions*.

I.2.b The simplified QM+MM scheme

i) Decoupling hypothesis

Among the various alternative strategies that can be followed to reduce the computational cost is the so-called QM+MM approach[16b]. It is based on the assumption that reorganization of the inner-sphere upon redox change is largely caused by local electronic and nuclear relaxation and that it is decoupled from the reorganization of the environment. Alternatively said, we ignore mechanical and electrostatic coupling between the cofactor and its environment. This hypothesis has been tested in various redox enzymes such as heme proteins or cryptochromes by comparison with MD simulations based on hybrid QM/MM PES[17]. It may not be true for other proteins like the photoreactive center where the redox cofactors can be extremely polarizable. Adopting the QM+MM approach to simulate redox properties is a choice that requires precautions. The mechanical embedding formulation of the DFT/MM energy provides a good starting point to reach the QM+MM formalism. The energy reads, for a non-polarizable FF:

$$E_x^{DFT/MMme} = T_s[\rho_x] + V_{nuc}[\rho_x] + J[\rho_x] + E_{xc}[\rho_x] + E_x^{cl} + E_x^{MM} \quad (18)$$

$$E_x^{cl} = \sum_{A \epsilon QM} \sum_{K \in MM} V_{A-K} \quad (19)$$

$$E_x^{DFT/MMme} = E_x^{DFTgp} + E_x^{cl} + E_x^{MM} \quad (20)$$

The main difference with Eq. 10 is that the interaction term (E_x^{cl}) between the QM and MM regions is calculated not from the electron density but by classical electrostatics. With a FF relying on monopolar approximations (charges) the V_{A-K} interaction term between QM atom A and MM atom K is calculated by Coulomb's law. But with more advanced force fields relying on permanent multipoles V_{A-K} takes more complex forms. Consequently the Kohn-Sham Hamiltonian doesn't include embedding by the environment. The electron density is not polarized by the MM region. This is a "gas phase calculation" leading to energy E_x^{DFTgp}. Importantly the QM+MM approach can be made compatible with polarizable FF as long as the cofactor itself remains non-polarizable. Otherwise the decoupling assumption underlying the QM+MM idea breaks down.

ii)Inner-sphere contribution

Thanks to this decoupling hypothesis the inner sphere contribution (ΔE_{in}) can be estimated from the energies of the isolated cofactor in the gas phase.

$$IP = E_2(R_1) - E_1(R_1) \quad (21)$$

$$EA = E_2(R_2) - E_1(R_2) \quad (22)$$

$$\lambda_{inner,1} = E_1(R_2) - E_1(R_1) \quad (23)$$

$$\lambda_{inner,2} = E_2(R_1) - E_2(R_2) \quad (24)$$

$$\Delta E_{in} = IP - \lambda_2 = E_2(R_2) - E_1(R_1) \quad (25)$$

In these equations $E_x(R_y)$ denotes the gas phase DFT energy for the cofactor in state x for parametric nuclear coordinates corresponding to the minimum of the potential energy surface of state y.

iii) Outer-sphere contribution

The outer sphere contribution is more involved since the environment has many more atoms and a larger number of degrees of freedom. On the other hand, it involves only classical terms that can be evaluated by classical MD simulations, eventually polarizable. Thus by assuming a decoupling between the inner- and outer-spheres the major issue to finding the optimum balance has been partially lifted because the computationally time consuming quantum mechanical part is separated from the sampling of the environment by MD simulations. If the QM+MM approximation is valid, the main task for improving reliability of redox calculations is to improve the evaluation of electrostatic interactions for the outer-sphere. This largely relies on the quality of the force fields that are used. This important aspect will be surveyed now.

I.3 Force fields for redox property simulations

I.3.a. First generation force fields

Classical Molecular dynamics simulate the time evolution of energy of a system as a function of its atomic coordinates. The accuracy of such simulations relies on the set of potential energy functions and parameters referred to as a force field. First generation FF incorporate a relatively simple potential energy functions as shown in Eq 26. The first three terms correspond to covalent (bonded) interactions, *i.e*. bonds, angles and torsions, while the last terms are describing non-bonded interactions. Electrostatic energy is described with Coulomb interactions between point charges q_K on each atom. Van der Waals (vdW) interactions are often represented by a Lennard-Jones 6-12 potential.

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsion} k_\phi[cos(n\phi + \delta) + 1] + \sum_{\substack{nonbond \\ paris}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right] \quad (26)$$

First molecular dynamics simulations of proteins were carried out in 1977[18]. Over the past 30 years, a large number of FFs have been developed, including AMBER[19], CHARMM[20], GROMOS[21], OPLS[22] and many others. They can be used with massively optimized codes for production runs[23]. These force fields share similar functional forms of Eq. 26. This generation of FF is also referred as additive FF and some developments are still ongoing[24]. They have been widely used in the study of proteins[25]. However, as many simulations have shown their limitations, advanced FF for proteins using more accurate potential energy functions are needed especially to get a better accuracy in the calculations of electrostatic effects[26].

I.3.b. On the importance of electrostatic induction

It is well known that taking into account electronic polarization effects is crucial when modeling the redox free energy[16b, 27]. Polarization refers to the fact that the electron clouds of the environment molecules are deformed by the change of the redox state of the cofactor. There are three well-known theoretical models to include polarization effects in FF: fluctuating charges model, Drude oscillator model and induced dipole model. We introduce here their theoretical models in detail including their history, the mathematical formulas to include polarization, the corresponding energy term and some examples of the existing force fields. Then we mention their pros and cons in the case of modeling redox potentials of proteins.

i) Fluctuating charge model

The fluctuating charge (FQ) model[28] also known as charge equilibration or electronegativity equalization model[28e], uses the same partial charge description as traditional non-polarizable force fields. The difference is that these partial charges on each atom are allowed to change in order to adapt to different electrostatic environments during the simulation. The set of partial charges is calculated by minimizing the electrostatic energy of the given structure. The charge conservation is ensured by applying the principle of electronegativity equalization. The charge-dependent energy for a system of M molecules containing N_i atoms per molecule is expressed as

$$E(Q,R) = \sum_{i=1}^{M}\sum_{\alpha=1}^{N} \chi_{i\alpha} Q_{i\alpha} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\alpha=1}^{N_i}\sum_{\beta=1}^{N_i} J_{i\alpha i\beta} Q_{i\alpha} Q_{j\beta} + \frac{1}{2}\sum_{i=1}^{MN'}\sum_{j=1}^{MN'} \frac{Q_i Q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{j=1}^{M} \lambda_i \left(\sum_{i=1}^{N} Q_{ij} - Q_j^{total} \right) \quad (27)$$

where χ is the atomic electronegativity that controls the direction of electron flow, J is the atomic hardness that represents the resistance to deformation of electron flow, Q_i is the partial charge on atomic site i, λ_i is the Lagrange multiplier which ensures the conservation of the total charge. These parameters are optimized to reproduce the molecular dipole moments and the associated molecular polarization response. Either extended Lagrangian or self-consistent iteration can be used to compute the fluctuating charges in the MD simulations. The first version of this model was developed for water in the year of 1985[29]. Later on, it has been developed by Patel, Brooks and coworkers within the CHARMM program named CHARMM-FQ[30] and has been applied to simulations of biomolecules[31].

ii) Drude oscillator model

In the Drude oscillator model[32], electronic polarization is based on the presence of a Drude particle attached to its parent atom *via* a harmonic spring with a defined force constant. This force is associated with the electric field felt by the Drude particle

The Drude oscillator, named after Paul Drude in 1902[33], also is also known in the literature as the Shell model[32a] or the Charge-On-Spring model[34]. Originally, it was designed to study ionic materials in the solid state. In this model, electronic polarization is accounted for by the presence of an auxiliary particle, called the Drude particle, which is attached to its parent atom via a harmonic spring, the force constant of which is k_D. The Drude particle carries a charge q_D, and the charge of the parent atom A is replaced by $q_A = q - q_D$ to preserve the net charge of the atom–Drude (A-D) pair (q). q_D is defined according to the isotropic atomic polarizability of the parent atom $\alpha = \frac{{q_D}^2}{k_D}$. In the presence of an electric field $\boldsymbol{F}$, the Drude particle oscillates around a displaced position $d = \frac{q_D E}{k_D}$. The induced atomic dipole is

$$\boldsymbol{\mu} = q_D d = \frac{{q_D}^2 \boldsymbol{F}}{k_D}.$$

After adding Drude particles, the functional form of the force field becomes that of a polarizable FF. The energy terms for intramolecular energy (bond lengths, angles, and dihedrals) and the vdW

interactions (Lennard-Jones ''12–6'' nonpolar contribution) remain the same as an additive FF. Only the electrostatic interactions are changed. Except the electrostatic interactions between atom-atom(V_{AA}), the electrostatic energies between atom-Drude particle (V_{AD}), and Drude-Drude particles (V_{DD}), have to be computed. One last term must to be added is the harmonic self-energy of the Drude oscillators, which represents the polarization energy (V_{pol}).

$$E_{ele} = V_{AA} + V_{AD} + V_{DD} + V_{pol} \tag{28}$$

$$E_{ele} = \sum_{i<j}^{N} \frac{q_{A(i)} q_{A(j)}}{|r_A(i) - r_A(j)|} + \sum_{i<j}^{N,N_D} \frac{q_{D(i)} q_{A(j)}}{|r_D(i) - r_A(j)|} + \sum_{i<j}^{N_D} \frac{q_{D(i)} q_{D(j)}}{|r_D(i) - r_D(j)|} + \frac{1}{2} \sum_{i}^{N_D} k_D (r_D(i) - r_A(i))^2 \tag{29}$$

One version of polarizable FF based on the Drude oscillator model is named "Drude-2013". It was built on its origins in the CHARMM additive FF. Its development and applications have recently been reviewed[35]. In the Drude-2013 polarizable FF[36], the Drude particles are only associated to non-hydrogen atoms for the sake of computational efficiency. The restoring force constant k_D is assumed to be the same for all atoms with a fixed value of 1000 kcal/mol/Å^2, such that the charge q_D is the parameter that governs the magnitude of α for a given atom. In addition, the model includes virtual particles representative of lone pairs, typically located on hydrogen bond-acceptor atoms. The anisotropic molecular polarization can be achieved by using a matrix form of the force constant k_D and decomposing the displacement distance in three dimensions[37]. The combination of lone pairs and anisotropic polarization leads to an improved description of hydrogen bonding in polar compounds and interactions with ions as a function of orientation[37]. The interactions between induced dipoles (but not charge-dipole interactions) are explicitly included for 1–2 and 1–3 atom pairs with short-range Thole damping to avoid a polarization catastrophe[38].

iii) Induced dipole model

The third method presented here for including polarization effects into force fields is the induced dipole model. The theory of atom point dipoles was first introduced by Silberstein[39] and then applied by Applequist[40]. In this model, a classical point dipole moment is induced at each polarizable atom site according to the electric field felt by that site. The induced dipole at each atomic site is computed as $\boldsymbol{\mu_i} = \alpha_i \boldsymbol{F_i}$, where α_i represents the atomic polarizability. α_i can be generally treated as an isotropic quantity or an anisotropic tensor. $\boldsymbol{F}_i$ is the total electric field at atom i and consists of the electric field created by the other permanent multipoles ($\boldsymbol{F_i^0}$) plus the field of the other induced dipoles in the systems ($\boldsymbol{F}_i^{ind}$).

In the case of point charge model (for example AMBERff02[41]),

$$M_j = q_j,\ T_\alpha^{ij} = \frac{1}{r_{ij}} \text{ and } \quad T_{\alpha\beta}^{ij} = \frac{1}{r_{ij}^3}\boldsymbol{I} - \frac{3}{r_{ij}^5}\begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \tag{30}$$

where M_j is the charge on atom center *j.* T is the interaction tensor between sites i and j. In other FFs, M_j may be developed till higher orders of multipoles. This combination of polarization and multipolar electrostatics will be described in section III.

Since the induced dipoles alter the field at each site, the procedure must be iterated to generate a self-consistent set of "mutual" induced dipoles arising from the mutual polarization until the induced dipoles at each site reach convergence. This is computationally demanding and is typically a limiting factor in efficiency of such simulations. A number of approaches has been contemplated to overcome the limitations of the SCF procedure[42].

This model reproduces well the average polarizabilities, but the polarization becomes infinite when the induced dipoles interaction distance is too small. To correct this 'polarization catastrophe' phenomenon, Thole proposes to use a damping factor[38].

iv) Pros and cons

The advantage of the Fluctuating Charge model is that the polarization is obtained without introducing new interactions. Thus no additional term has to be computed compared to non-polarizable FFs. However it needs a much smaller time step. The major disadvantage is that the charge-flow is limited by the chemical connectivity. It cannot represent polarization that does not occur in the direction of bonds. Thus it cannot capture the out of plane polarization (such as benzene, bifurcated hydrogen bonding). As these types of interactions are important in protein interactions, we have considered that this is not a relevant choice for the simulation of redox potentials of heme proteins.

The computational cost of the Drude model originates in the large increase in the number of particles and lone pairs in the simulated system. Compared to the classical induced dipole method, the Drude model based FF involves less complex numerical algorithms since a point charge framework is retained. With the implementation in the NAMD package, the computational cost is about 1.2 to 1.8 times that of fixed-charge models[23]. The Drude-2013 FF has been applied to many systems including biomolecules, yielding quantitative improvements over additive force fields thanks to the inclusion of cooperative effects. In the case of simulated redox potentials, the Drude FF was applied to evaluate the reorganization free energy for electron self-exchange in aqueous Ru^{II} to Ru^{III}. The reorganization free energy was reduced by 22% in comparison to a non-polarizable water model[27a].

In the induced dipole model[43], since the induced dipoles alter the field at each site, the procedure must be iterated until the induced dipoles at each site reach convergence. This is computationally demanding and is typically a limiting factor in the efficiency of such MD simulations. At the same time, with the iterative scheme one better reproduces anisotropy and non-additivity of the molecular polarization response across many different compounds. This is the major advantage over the Drude model for instance. The conformational dependence of electrostatics can be significant. One well known FF of this model is a point charge model named AMBERff02[41]. It has been applied to simulate redox properties in several papers[26b, 44]. Thus we would like to test this model too.

I.3.c. Electrostatic Multipole based force fields

In most FF, electrostatics between molecules are approximated by Coulomb's law between point charges (and eventually as we just saw by induction). Yet, it has been shown that the error on the computation of the molecular electrostatic potential can be reduced by orders of magnitude upon complementing atomic monopole interactions by interactions involving permanent dipole and quadrupole moments[45]. One may anticipate that FF going beyond the point charge approximation may significantly improve the accuracy of the computed redox potentials. Several FFs have multipole moment descriptions such as AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications)[46], SIBFA (Sum of Interactions Between Fragments *Ab Initio* Computed)[47] or NEMO (Non-Empirical Molecular Orbital)[48]. Among them, AMOEBA is the most widely used FF because of its reasonable computational cost. A first decisive step toward the use of AMOEBA to the evaluation of redox properties of heme proteins, namely the parameterization of AMOEBA for the heme cofactor, will be reported in Chapter 2.

In summary, in Section I we have detailed the formalism for simulating redox potentials of cofactors embedded in protein matrices from microscopic considerations. We highlighted current weaknesses of the force fields which limit the predictability of these approaches to real systems. The bet of this PhD thesis is that significant progress toward more accurate calculations will be achieved along with a significant improvement of electrostatic interactions between the cofactors and the environment. Chapter 2 will report contributions toward that objective, by employing to AMOEBA simulations. For the moment we report in Section II application of the LRA formalism for a series of heme proteins. This will enable the reader to understand how this formalism actually works in practice. For simplicity we have chosen a non-polarizable force field.

II. THE LRA MACHINERY IN ACTION

In Section II we illustrate the Marcus Theory/Linear Response Approximation on actual hemeproteins. These results will allow us to understand how the LRA can be validated from microscopic simulations and what kind of information can be extracted from them at the atomic level. For the sake of simplicity the simulations are carried out with non-polarizable FF. The QM+MM strategy has been tested in the case of heme proteins and was found to be adequate[17a].

II.1 Model systems of heme proteins

II.1.a selection of heme proteins

The main line for Chapter 1-2 is to seek for alternative FF that would enable more reliable estimates of protein redox cofactors. The accuracy of calculated redox potentials depends on many factors, not only on the FF. To attempt separating errors from FF potential functions from other factors (extensiveness of computational sampling, limit of validity of the LRA, polarization of the cofactors…), we have chosen a series of six proteins bearing in mind the following points. First it has been demonstrated by others that in general the QM+MM strategy applies well to hemeproteins[16b]. This simplifies the modelling and will put emphasis on MM FF electrostatic evaluation. Second the chosen proteins are small. Therefore we hope to alleviate as much as possible the problem of extensive conformational sampling[†]. Third the redox potential for these proteins are known from experiments. Finally the redox potentials are scattered over more that 500 mV, from -210 mV to +350 mV. Figure 2 depicts the crystal or NMR structures obtained from the Protein Data Bank (PDB)[49].

Table 1 lists some characteristics of the cytochromes. In particular the entry "ligands" indicates the amino acid residues linked in axial positions to the heme. The experimental redox potential measured with respect to the Standard Hydrogen Electrode (SHE) are reported. The six cytochromes can be classified into two types. As shown in Figure 2(a) and (b) are folded within a globin motif and belong to type of cytochrome.c (Cyt.c), while Figure 2 (c) to (f) are formed of alpha helices and beta sheets (αβ) and belong to type of cytochrome.b (Cyt.b). Their cofactor prosthetic group and ligands are different too. Cyt.c (label a,b) consist of a heme and two ligands, histidine and methionine residues connected to the iron cation, with two cysteine residues covalently connected through thioether linkages to heme. On the other hand Cyt. b (label c,d,e,f) consist of heme ligated by the side chains of two histidine residues.

[†] Indeed, some complex redox proteins, involving multiple domains may necessitate hundreds of nanoseconds of MD simulations to converge redox potentials.

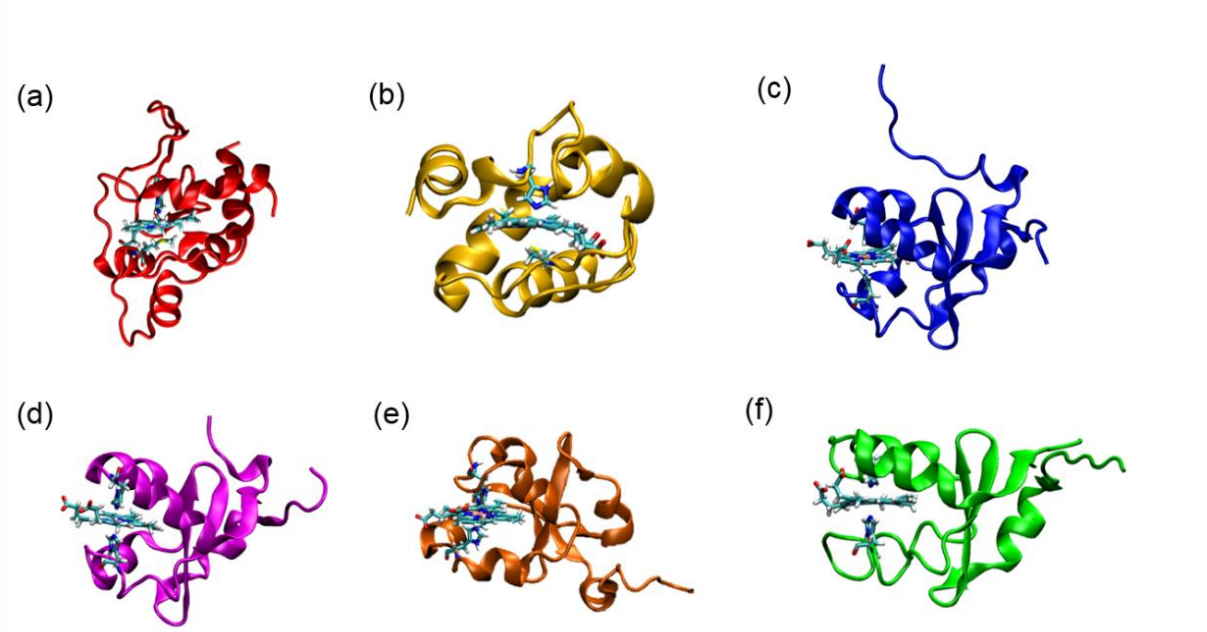


Figure 2. Cytochromes selected to illustrate the LRA/QM+MM machinary. (a) Globin (*R. capsulatus* C2), (b) Globin (*P. aeruginosa*, C_{551}), (c) αβ (*B. Taurus*), (d) αβ (*Rattus* N. V451/V61I), (e) αβ (*Rattus N.*), (f) αβ (*E. vacuolata*)

Three cytochrome b5 (label c,d,f) with similar redox potential (difference of around 50 mV) are studied to investigate how modest our method in reproducing the accuracy of electrochemistry. The experimental values are measured relative to the standard hydrogen electrode (E_{exp}^{SHE}). The oxidation free energies for SHE are shifted by 4.44 V in the experimental data. The last column in Table 1 (E_{exp}^{0}) are values to be compared with simulation results.

Table 1. Information of selected models for six Cytochromes.

Label	type	Organism	PDB	#RES	ligands	E_{exp}^{SHE}(mV)	E_{exp}^{0} (V)
a	c2	*R. capsulatus*	1c2r	116	His-Met	350	4.790
b	c551	*P. aeruginosa*	351c	82	His-Met	270	4.710
c	b5	*B. Taurus*	1cyo	93	His-His	-10	4.430
d	b5	*Rattus n.. V451/V61I*	1eue	94	His-His	-63	4.377
e	b5	*Rattus n.*	1b5a	86	His-His	-102	4.338
f	b558	*E. vacuolata*	1cxy	90	His-His	-210	4.230

Labels: identifiers for Figure1,5 and 6. PDB: Protein Data Band code. #RES: number of residues in the protein structure. E_{exp}^{SHE}: the experimental redox potential relative to a reference electrode (SHE)[50]. The uncertainty is around $\pm 30mV$. E_{exp}^{0}: The experimental relative result plus the potential of SHE=4.44 V.

II.1.b Structure preparation

The structures have been obtained from the Protein Data bank (PDB entry codes in Table 1). All X-ray structures had been obtained with high resolution (around 2 Å). The X-ray structure of 1c2r is a dimer, although the protein is monomeric in solution. Only the first chain (A) was simulated. The PDB structure of 1cyo misses the last five residues and the structure 1cxy misses the first four residues and the last five residues. This situation is likely to be related to the flexibility of these terminal loops that

prevents structural identification. However since the sequence is known for these proteins, the missing residues were added with the CHARMM package[51] (version 35). The other structures have a complete structure. The structure of 1B5A was obtained from NMR measurements which had the position of hydrogen atoms, the other X-ray structures were hydrogenated with the HBUILD module of the CHARMM package and solvated in a water box. Three different water box sizes were tested: 90, 100 and 110 $\AA^3$ (with TIP3[52] water molecules). All crystallographic water molecules were retained. The *p*Ka values of ionizable amino-acid side chains were determined using the Propka server[53]. It was found that all amino acid side chains adopt standard protonation states at neutral pH. Therefore glutamic acid and aspartic acid side chains were deprotonated and all lysine side chains were protonated. We added Na^+ and Cl^- counter ions to ensure electrical neutrality and to reach an ionic strength of 0.015 mol/L. The CHARMM22 FF[54] with CMAP correction (to improve backbone torsions) was chosen for the classical MD simulations together with the TIP3P FF for water. The FF parameters for the heme with deprotonated propionate and for the axial ligands were taken from all22_prot_heme[55]. The energies of the whole systems were minimized for 2500 steps with CHARMM.

II.2 Inner sphere energy: QM calculations in gas phase

Within the QM+MM methodology the inner-sphere is modelled from gas phase calculations of the redox cofactor. The structures of the heme prosthetic group treated with QM calculations are depicted in Figure 3. For these six models, we have two different types of inner spheres. Figure 3(a) is the inner-sphere of Cyt.c. The amino acid backbone atoms of histidine, methionine and cysteine have been truncated as methyl-imidazole (ImMe), ethyl-methyl-sulfide (EMS) and two methyl-methyl-sulfide (MMS) respectively. Figure 3(b) is the inner sphere of Cyt.b modeled as iron porphyrin (PFe) with two ImMe axial ligands. We have protonated the two propionate groups of heme. This is mandatory to control the electronic structure of the iron cation in the complex. Indeed in the gas phase, contrary to protein or solvent media, the molecular orbitals of the carboxylate groups are not stabilized and are of similar energies to the iron *3d* orbitals. Therefore the self-consistent-field iterations involved in DFT calculations for the ferric state tend to oxidize the carboxylate functions leading to the Fe(II)COO• electronic structure instead of the expected Fe(III)COO^- electronic structure. This problem is not present with full QM/MM where the negative carboxylate functions are stabilized by counter-ions or hydrogen bond donors. Protonation avoids this inconvenience of the QM+MM framework.

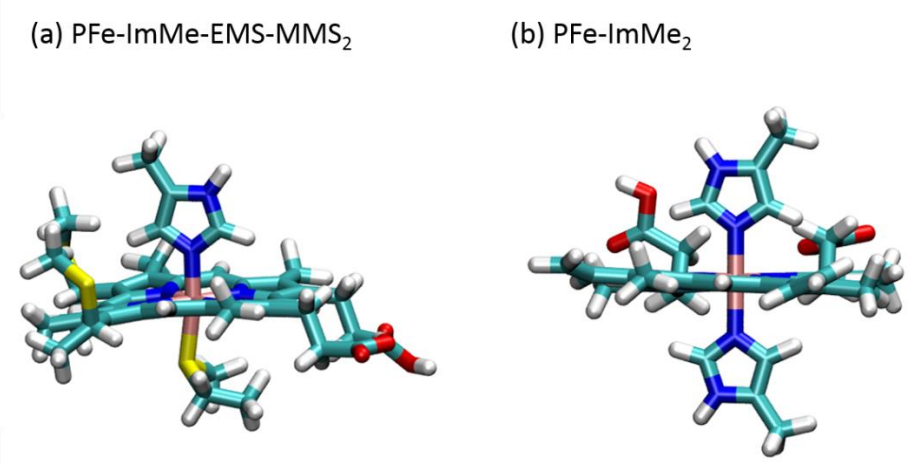


Figure 3. Inner sphere structure optimized by DFT. (a) Inner sphere of Cyt.c: PFe-ImMe-EMS-MMS$_2$ (b) Inner sphere of Cyt. b: PFe-ImMe2.

The structures have been optimized with deMon2k[56] with the OPTX-PBE functional[57]. This functional was shown to perform well for transition metal cations[58]. The DZVP-GGA[59] basis set is used for all atoms. The auxiliary basis sets GEN-A2 (for C, H and O) and GEN-A2* (for Fe, N and S, the star denotes the inclusion of *f* and *g* auxiliary functions) have been employed to expand the auxiliary electronic densities as employed in the so-called auxiliary DFT framework. An adaptive DFT grid of fine accuracy has been used for the numerical evaluations of the XC energy and potential and a multipolar expansion scheme has been employed to evaluate the long-range Coulomb integrals. The relative energy to the singlet states are summarized in Table 2. The singlet and doublet states are the most stable for the ferrous and ferric states respectively. This is true for both complexes, in agreement with experimental values. The energies of the inner sphere are calculated as Fe^{II} the initial state (1) and Fe^{III} the final state (2) (Eq. 25). We obtained the ΔE_{in} for PFe-ImMe2 of 4.827 eV and PFe-ImMe-EMS-MMS2 of 5.304 eV.

Table 2. Relative minimum energies of two different type of cofactors in oxidized and reduced forms calculated with DFT. Energy of each spin state are shown in eV. The results of singlet state are taken as reference.

Oxidized form	**doublet**	**quadruplet**	**sextet**
PFe^{III}-$ImMe_2$	4.827	5.136	5.436
PFe^{III}-ImMe-EMS-MMS_2	5.304	5.352	5.869
Reduced form	**singlet**	**triplet**	**quintuplet**
PFe^{II}-$ImMe_2$	0.000	0.680	0.799
PFe^{II}-ImMe-EMS-MMS_2	0.000	0.413	0.696

II.3 Outer sphere energy: sampling from MD simulations

II.3.a Computational setup

i) Equilibration

The outer sphere contributions are obtained from MD simulations of the solvated proteins. Since the protein structures are different for each protein, MD simulations for the six proteins, each in the two states have been carried out. The CHARMM22 FF with CMAP corrections have been used[54]. The point charges of heme cofactor were modified for the two different redox states: Fe^{II} and Fe^{III}. Atom charges were obtained from population analysis of the DFT electronic densities of optimized structures with the iterative Hirshfeld scheme[60]. The optimized structures were computed with two propionates protonated. However, in the MD simulations, the cofactor is not protonated. The charges of COO^- of the two propionates were included in the outer-sphere using the parameters from CHARMM FF (atomic charges of 0.62 for C and -0.76 for O). The extra charges were added to the CH_2 connected to COO^- to ensure a total charge of -2 and -1 for Fe^{II} and Fe^{III} states, respectively. The point charges used for cofactors PFe-$ImMe_2$ and PFe-ImMe-EMS-MMS_2 are summarized in Tables I and II of Annex I.

MD simulations have been carried out with the NAMD software[61] (version 2.0). Equilibration phase has been carried out in the isothermal–isobaric ensemble (NPT) under periodic boundary conditions. The velocity Verlet integrator has been used with a 2 fs time step. Langevin barostat was applied with a target pressure of 1 bar. The solvated protein was initially minimized for 10,000 steps and subsequently equilibrated with all protein atoms kept frozen. The temperature was increased slowly from 25 K to 298 K with increase of 50 K each step during 20 ps MD. The temperature was then fixed to 298 K with a Langevin thermostat. The protein was then slowly released by applying harmonic restraints around the initial positions with force constants of 10.0, 5.0 and 1.0 kcal $mol^{-1}Å^{-2}$. The duration of each of these runs was 100 ps. Eventually, all position restraints were dropped and the protein was equilibrated for 500 ps. After this equilibration, the volume was held with the same temperature. The particle-mesh Ewald (PME) method[62] was employed for the calculation of electrostatic interactions to avoid truncation of these long-range forces. Nonbonding interactions were treated using a cutoff of 12.0 Å. Production runs in the canonical ensemble (NVT) were then carried out for each state. Geometries were saved every ps during each simulation. Thus geometries for each state were collected form each trajectory trj(Fe^{II}) and trj(Fe^{III}).

ii) Numerical uncertainties

The vertical energy gap of each trajectory was then calculated. The uncertainty on the energy gap is calculated as:

$$u = \frac{\sigma}{\sqrt{N_{eff}}} \quad (31)$$

$$U = \frac{u_1 + u_2}{2} \quad (32)$$

where σ is the standard deviation of the energy gap. N_{eff} is the effective statistical sample. It has been calculated with the program R[63] and the CODA package[64]. The uncertainty U of ΔG^0_{out} and λ^{st} can be calculated as the average of the uncertainty of the energy gap of each state (u_1, u_2).

II.3.b Box sizes, simulation lengths, starting structures

Several simulation parameters need to be tested. For example the size of the water box, the effect of the initial structure or the length of simulations are likely to influence the computed redox potentials or the reorganization energies.

i) Simulation box size

The choice of the size of water solvation boxes complies to various rules. Of course one should choose large enough water boxes to solvate the system entirely, while trying to minimize its size to reduce the computational cost. One should also ensure that distances between periodic images are larger than the cutoff used for non-bonding interactions. This means that the minimal distance between the solute protein and the wall should be bigger than half the cutoff used. If not, we get duplicate force evaluations from the images. When using an NPT ensemble, the box dimension will fluctuate due to the pressure and thus it is better to have an even bigger size than the minimal one.

Cyt. c551 (PDB: 351c) was chosen as an example. The cell size of Cyt.c551 is 30*50*50 Å. We decided to test with a cube box since we will used it for other proteins. Three different water boxe sizes were chosen: (90 Å)^3 , (100 Å)^3 and (110 Å)^3. For each of them we run 30 ns MD in each state. The first 4 ns were discarded to evaluate the energy gaps. Outer-sphere free energies $\Delta G^0_{outer-sphere}$ and reorganization energies λ^{st} were obtained as described before and the results are summarized in Table 3. These tests have been carried out with a preliminary set of FF parameters for the heme, that we finally refined for the production runs (see below).

As shown in Table 3, the uncertainties are quite small, around 0.006 eV. This indicates that we have enough vertical gap values to have a good statistical knowledge on the average energy gaps, therefore on ΔG^0_{out} and λ^{st}. The value of G^0_{out} are very similar in each case, with a difference of around 0.01 eV. However, we are not sure if the ΔG^0_{out} is converged during 30 ns.

Table 3. Free energies and reorganization energies obtained with different sizes of water box with Cyt. c551 in eV.

Water box edge (Å)	ΔG^0_{out}	λ^{st}	U
90	0.223	0.243	0.004
100	0.234	0.208	0.006
110	0.238	0.237	0.007

For this, the variation of ΔG^0_{out} were calculated for each case. As mentioned before, the energy gap of the first 4 ns was dropped. The free energy obtained from all dynamics is used as reference (G^0_{ref}). Free energy obtain with simulation time of t (ΔG^0_t) is calculated during the trajectories. The difference of them gives the fluctuation of ΔG^0_{out}. Results are shown in Figure 4.

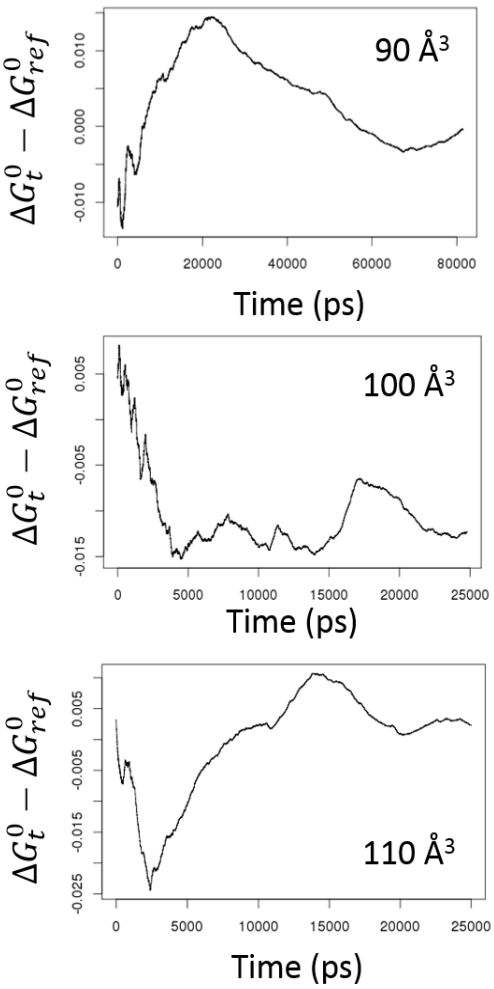


Figure 4. Fluctuation of free energy during dynamics with different sizes of water box.

ii) Simulation length

As the beginning we ran 30 ns for each MD. The variation in the case of the water box with (90 Å)^3 is around 0.01 eV. Thus we decided to continue the dynamics until 80 ns. We observed that the fluctuation converges at the end of 80 ns. However, this could be a too expensive simulation time. In the case of (100 Å)^3 and (110 Å)^3, the fluctuation converges within 30 ns with a value within 0.005 eV. Thus we decided to choose a water box of (100 Å)^3 which has the same accuracy as (110 Å)^3 but which is less expensive. With this water box, we have around 32,100 TIP3[52] water molecules. The distance between the protein and the wall is at least around 25 Å which is c.a. twice bigger than the cutoff of 12 Å. One thing to be mentioned is that these results were obtained out a wrong version of inner sphere charge. However the total charge of the two states was correct. Since it is quite expensive to redo all these tests, we assumed that with the new version of charge distribution, the result will be similar. All other simulations were then carried out with a water box of (110 Å)^3 and with the right version of charge distribution.

II.3.c Stabilities of MD simulations

We now report results for the six proteins depicted in Figure 2. We run 30 ns dynamics for each redox state. The root-mean-square deviation (RMSD) of protein backbone during the dynamics is shown in Figure 5. The PDB structures after energy minimization (section II.1.b) are taken as reference. The

RMSD of the ferrous state is in black while that for the ferric state is in red. Labels (a-f) for six cytochromes are the same as in Table 1.

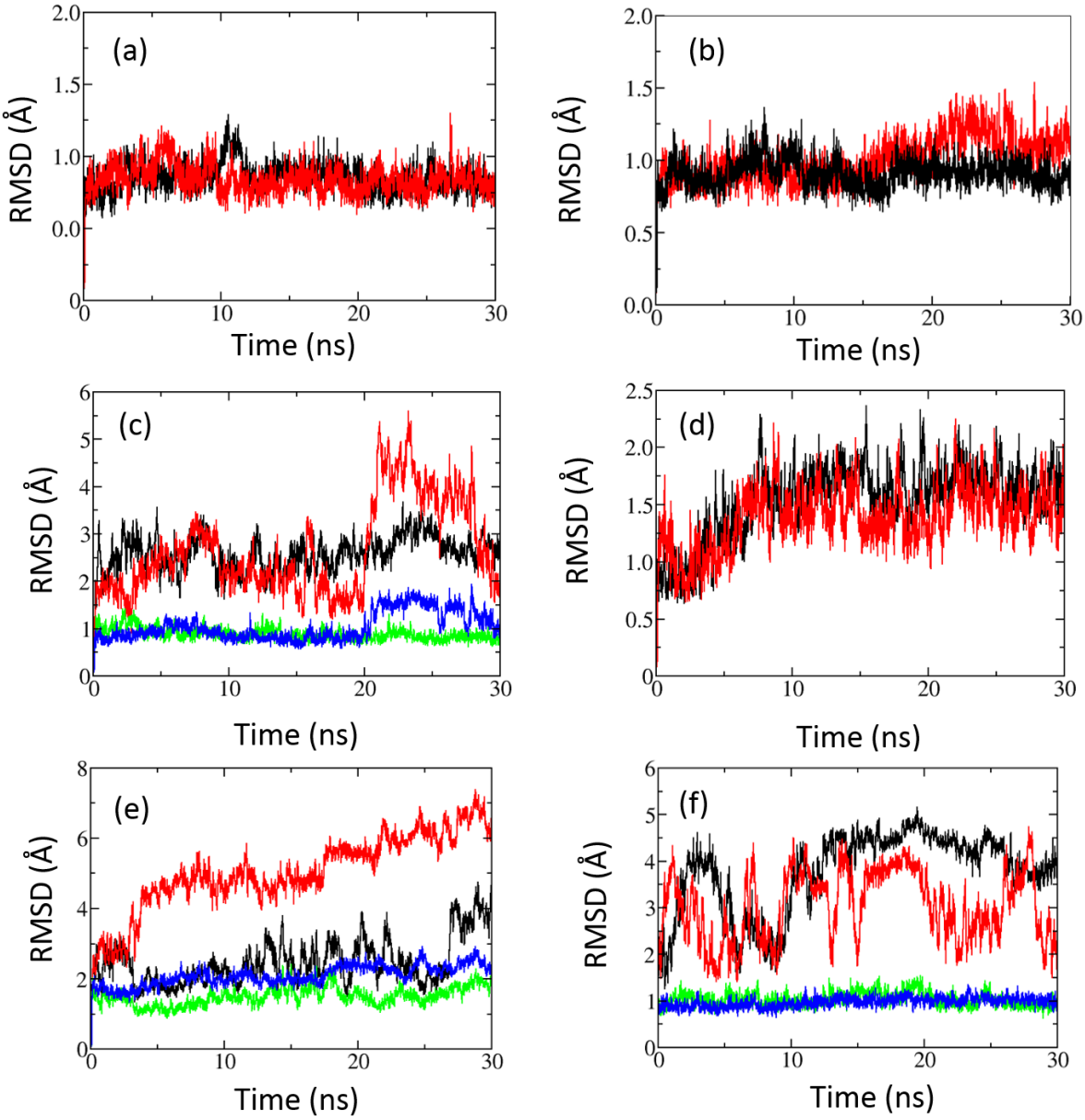


Figure 5. Root-mean-square deviation (RMSD) of six cytochromes protein backbone during 30ns molecular dynamics. The label is the same as in Table 1. Black and green is for dynamics in Fe^{II} form. Red and blue is dynamics in Fe^{III} form. Black and red is RMSD of whole proteins backbond while Green and Blue is RMSD of protein excluding flexible residues at the two end of proteins sequence. (c) we cut last five residues which were not resolved in PDB. (e) we cut the last seven residues. (f) we cut the first four and last five residues which was no resolute in PDB.

RMSD for proteins (a), (b) and (d) are within 2 Å, which means that the dynamics are rather stable. For proteins (c) (e) and (f), the RMSD are much higher. For Protein(c) and (f) we added residue coordinates that were misssing in the PDB structure, while protein (e) had been obtained by Nuclear Magnetic Resonance. We recomputed the RMSD exlucding the terminal residues (green and blue curves). These curves fluctuate around 2 Å like for the other proteins. This indicates that the large RMSD were due to

these terminal residues which adopt very flexible conformations. This is consistent with the fact thah the position of these atoms could not be obtianed in X-ray sturctures.

In conclusion, apart from the inherent flexibility of the terminal loop for some proteins, the RMSD of all these dynamics are less than 2.5 Å. This indicated that MD simulations produced stable structures for protein backbone atoms.

II.3.d Marcus parameters

i) On the validity of the Linear Response Approximation

As a first step, we wish to check whether the hypothesis of Gaussian distributions for the energy gap is correct for the proteins considered here. This is equivalent to the hypothesis of parabolic shapes of the free energy functions. We show on Figure 6 the free energy curves. For each graph and each diabatic state, two curves are plotted. The dashed-dotted curves correspond to parabolas obtained from the fitting of the probability distribution of p_x (see Eq 3). The linear regression coefficients reported in Table 4 are all close to 1.00. This indicates that the Gaussian approximation for p_x is justified (Eq. 5) for all the proteins in all redox states considered here. Note that most of the fitting points are located close to the energy minima of the parabolas (circles on the graphs).

Table 4: Linear regression coefficients associated to the linear fitting of the probability distributions (p_x for $x = 1$ or 2) to Gaussian functions (Eq. 3).

Label	$R^2_{(1)}$	$R^2_{(2)}$
a	0.998	0.997
b	0.998	0.997
c	0.997	0.998
d	0.997	0.996
e	0.997	0.995
f	0.996	0.998

Having recognized the Gaussian approximation, if the systems are truly ergodic we should have the same curvatures for both the initial and final states (equivalently having $\lambda_1^{var} = \lambda_2^{var}$ see section I.1). Eq. 4 translates this by stating that the difference between the two free energy curves equals the value of the free potential energy gap (the reaction coordinate), for any value of the reaction coordinate. The triangles on the graphs are obtained thus from application of Eq. 4. For example, for a given graph, all the points directly obtained from the simulations that generate the $G_1(\Delta E_{12})$ curve are shifted by (ΔE_{12}) to generate new points for the $G_2(\Delta E_{12})$ curve, by virtue of Eq. 4. Therefore the plain curves which are obtained by a linear regression fitting on both the circles and triangles are expected to match the dashed-dotted curves if the system is effectively ergodic. This is true for many systems (for example

proteins a, c, e), but can be somehow different for other proteins (b). This is in line with the small differences between λ_1^{var} and λ_2^{var} (Table 5) for proteins c, d, f, and more pronounced differences for proteins b for instance. *Does this mean this reduction is taking place beyond the ergodic approximation*? It must be remarked that it is difficult to test if the system is ergodic or not because it strongly depends on the capability of the MD methodology we have used to fully explore the conformation space. Apparent lack or ergodicity may simply be due to a lack of conformational sampling. It is hard to claim a non-ergodic system here. Instead these results illustrate the difficulties to reach complete sampling of the conformational space when running this kind of simulations.

Table 5. Outer sphere energies and reorganization energy of six heme proteins in eV. ΔG_{out}^0 calculated from Eq. 7. λ^{st} and λ^r are calculated from Eq.8 and 9 respectively.

Label	PDB	ligand	ΔG_{out}^0	λ^{st}	λ_1^{var}	λ_2^{var}	λ^r	U
a	1c2r	His-Met	0.581	0.763	0.721	0.842	0.766	0.006
b	351c	His-Met	0.570	0.826	0.667	0.882	0.828	0.005
c	1cyo	His-His	0.216	0.838	0.839	0.879	0.832	0.004
d	1eue	His-His	0.179	0.833	0.838	0.891	0.838	0.006
e	1b5a	His-His	0.199	0.828	0.863	1.048	0.828	0.005
f	1cxy	His-His	-0.077	0.822	0.889	0.808	0.825	0.005

ii) Stokes and Marcus reorganization energies

Assuming the lack of convergence between λ_1^{var} and λ_2^{var} is due to insufficient sampling and that the LRA approximation holds here, we may estimate ΔG_{out}^0 and λ^{st} and λ^M by equations 7, 8 and 9. (Table 5). The statistical uncertainties for ΔG_{out}^0 and λ^{st} are listed in the last column of Table 5. There are quiet small at around 0.005 eV. We observed that the ΔG_{out}^0 of Cyt.c are around 0.3 eV higher than the other four Cyt.b. The reorganization energies λ^{st} are almost the same at around 0.8 eV, except for the first Cyt.c with a lower reorganization energy at 0.763 eV. In most of the case $\lambda_1^{var} = \lambda_2^{var} = \lambda^{st} = \lambda^M$; for example in the case of c, d and f. In the case of a, b and e the differences are bigger at around 0.1 to 0.2 eV. This could come from not enough sampling of the potential energy surface.

We see clearly that the ΔG_{out}^0 of protein a and b are bigger than the others. The value of protein f is the smallest one and the energy is negative.

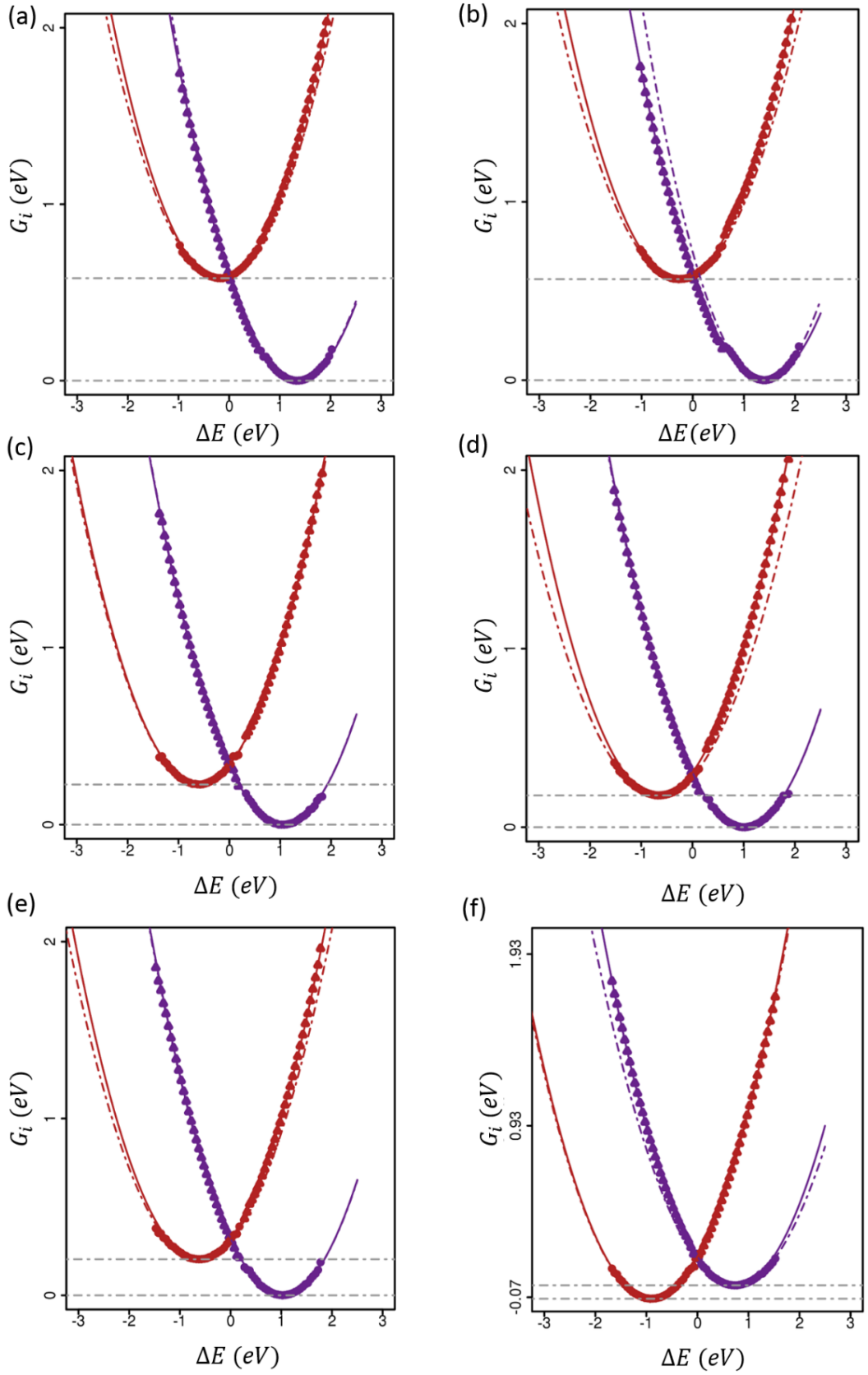


Figure 6. Diabatic free energy curves of Eq.3 for initial (purple) and final (red) states obtained from the energy gap calculation during 30 ns. The circles correspond to the points obtained directly from MD simulations. The triangles are obtained from the circles by application of Eq. 4, that is enforcing the ergodic hypothesis. The dashed-dotted curves correspond to linear fitting of a parabolic free energy curve on circles. The plain curves correspond to linear fitting of a parabolic free energy curves on the circles and triangles (i.e. enforcing ergodic hypothesis).

II.4 Analyses of computed Marcus theory parameters

II.4.a Comparison to experimental values

The result of simulated inner sphere energies (ΔE_{in}), outer sphere energies(ΔG^0_{out}) and theoretical redox potential (E^0_{th}) are summarized in Table 6. The theoretical redox potential is calculated as $E^0_{th} = (\Delta E_{in} + \Delta G^0_{out})/F$ .These results are compared with experimental results (E^0_{exp}), the difference between simulations and experiment are calculated (diff).

Table 6. Summary of inner sphere and outer sphere energies of six heme proteins as well as the theoretical and experimental redox potential.

Label	PDB	ligand	ΔE_{in} (eV)	ΔG^0_{out} (eV)	E^0_{th}(V)	E^0_{exp}(V)	diff(V)
a	1c2r	His-Met	5.304	0.581	5.886	4.790	1.096
b	351c	His-Met	5.304	0.570	5.874	4.710	1.164
c	1cyo	His-His	4.827	0.216	5.043	4.430	0.613
d	1eue	His-His	4.827	0.179	5.006	4.377	0.629
e	1b5a	His-His	4.827	0.199	5.026	4.338	0.688
f	1cxy	His-His	4.827	-0.077	4.750	4.230	0.520

Instead of obtaining the exact absolute values, we consider the relative differences of the potentials. The comparison curve between experimental and computed redox potential is plotted in Figure 7. The blue curve is the ideal result where simulation results equal to experimental results (y=x), while the read curve is the result we obtained. Overall, theoretical results reproduce well the trend of experimental results. However there is a large difference between Cyt.c (label a,b) and Cyt.b (label c,d,e and f) of around 0.5 V. This difference comes from both inner and outer sphere, especially from the inner part (0.46 eV). Restricting our attention to the Cyt. b. (the ones with the four lowest redox potentials) a linear regression factor R^2=0.8377 is obtained. The tendency is correct except the result of the second point. The differences with the experimental result of these two points (d and e) are quite small (39mV), a value which is just slightly higher than the uncertainty (±30mV).

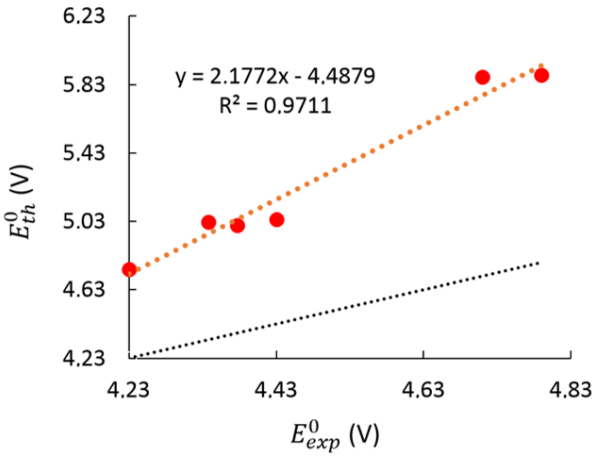


Figure 7. Comparison of experimental and computed redox potentials. The dashed line is the ideal result($y = x$). The red line is the plot with E^0_{exp} against E^0_{th}.

The presented method gives an overall agreement with experimental results. However, it is not accurate enough to distinguish a difference of 0.05 V. The difference between Cyt.c and Cyt.b may also be a problem of FF. Since in the inner sphere of Cyt.c there is a sulfur atom of the MET ligand, which is a highly polarizable atom, with a non-polarizable FF, one may miss important information.

II.4.b Energy decompositions at the level of amino acids

i) Free energies of reduction

In order to better understand why there is a big difference between redox potential of Cyt.c and Cyt b, we chose one cytochrome from each group, Cyt. c551 (PDB code: 351c) and Cyt. b5 (PDB code: 1b5a) to analyze the contribution of redox free energy ΔG^0_{out} and reorganization energy λ^{st} of each amino acid. The method is the same as we described in section I.3.b. Each energy is calculated as below:

$$\Delta E_{12} = E_{aa-Fe^{III}} - E_{aa-Fe^{II}} \quad (33)$$

$$\Delta G^0_{aa-Fe} = \frac{1}{2}(\langle \Delta E_{12} \rangle_1 + \langle \Delta E_{12} \rangle_2) \quad (34)$$

$$\Delta \lambda^{St}_{a-Fe} = \frac{1}{2}(\langle \Delta E_{12} \rangle_1 - \langle \Delta E_{12} \rangle_2) \quad (35)$$

where $aa - Fe^{III/II}$ indicates the interaction of each amino acid (aa) with the heme cofactor in different redox states. Results of free energy and reorganization energy are shown in Figure 8 and Figure 9 respectively.

In Figure8, the color corresponds to their contribution to ΔG^0_{out}. Blue is positive and red is negative. If the value is negative (red), it means that $E_{aa-Fe^{III}} < E_{aa-Fe^{II}}$. It favors the Fe^{III} state and vice versa. In both 1b5a and 351c, the amino acids in blue are arginine (ARG) and lysine(LYS) which have positive charge, while in red are aspartic acid (ASP) and glutamic acid (GLU) which are negatively charged. This is reasonable because the negative charge stabilized more the Fe^{III} state than the Fe^{II} state. The neutral amino acids do not have a big influence on ΔG^0_{aa-Fe}.

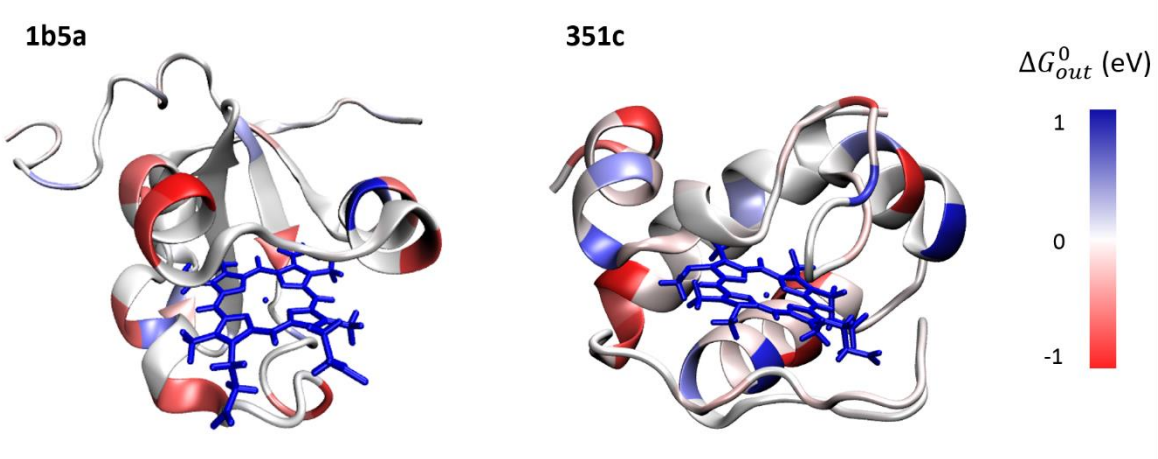


Figure 8. Structure of Cyt.c (351c) and Cyt.b (1b5a). The amino acids are colored according to their contribution to outer sphere energy.

The exact value of the ten most positive and negative free energy residues are summarized in Table 7. In the case of 1b5a, the highest contribution comes from ARG68 (0.895 eV), the lowest comes from GLU59 (-0.845 eV). For 351c, the highest comes from LYS33 (0.778 eV) and the lowest comes from GLU 70 (-0.592 eV). They are very close to the heme cofactor.

Table 7. Ranking of protein residues according to their contribution to free energy of 1b5a and 351c in eV.

Cytochromes	1b5a		351c	
Positive Ranking	Residue	ΔG^0_{out}	Residue	ΔG^0_{out}
1	ARG68	0.895	LYS33	0.778
2	ARG47	0.551	ARG44	0.739
3	LYS72	0.499	LYS21	0.670
4	LYS34	0.408	LYS10	0.595
5	LYS89	0.389	LYS49	0.473
6	LYS16	0.373	LYS8	0.430
7	LYS28	0.360	LYS28	0.427
8	ARG19	0.208	LYS76	0.360
9	ARG84	0.202	GLY24	0.163
10	LYS2	0.196	GLY51	0.123
Negative Ranking				
-1	GLU59	-0.846	GLU70	-0.592
-2	GLU38	-0.731	GLU43	-0.557
-3	ASP60	-0.719	ASP29	-0.549
-4	ASP66	-0.694	GLU4	-0.497
-5	GLU43	-0.632	ASP68	-0.476
-6	GLU44	-0.603	GLU41	-0.442
-7	ASP31	-0.567	ASP19	-0.440
-8	GLU48	-0.549	ASP69	-0.376
-9	ASP53	-0.542	ASP29	-0.337
-10	GLU56	-0.522	PRO62	-0.134

$$\Delta E_{12} = \sum_{i}^{aa} \sum_{j}^{heme} \frac{q_i(q_j^{Fe^{III}} - q_j^{Fe^{II}})}{r_{ij}} \quad (36)$$

As shown in Eq. 36, the energy is coming from Coulomb interaction between charges on atoms of each aa (q_i) and charges on atoms of heme cofactor of different states (q_j). As the distance between the charges of amino acid and heme increases, the contribution decreases. Thus the closer the amino acid is to the heme center, the stronger its contribution. Therefore, free energy contributions coming from interactions between protein and heme cofactors depend on how many charged amino acids in the protein and their distance to the heme center. As shown in Figure 2, protein of 1b5a forms αβ, while 351c contains only α[‡]. They have very different shapes. In addition, the heme cofactor of 1b5a is more accessible to water molecules while for 351c, the heme cofactor is buried in the protein center. The free energy comes from interaction of water and the heme cofactor could be a source of this difference too. We see that the accuracy of this interaction is highly dependent on the accuracy of the evaluation of Coulomb interactions between proteins and Heme cofactors.

ii) Stokes reorganization energies

The same analyzes have been done for the reorganization energy. This time the color bar is 10 time smaller than for the free energy. This is obtained from the difference of the energy gap of each amino acide (Eq. 31).

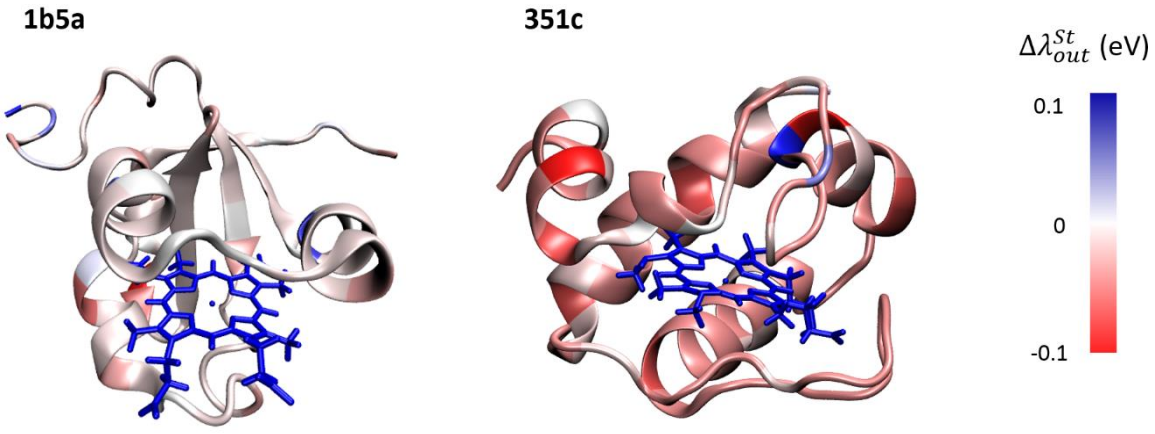


Figure 9. Structure of Cyt. c and Cyt, b. The amino acids are colored according to their contribution to reorganization energy.

The residues giving the ten most positive and negative reorganization energies are listed in Table 8. For the 1b5a contribution, the highest value comes from LEU94 (0.059 eV) which is the last amino acid at the end of the protein. The lowest comes from LYS16 (-0.037 eV). For 351c, the highest comes from

[‡] α means helices, while β means β sheets.

TYR27 (0.061) and the lowest comes from LYS28 (-0.016eV). This is different than in the case of the redox free energy, although most of the important contributions of the reorganization energy are coming from charged residues, but there are some neutral aa which give important contributions too. For example, LEU94 which is a neutral amino acid gives the highest contribution. This could be due to the great flexibility of these residues and we haven't sampled all possible conformations. As shown in Figure 5(e), the RMSD of these residues are quite large especially for the Fe^{III} state (5 Å). More simulations are needed. This could also explain the result in Table 5 for which λ_2^{var} of 1b5a is larger by c.a. 0.2 eV than λ_1^{var}.

Table 8. Ranking of protein residues according to their contribution to reorganization energy of 1b5a and 351c in eV

Cytochromes	1b5a		351c	
Positive Ranking	Residue	$\boldsymbol{\lambda^{st}}$	Residue	$\boldsymbol{\lambda^{st}}$
1	LEU94	0.059	TYR27	0.061
2	SER71	0.048	LYS21	0.041
3	GLU92	0.040	LYS82	0.028
4	ASP53	0.037	GLU4	0.022
5	LYS34	0.022	CYS12	0.017
6	LYS89	0.020	CYS15	0.016
7	GLN49	0.018	GLU70	0.013
8	ASP3	0.017	ASP29	0.013
9	LYS14	0.011	ASN64	0.012
10	LYS72	0.011	HIS16	0.012
Negative Ranking				
-1	LYS16	-0.037	LYS28	-0.016
-2	ASP17	-0.022	LYS8	-0.016
-3	GLU48	-0.021	LYS33	-0.013
-4	ASP31	-0.014	LYS49	-0.011
-5	ASP17	-0.011	ASP69	-0.010
-6	GLU44	-0.011	LYS10	-0.007
-7	HIS26	-0.010	LYS76	-0.007
-8	GLU38	-0.009	ASP2	-0.002
-9	GLU11	-0.009	GLY54	-0.001
-10	LYS86	-0.006	GLN37	-0.001

In order to analyze the differences coming from the ligands which are linked with iron, their contributions to free energy and organization energy are summarized in Table 9.

We observe that their contributions are not that high, out of the first-ten ranking. The two HIS in the case of 1b5a give very similar contribution to ΔG_{out}^{0} which is 0.1 eV bigger than the HIS contribution in 351c. The two CYS connected to heme cofactors give negative contributions to the free energy, which means that they help to stabilize the Fe^{III} state. A recent publication has shown that the insertion of cysteine linkages accelerates electron flow through tetra-heme protein[17b]. The contribution of

reorganization energy of two HIS in 351c is quite small, less than 0.1 eV. They are quite rigid due to the fact that they are linked to the heme center.

Table 9. Ranking of ligands of 1b5a and 351c according to their contribution to free energy and reorganization energy in eV

	1b5a			351c	
Ranking	ligands	ΔG^{0}_{out}	Ranking	ligands	ΔG^{0}_{out}
15	HIS39	0.134	16	HIS16	0.019
14	HIS36	0.134	61	MET61	0.111
			12	CYS12	-0.064
			15	CYS15	-0.004
Ranking	ligands	λ^{st}	Ranking	ligands	λ^{st}
37	HIS39	0.002	33	HIS16	0.012
13	HIS36	0.007	11	MET61	0.002
			66	CYS12	0.017
			43	CYS15	0.016

CONCLUSION

In this Chapter, we have introduced theory and computational set up for simulating redox potentials of heme proteins within the Linear Response Approximation. Six cytochromes with experimental redox potential were chosen as model systems. We simulated the redox free energy in the framework of Marcus theory. The QM+MM method was applied to evaluate the inner sphere and outer sphere contributions. DFT computations were performed and we observed a large difference of around 0.5 eV for two different inner sphere energies for Cyt.c and Cyt.b. Outer sphere energies were obtained by sampling 30 ns MD simulations of different states performed with a non-polarizable FF. Diabatic free energy curves were plotted and we showed that the LRA is valid in most cases. There are some differences in certain cases which may be due to lack of conformational sampling.

Compared to experimental results, our simulations reproduce the global trends. However, the energy shifts between Cyt.c and Cyt.b are quite different (by around 0.5 V). This difference could come from the difference in the inner sphere structures which are quite different. It could also come from the outer-sphere in which there is a highly polarizable sulfur atom in the inner sphere. However, with the non-polarizable point charge model, one cannot describe the electrostatic interaction correctly. The four Cyt.b reproduce well experimental results with a quite good correlation. However, a difference of around 50 mV is out of reach of the accuracy of the presented method.

With this QM+MM method, we are able to analyze the contributions of the redox free energy by each amino acid. It is shown that the main contributions come from charged amino acids which are close to the heme cofactor. The accuracy of these simulations is highly dependent on the electrostatic interaction description between heme and environment. The next step will be to test more

sophisticated polarizable FFs and we expect to be able to improve the reliability of this simulation method.

REFERENCES

1. Warshel, A., Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *J. Phy. Chem.* **1982,** *86* (12), 2218.
2. Tachiya, M., Relation between the electron-transfer rate and the free energy change of reaction. *J. Phy. Chem.* **1989,** *93* (20), 7050.
3. Carter, E. A.; Hynes, J. T., Solute-dependent solvent force constants for ion pairs and neutral pairs in a polar solvent. *J. Phy. Chem.* **1989,** *93* (6), 2184.
4. (a) Gunner, M. R.; Honig, B., Electrostatic control of midpoint potentials in the cytochrome subunit of the Rhodopseudomonas viridis reaction center. *Proceedings of the National Academy of Sciences.* **1991,** *88* (20), 9151; (b) Ishikita, H.; Loll, B.; Biesiadka, J.; Saenger, W.; Knapp, E.-W., Redox Potentials of Chlorophylls in the Photosystem II Reaction Center. *Biochemistry.* **2005,** *44* (10), 4118.
5. Mao, J.; Hauser, K.; Gunner, M. R., How Cytochromes with Different Folds Control Heme Redox Potentials. *Biochemistry.* **2003,** *42* (33), 9829.
6. Warshel, A.; Papazyan, A.; Muegge, I., Microscopic and semimacroscopic redox calculations: what can and cannot be learned from continuum models. *J. Bio. Inorg. Chem.* **1997,** *2* (1), 143.
7. Churg, A. K.; Weiss, R. M.; Warshel, A.; Takano, T., On the Action of Cytochrome-C - Correlating Geometry Changes Upon Oxidation with Activation-Energies of Electron-Transfer. *J. Phys. Chem-Us.* **1983,** *87* (10), 1683.
8. Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A., Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins.* **2000,** *39* (4), 393.
9. Oliveira, A. S. F.; Teixeira, V. H.; Baptista, A. M.; Soares, C. M., Reorganization and conformational changes in the reduction of tetraheme cytochromes. *Biophys. J.* **2005,** *89* (6), 3919.
10. Bhattacharyya, S.; Stankovich, M. T.; Truhlar, D. G.; Gao, J. L., Combined quantum mechanical and molecular mechanical simulations of one- and two-electron reduction potentials of flavin cofactor in water, medium-chain acyl-CoA dehydrogenase, and cholesterol oxidase. *J. Phys. Chem. A.* **2007,** *111* (26), 5729.
11. de la Lande, A.; Gillet, N.; Chen, S.; Salahub, D. R., Progress and challenges in simulating and understanding electron transfer in proteins. *Arch. Biochem. Biophys.* **2015,** *582*, 28.
12. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227.
13. Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Phys. Rev.* **1964,** *136*, B864.
14. Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140*, A1133.
15. LeBard, D. N.; Kapko, V.; Matyushov, D. V., Energetics and Kinetics of Primary Charge Separation in Bacterial Photosynthesis. *J. Phys. Chem. B* **2008,** *112* (33), 10322.
16. (a) Gillet, N.; Lévy, B.; Moliner, V.; Demachy, I.; de la Lande, A., Electron and Hydrogen Atom Transfers in the Hydride Carrier Protein EmoB. *J. Chem. Theory Comput.* **2014,** *10* (11), 5036; (b) Blumberger, J., Free energies for biological electron transfer from QM/MM calculation: method, application and critical assessment. *Phys. Chem. Chem. Phys.* **2008,** *10* (37), 5651.
17. (a) Tipmanee, V.; Blumberger, J., Kinetics of the Terminal Electron Transfer Step in Cytochrome c Oxidase. *J. Phy. Chem. B* **2012,** *116* (6), 1876; (b) Jiang, X.; Futera, Z.; Ali, M. E.; Gajdos, F.; von Rudorff, G. F.; Carof, A.; Breuer, M.; Blumberger, J., Cysteine Linkages Accelerate Electron Flow through Tetra-Heme Protein STC. *J. Am. Chem. Soc.* **2017,** *139* (48), 17237; (c) Breuer, M.; Zarzycki, P.; Blumberger, J.;

Rosso, K. M., Thermodynamics of Electron Flow in the Bacterial Deca-heme Cytochrome MtrF. *J. Am. Chem. Soc.* **2012,** *134* (24), 9868.
18. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267*, 585.
19. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179.
20. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **1998,** *102* (18), 3586.
21. Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P., Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008,** *10* (19), 2747.
22. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225.
23. Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B., High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *Phys. Chem. Lett.* **2011,** *2* (2), 87.
24. Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Current Status of Protein Force Fields for Molecular Dynamics. *Methods in molecular biology (Clifton, N.J.)* **2015,** *1215*, 47.
25. (a) D., M. A., Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **2004,** *25* (13), 1584; (b) Guvench, O.; MacKerell, A. D., Comparison of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins*, Kukol, A., Ed. Humana Press: Totowa, NJ, **2008**; pp 63.
26. (a) Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications. *Theor. chem. acc.* **2009,** *124* (1-2), 11; (b) Warshel, A.; Kato, M.; Pisliakov, A. V., Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007,** *3* (6), 2034.
27. (a) Blumberger, J.; Lamoureux, G., Reorganization free energies and quantum corrections for a model electron self-exchange reaction: comparison of polarizable and non-polarizable solvent models. *Mol. Phy.* **2008,** *106* (12-13), 1597; (b) Sterpone, F.; Ceccarelli, M.; Marchi, M., Linear Response and Electron Transfer in Complex Biomolecular Systems and a Reaction Center Protein. *J. Phy. Chem. B* **2003,** *107* (40), 11208; (c) King, G.; Warshel, A., Investigation of the free energy functions for electron transfer reactions. *J. Chem. Phys.* **1990,** *93* (12), 8682; (d) Ceccarelli, M.; Marchi, M., Simulation and Modeling of the Rhodobacter sphaeroides Bacterial Reaction Center II: Primary Charge Separation. *J. Phy. Chem. B.* **2003,** *107* (23), 5630; (e) Ungar, L. W.; Newton, M. D.; Voth, G. A., Classical and Quantum Simulation of Electron Transfer Through a Polypeptide. *J. Phy. Chem. B.* **1999,** *103* (34), 7367.
28. (a) Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986,** *108* (15), 4315; (b) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. J., Fluctuating Charge Force-Fields for Aqueous-Solutions. *J. Mol. Liq.* **1995,** *65-6*, 31; (c) Rappe, A. K.; Goddard, W. A., Charge Equilibration for Molecular-Dynamics Simulations. *J. Phys. Chem-Us.* **1991,** *95* (8), 3358; (d) Rick, S. W.; Stuart, S. J.; Berne, B. J., Dynamical Fluctuating Charge Force-Fields - Application to Liquid Water. *J. Chem. Phys.* **1994,** *101* (7), 6141; (e) Sanderson, R. T., An Interpretation of Bond Lengths and a Classification of Bonds. *Science.* **1951,** *114* (2973), 670.
29. Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **1986,** *108* (15), 4315.
30. (a) Patel, S.; Brooks, C. L., Development and application of a CHARMM fluctuating charge protein force field. *Abstr. Pap. Am. Chem. S.* **2004,** *227*, U1010; (b) Patel, S.; Brooks, C. L., Fluctuating

charge force field for methanol: Liquid-vapor interfacial properties. *Abstr Pap Am Chem S* **2004,** *227*, U319; (c) Patel, S.; Brooks, C. L., CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004,** *25* (1), 1; (d) Patel, S.; Mackerell, A. D.; Brooks, C. L., CHARMM fluctuating charge force field for proteins: II - Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004,** *25* (12), 1504.

31. (a) Bauer, B. A.; Patel, S., Recent applications and developments of charge equilibration force fields for modeling dynamical charges in classical molecular dynamics simulations. *Theor. chem. acc.* **2012,** *131* (3); (b) Lucas, T. R.; Bauer, B. A.; Patel, S., Charge equilibration force fields for molecular dynamics simulations of lipids, bilayers, and integral membrane protein systems. *Biochimica et biophysica acta.* **2012,** *1818* (2), 318.

32. (a) van Maaren, P. J.; van der Spoel, D., Molecular dynamics simulations of water with novel shell-model potentials. *J. Phy. Chem. B.* **2001,** *105* (13), 2618; (b) Lamoureux, G.; Roux, B. t., Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003,** *119* (6), 3025.

33. Drude, P.; Millikan, R. A.; Mann, R. C. The Theory of Optics; Longmans, Green, and Co.: New York, **1902**.

34. Kunz, A. P. E.; van Gunsteren, W. F., Development of a Nonlinear Classical Polarization Model for Liquid Water and Aqueous Solutions: COS/D. *J. Phys. Chem. A.* **2009,** *113* (43), 11570.

35. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., Jr., An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016,** *116* (9), 4983.

36. Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D., Jr., Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013,** *9* (12), 5430.

37. Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B., Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2006,** *2* (6), 1587.

38. Thole, B. T., Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981,** *59* (3), 341.

39. Silberstein, L., Dispersion and the size of molecules of hydrogen, oxygen, and nitrogen. *Philos. Mag. Series 6* **1917,** *33*, 215.

40. Applequist, J.; Carl, J. R.; Fung, K. K., Atom Dipole Interaction Model for Molecular Polarizability - Application to Polyatomic-Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972,** *94* (9), 2952.

41. Piotr, C.; James, C.; Peter, K., Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N‐methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001,** *22* (10), 1048.

42. (a) Wang, W.; Skeel, R. D., Fast evaluation of polarizable forces. *J Chem Phys* **2005,** *123* (16); (b) Aviat, F.; Levitt, A.; Stamm, B.; Maday, Y.; Ren, P.; Ponder, J. W.; Lagardère, L.; Piquemal, J.-P., Truncated Conjugate Gradient: An Optimal Strategy for the Analytical Evaluation of the Many-Body Polarization Energy and Forces in Molecular Simulations. *J. Chem. Theory Comput.* **2017,** *13* (1), 180.

43. Applequist, J.; Carl, J. R.; Fung, K.-K., Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972,** *94* (9), 2952.

44. Tipmanee, V.; Oberhofer, H.; Park, M.; Kim, K. S.; Blumberger, J., Prediction of Reorganization Free Energies for Biological Electron Transfer: A Comparative Study of Ru-Modified Cytochromes and a 4-Helix Bundle Protein. *J. Am. Chem. Soc.* **2010,** *132* (47), 17032.

45. (a) Vigné‐Maeder, F.; Claverie, P., The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *J. Chem. Phys.* **1988,** *88* (8), 4934; (b) 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* (5), 1968; (c) E., W. D., Representation of the molecular electrostatic potential by atomic multipole and bond dipole models. *J. Comput. Chem.* **1988,** *9* (7), 745.
46. (a) Ren, P. Y.; Ponder, J. W., Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002,** *23* (16), 1497; (b) Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B.* **2003,** *107* (24), 5933.
47. (a) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P., Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.* **2007,** *3* (6), 1960; (b) Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C., Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A.* **2003,** *107* (48), 10353.
48. Hermida-Ramon, J. M.; Brdarski, S.; Karlstrom, G.; Berg, U., Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003,** *24* (2), 161.
49. H. M. Berman, J. W., Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne., The Protein Data Bank *Nucleic Acids Res.* **2000,** *28*, 235.
50. Zhong, Z.; Gunner, R. M., Analysis of the electrochemistry of hemes with Ems spanning 800 mV. *Proteins: Struc., Func., and Bioinf.* **2009,** *75* (3), 719.
51. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30* (10), 1545.
52. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926.
53. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32* (Web Server issue), W665.
54. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8* (9), 3257.
55. (a) Felix, A.; Emad, T.; Jerome, B.; Zaida, L. S., Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput.l Chem.* **2004,** *25* (13), 1613; (b) Northrup, S. H.; Pear, M. R.; Morgan, J. D.; McCammon, J. A.; Karplus, M., Molecular dynamics of ferrocytochrome c: Magnitude and anisotropy of atomic displacements. *J. Mol. Biol.* **1981,** *153* (4), 1087.
56. A.M. Koster, G. G., A. Alvarez-Ibarra, P. Calaminici, M.E. Casida, J. Carmona-Espindola, V.D. Dominguez, R. Flores-Moreno, G.U. Gamboa, A. Goursot, T. Heine, A. Ipatov, A. de la Lande, F. Janetzko, J.M. del Campo, D. Mejia-Rodriguez, J. U. Reveles, J. Vasquez-Perez, A. Vela, B. Zuniga-Gutierrez, and D.R. Salahub, deMon2k, Version 4, The deMon developers, Cinvestav, Mexico City **2016**.
57. (a) Handy, N. C.; Cohen, A. J., Left-right correlation energy. *Molecular Physics* **2001,** *99* (5), 403-412; (b) Swart, M.; Ehlers, A. W.; Lammertsma, K., Performance of the OPBE exchange-correlation functional. *Mol. Phys.* **2004,** *102* (23-24), 2467.
58. (a) Han, W.-G.; Noodleman, L., Structural Model Studies for the Peroxo Intermediate P and the Reaction Pathway from P → Q of Methane Monooxygenase Using Broken-Symmetry Density Functional Calculations. *Inorg. Chem.* **2008,** *47* (8), 2975; (b) Swart, M., Accurate Spin-State Energies for Iron Complexes. *J. Chem. Theory Comput.* **2008,** *4* (12), 2057; (c) Conradie, M. M.; Conradie, J.; Ghosh, A., Capturing the spin state diversity of iron(III)-aryl porphyrins: OLYP is better than TPSSh. *J. Inorg. Biochem.* **2011,** *105* (1), 84.

59. Calaminici, P.; Janetzko, F.; Köster, A. M.; Mejia-Olvera, R.; Zuniga-Gutierrez, B., Density functional theory optimized basis sets for gradient corrected functionals: 3d transition metal systems. *J. Chem. Phys.* **2007,** *126* (4), 044108.
60. (a) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R., Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007,** *126* (14), 144111; (b) de la Lande, A.; Clavaguéra, C.; Köster, A., On the accuracy of population analyses based on fitted densities. *J. Mol. Model.* **2017,** *23* (4), 99.
61. C., P. J.; Rosemary, B.; Wei, W.; James, G.; Emad, T.; Elizabeth, V.; Christophe, C.; D., S. R.; Laxmikant, K.; Klaus, S., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781.
62. (a) Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092; (b) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577.
63. Team, R. C., R: A language and environment for statistical computing. . *R Foundation for Statistical Computing, Vienna, Austria.* **2013**.
64. Plummer, M. B., Nicky; Cowles, Kate and Vines, Karen (2006). CODA: convergence diagnosis and output analysis for MCMC. R News, 6(1) pp. 7.

CHAPTER 2

Beyond the Point Charge Description

AMOEBA Polarizable Force Field

Beyond the Point Charge Description
AMOEBA Polarizable Force Field

Much effort has been focused on including the response of the molecular environment by an explicit representation of polarization. The quality of the description of the permanent moment distribution is also very important to reproduce the global electrostatic interactions. This approach is inherently limited because point charges are intrinsically isotropic while real atoms are anisotropic in the electron distribution due to the presence of lone pairs, π clouds, and σ holes. In some FFs extra point charges are added at different positions than atom nuclei. This allow the simulation of lone pairs[1] or centers of bonds[2]. An alternative is to compute the electrostatic potential created by a molecule from multipole distributions including not only monopoles (point charges) but also dipoles, quadrupoles or even higher-order terms. With these multipole moments, the non-spherical components of atomic charge density are better described and the anisotropy of the system is naturally captured. This can describe more accurately the flexibility of the interactions which is crucial especially around polar molecules. Several FFs have multipole moment descriptions such as AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications)[3], SIBFA (Sum of Interactions Between Fragments *Ab Initio* Computed)[2, 4] or NEMO (Non-Empirical Molecular Orbital)[5].

Among them, AMOEBA is the most widely used FF because of its reasonable computational cost. It has been historically developed by Ponder, Ren and co-workers[3, 6]. AMOEBA was first implemented in the TINKER modeling package[7]. Recently, a parallelized version was implemented in OpenMM[8] using Open Multi-Processing (openMP) and recently updated to Graphics Processing Unit (GPU) using the CUDA programming language. Another highly parallelized version named TINKER-HP[9] using Message Passing Interface (MPI) was developed very recently by the group of J. -P. Piquemal (Sorbonne University Paris). These algorithmic developments open the door toward the application of AMOEBA to the simulation of large biomolecular systems, in particular in the context of the present PhD thesis, to the simulation of redox properties of proteins.

In the past decade, AMOEBA parameters of many molecules have been developed including water[3b, 10], ions[11], small molecules[12], biomolecules (peptides[13], protein[14], lipid[15], nucleic acids[16]). Many applications of these parameters have been reported such as the computation of ion hydration free energy[11a, 11d, e, 16-17], the structures and the thermodynamics of organic molecules[18], and protein-ligand binding[19]. These applications have demonstrated that AMOEBA is able to perform well both in gas and solution phases for various chemical properties. Based on these relevant results, we considered that

AMOEBA could be a good potential choice to simulate redox potentials of heme proteins. Toward this objective an unavoidable step is to have access to force field parameters for the heme. However, to the best of our knowledge such parameters were not available in the literature at the beginning of this PhD work. Therefore a key step of this PhD work was to create a reliable set of parameters for the heme cofactor in both its ferrous and ferric forms.

In Section I, we introduce the formalism of AMOEBA. We then describe in Section II the standard parametrization procedure of AMOEBA that we followed for the heme cofactor. Finally in Section III, a published article in the *Journal of Chemical Theory and Computation*[20] describing the parameterization work is reproduced.

I. AMOEBA FORMALISM

The AMOEBA force field has the following general functional form for the interactions among atoms.

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{opp} + U_{torsion} + U_{vdW} + U_{ele}^{per} + U_{ele}^{ind} \quad (1)$$

The first five terms describe the short-range valence interactions bond stretching (U_{bond}), angle bending (U_{angle}) bond-angle cross term ($U_{b\theta}$), out-of-plane bending (U_{opp}), and torsional rotation ($U_{torsion}$) while the last three terms are respectively the non-bonded van der Waals (U_{vdW}) and electrostatic contributions including interaction among permanent multipoles (U_{ele}^{per}) and electronic induction (U_{ele}^{ind}).

I.1 Short-range valence interactions

The AMOEBA potential includes full intramolecular flexibility, which is important to explore different conformations of large and flexible molecules. The functional forms for bond stretching and angle bending were adapted from the MM3 force field[21] and include anharmonicity through the use of higher-order deviations from ideal bond lengths and angles.

The potential for bond stretching takes the form of Taylor serial development at 4th order of the Morse potential, with K_b being the bond force constant, *b* the both length and b_0 the reference value.

$$U_{bond} = K_b(b + b_0)^2 \left[1 - 2.55(b - b_0) + \left(\frac{7}{12}\right)2.55(b - b_0)^2\right] \quad (2)$$

This potential function mimics the anharmonic shape of a bond stretching. According to the authors of MM3, the constant of order 3 is sufficient in the vast majority of cases to bring in the effect

anharmonicity, the term of order 4 being justified only as a safeguard to avoid problematic situations of dissociative states when the structure encountered is far from standard distances.

The potential for angle bending is described by a potential at 6th order, with $\boldsymbol{\theta}$ this angle and θ_0 its reference value.

$$U_{angle} = K_\theta(\theta + \theta_0)^2\,[1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5}(\theta - \theta_0)^2 - 7. \times 10^{-7}(\theta - \theta_0)^3 + 2.2 \times 10^{-8}(\theta - \theta_0)^4] \quad (3)$$

The coupling between bond and angle deformation takes the form:

$$U_{b\theta} = K_{b\theta}\,[(b - b_0) + (b' - b_0')](\theta - \theta_0) \quad (4)$$

with $K_{b\theta}$ the force constant, $(\theta - \theta_0)$ the angle deformation, $(b - b_0)$ and $(b' - b_0')$ the bond stretching of the two bonds forming the angle considered.

A Wilson-Decius-Cross function[22] is used at sp^2-hybridized trigonal centers to restrain the out-of-plane bending, where χ is the dihedral angle formed by the four sequentially bonded atoms.

$$U_{oop} = K_x\,x^2 \quad (5)$$

The torsion energy is described by a traditional Fourier expansion (a 1-fold through 6-fold trigonometric form) torsional functional. It is used to aid in merging the short-range "valence" terms with the long-range "non-bonded" interactions.

$$U_{torsion} = \sum_n K_{n\phi}\,[1 + cos(n\phi \pm \delta)] \quad (6)$$

ϕ is the dihedral angle computed from the p-orbital directions at the two trigonal centers, rather than from the usual bond vectors. It allows appropriately to increase the flexibility of atoms bonded to trigonal centers (e.g., aromatic hydrogen atoms). The torsional parameters may be refined after the determination of the non-bonded parameters with the hope that the improved AMOEBA intramolecular electrostatic model will lead to a more physical balance between the local and long-range interactions in the potential energy.

I.2 Long-range nonbonded interactions

The pairwise additive van der Waals interaction in AMOEBA adopts the buffered 14-7 functional form proposed by Halgren[23].

$$U_{vdW} = \varepsilon_{ij}\left(\frac{1.07}{\rho_{ij} + 0.07}\right)^7\left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2\right) \quad (7)$$

The potential is a function of the separation distance R_{ij} between the atoms i and j through the term

$$\rho_{ij} = \frac{R_{ij}}{R_{ij}^0} \tag{8}$$

R_{ij}^0 is the minimum energy distance. For heterogeneous atom pairs, $R_{ij}^0 = \frac{\left(R_{ii}^0\right)^3 + \left(R_{jj}^0\right)^3}{\left(R_{ii}^0\right)^2 + \left(R_{jj}^0\right)^2}$, the well-depth is given by $\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{\left(\varepsilon_{ii}^{1/2} + \varepsilon_{jj}^{1/2}\right)^2}$.

The 14-7 potential was shown to be more effective than the 12-6 Lennard-Jones (LJ) potential to reproduce interaction energies of rare gases[23]. By comparison with quantum chemical calculations, Halgren showed that the repulsive part in r^{-12} of the LJ leads to a repulsive potential that is too soft. It can be improved by the shape in r^{-14} of the buffered 14-7. The AMOEBA van der Waals parameters are usually derived by comparison with quantum mechanical interaction energies in gas phase and they are often refined using condensed phase simulations. Each atom in AMOEBA possesses a vdW site. For non-hydrogen atoms, the site is located at the position of the atomic nucleus. For a hydrogen atom connected to an atom X, the vdW site is shifted from the hydrogen center to atom X according to the percentage of the full bond length[24]. It was found to simultaneously improve the fit to accurate QM water dimer structures and energies for several configurations[21].

The electrostatic energy is calculated using derived multipoles composed of charge $q_{i,}$ dipoles $\mu_{i\alpha}$, and quadrupoles $Q_{i\beta\gamma}$.

$$M_i = \left[q_{i,}\, \mu_{i,x}, \mu_{i,y}, \mu_{i,z}, Q_{ixx}, Q_{ixy}, Q_{ixz,}\, , Q_{iyx}, Q_{iyy}, Q_{iyz,} Q_{izx}, Q_{izy}, Q_{izz,}\right]^t \tag{9}$$

There are typically five independent quadrupole components due to symmetry ($Q_{\alpha\beta} = Q_{\beta\alpha}$). Usually, the FF uses traceless quadrupoles ($\sum Q_{\alpha\alpha} = 0$). Furthermore, the μ_y, Q_{xy} and Q_{yz} components are zero except for chiral atoms such as the backbone C in amino acids. Therefore, most non-chiral atoms M_i^{nc} carry six unique, permanent electrostatic multipole parameters:

$$M_i^{nc} = \left[q_{i,}\, \mu_{i,x}, \mu_{i,z}, Q_{ixx}, Q_{iyy}, Q_{izx}, Q_{izz,}\right]^t \tag{10}$$

The interaction energy named U_{ele}^{per} between two multipole sites i and j is written as:

$$U_{ele}^{per} = \begin{bmatrix} q_i \\ \mu_{ix} \\ \mu_{iy} \\ \mu_{iz} \\ Q_{ixx} \\ \vdots \end{bmatrix}^T \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial x_j} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial x_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \frac{1}{R_{ij}} \begin{bmatrix} q_j \\ \mu_{jx} \\ \mu_{jy} \\ \mu_{jz} \\ Q_{jxx} \\ \vdots \end{bmatrix} \tag{11}$$

Electronic induction describes the polarization of electron density due to an external field. The induced dipole model is used by AMOEBA at atomic centers to describe many-body effects. The induced dipole at each atomic site is computed as $\mu_{i,a} = \alpha_i F_{i,a}$, where α_i is the atomic isotropic polarizability of atom i. a denotes the three Cartesian components. $\boldsymbol{F_i}$ is the electric field experienced by atom i. The iterative model is used, thus the induced dipole on each atom will further polarize all other atoms both within and outside the molecule such that $\boldsymbol{F}$ becomes the sum of the fields generated by both permanent multipoles and induced dipoles at sites other than atom i.

$$\mu_{i,a} = \alpha_i \left(\sum_{\{j\}} T_a^{ij} M_j + \sum_{\{j'\}} T_{ab}^{ij'} \mu_{j',b}^{ind} \right) \qquad \text{for } a, b = \{x, y, z\} \tag{12}$$

where $M_j = \left[q_{j,} \mu_{j,x}, \mu_{j,y}, \mu_{j,z}, Q_{j,xx}, Q_{j,xy}, Q_{j,xz,\cdots,} Q_{j,zz} \right]^t$ is the multipole distribution already described on atom center j. $T_a^{ij} = \left[T_a, T_{ax}, T_{ay}, T_{az}, \dots \right]$ is the interaction tensor between sites i and j.

The polarizability α_i for carbon, hydrogen, nitrogen and oxygen in AMOEBA are adopted from the Thole model[25]. The others such as aromatic carbon and hydrogen atoms and ions are usually computed from QM.

The above equations have to be solved iteratively for all atomic sites of the system at the same time. The damped induction approach initially developed by Thole is adopted in order to avoid the divergence of the polarization energy at short distances[25]. The damping is effectively achieved by smearing the atomic multipole moments in each pair of interaction sites. The smearing function for charges adopted by AMOEBA has the following functional form:

$$\rho = \frac{3\alpha}{4\pi} exp(-a\mu^3) \tag{13}$$

where $\mu = \frac{R_{ij}}{(\alpha_i \alpha_j)^{\frac{1}{6}}}$ is the effective distance as a function of linear separation R_{ij} and $\alpha_{i,} \alpha_j$ are atomic polarizabilities of sites i and j. The factor "a" is a dimensionless width parameter of the smeared charge distribution, and effectively controls the damping strength. In the original model of Thole, "a"

was chosen to be 0.572 for all atoms. In the first version of the AMOEBA water parameters, “a” was chosen to be 0.39 after fitting the interaction energies of a series of small water clusters. This value is used in AMOEBA for all atoms except for multiply-charged cations[11b-e, 17a]. In these cases, it is necessary to use stronger damping (a < 0.39) to better represent the electric field around the ions. Corresponding damping functions for charge, dipole, and quadrupole interactions were derived through their chain rule relationships[3b].

Many-body effects consist of both polarization and charge transfer (CT). SIBFA[2, 4] for instance accounts for CT terms, but CT is not explicitly taken into account in advanced FF. Usually, researchers consider CT to be a secondary effect that can be indirectly included in vdW and electrostatics parameters. But sometimes the lack of CT may be problematic, especially for transition-metal systems. In these cases, special care must be taken in the parameterization and the validation procedure. Another way to include full many-body effects is to use a model in which both charges and dipoles are fluctuating such as in the work of Friesner and Berne[26]. Furthermore when two atoms are close enough, their charge densities can overlap. As a consequence, the screening of the charge of a nucleus by its own electronic density decreases. This interaction is called 'electrostatic penetration' or 'charge penetration'. Different strategies have been proposed to include this effect in molecular modeling[27]. An optimized model has been added into the AMOEBA force field[28].

II. STANDARD PARAMETRIZATION STRATEGY

One of the challenges of advanced force fields is the development of parameter sets which is made complicated with AMOEBA owing to the complicated mathematical form of the potential energy function. In this Section, we describe the standard parameterization of the AMOEBA force field.

II.1 Define atom types and classes

In AMOEBA, parameters are classified by using different atom types and atom classes. Atoms with the same atom class will share parameters for short-range valence interactions (bond, angle, torsions ...) and the same vdW parameters. Within an atom class, atoms are further classified by types according to their chemical environment and consequently may have different parameters for multipole distribution and atomic polarizability. In the parameter set called AMOEBApro13[29], there are 42 atom classes and 258 atom types. The first step to create parameters of a new molecule is to define the atom type and class of each atom of the new fragment we want to add to the FF. Usually they are derived from a similar structure whose parameters exist already. However in some cases, one should

modify the parameters, for example to adjust the torsion parameters according to QM results[30]. Moreover, new atom types are sometimes necessary to describe diverse electrostatic interactions. A key step is to obtain an accurate multipole distribution of the molecule.

II.2 Extraction of atomic multipoles

Electrostatic multipoles are located on atomic centers and defined in the local frame for each site (atom or lone pair)[3b], constructed from neighboring atoms (Figure 1). While the molecule vibrates, turns or diffuses during the simulation, the atomic multipoles remain constant within the defined local context shown in Figure 1.

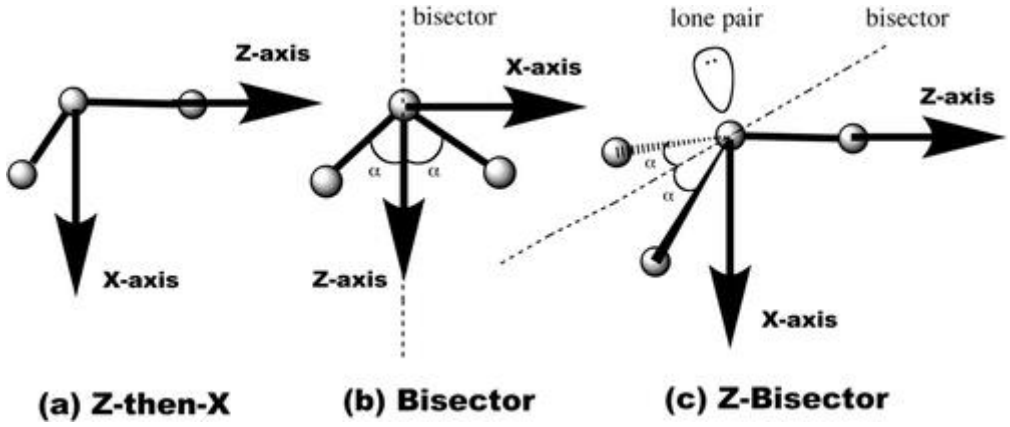


Figure 1. Local coordinated frame definitions for atomic multipole sites (a) The Z-then-X convention (b) The bisector convention (c) The convention Z-then-bisector (figure 1 in ref 12b)

The z-then-x convention in Figure 1 (a) is used for most sites for asymmetric triplets of atoms. By adding the orthogonal axis Y, chiral centers can be treated. The bisector convention (Figure 1 (b)) is used when the Z axis can be defined by the angle bisector formed with the two neighbors. It is typically used in the case of C_{2V} symmetry, such as water or a carboxylate group. The convention Z-bisector (Figure 1 (c)) is used for sites such as the sulfur atom of dimethylsulfoxide, which has a distinct primary Z axis and symmetry or pseudo-symmetry along a secondary direction. In each case, the positive Y axis is defined to create a cubic right-handed coordinate system. In the AMOEBA parameter set, we use a negative sign (-) to signal a multipole orientation in this bisector convention.

The distributed multipole analysis (DMA)[31] method is used to obtain the multipole distribution. This method was implemented in the GDMA software developed by Stone[32]. It evaluates the exact representations of charge density on Gaussian basis sets and approximates each of them by a multipole expansion on atomic nuclei sites and possible additional sites. However, the method is strongly basis-

set dependent. When the procedure is used with diffuse basis sets, the distribution of the multipoles leads frequently to non-physical values[33]. A new version of DMA was developed to solve this problem by using numerical quadrature for the diffuse functions, while retaining the original method for the more compact ones[33]. Furthermore, DMA multipoles truncated at quadrupoles is an approximation and could lead to errors in comparison with the electrostatic energy surface.[34] Ren and Ponder have proposed a method based on a fit of the electrostatic potential. They optimized the multipoles extracted in DMA on an electrostatic potential surface computed at a quantum chemistry level, usually MP2, associated to a large basis set if the size of the system allows it. This fitting optimizes only the dipole and quadrupole moments while the monopoles are kept fixed. The convergence criterion should not be set too tight to avoid over fitting. Ponder *et al*. have shown some examples of properties computed using both DMA and fitted multipoles. In some cases, free energy results have shown an improvement, but other results have not demonstrated an obvious difference[12c].

After parameters of the system are created, validation steps need to be carried out by comparing with ab initio simulations or experimental results. One can adjust the original parameters to enforce agreement with the reference results[35].

III. PARAMETERS OF THE HEME COFACTOR IN ITS FERROUS AND FERRIC FORMS

Our first objective toward simulation of heme proteins with AMOEBA has been the derivation of parameters for heme. This work has been described in an article published in 2018 in the Journal of Chemical Theory and Computation[20]. It is reproduced in the following pages.

Abstract

"We report the first parameters of the heme redox cofactors for the polarizable AMOEBA force field in both the ferric and ferrous forms. We consider two types of complexes, one with two histidine side chains as axial ligands and one with a histidine and a methionine side chains as ligands. We have derived permanent multipoles from second-order Møller-Plesset perturbation theory (MP2). The sets of parameters have been validated in a first step by comparison of AMOEBA interaction energies of heme and a collection of biologically relevant molecules with MP2 and Density Functional Theory (DFT) calculations. In a second validation step we consider interaction energies with large aggregates comprising around 80 H_2O molecules. These calculations are repeated for thirty structures extracted from semi-empirical PM7 MD simulations. Very encouraging agreement is found between DFT and AMOEBA force field which results from an accurate treatment of electrostatic interactions. We finally report long (10 ns) MD simulations of cytochromes in two redox states with AMOEBA testing both the 2003 and 2014 AMOEBA water models. These simulations have been carried out with the TINKER-HP (High Performance) program. In conclusion, owing to their ubiquity in biology we think the present work opens a wide array of applications of the polarizable AMOEBA force field on hemeproteins."

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 2705−2720

pubs.acs.org/JCTC

AMOEBA Polarizable Force Field Parameters of the Heme Cofactor in Its Ferrous and Ferric Forms

Xiaojing Wu,[†] Carine Clavaguera,[*,†] Louis Lagardère,[‡,§] Jean-Philip Piquemal,[∥,⊥,#] and Aurélien de la Lande[*,†]

[†]Laboratoire de Chimie Physique, Université Paris Sud - CNRS, Université Paris Saclay, 15 Avenue Jean Perrin, 91405 Orsay Cedex, France

[‡]Sorbonne Université, CNRS, Institut Parisien de Chimie Physique et Théorique (IP2CT), 4 Place Jussieu, F-75005, Paris, France

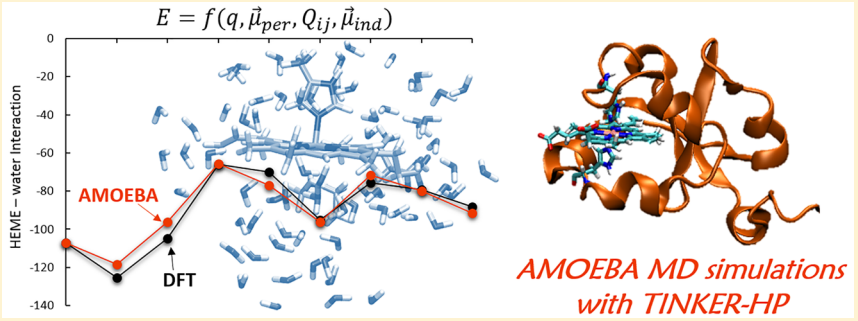
[§]Sorbonne Université, Institut des Sciences du Calcul et des Données (ISCD), 4 place Jussieu, F-75005, Paris, France

[∥]Sorbonne Université, CNRS, Laboratoire de Chimie Théorique (LCT), 4 Place Jussieu, F-75005, Paris, France

[⊥]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

[#]Institut Universitaire de France, 75005, Paris, France

Supporting Information



ABSTRACT: We report the first parameters of the heme redox cofactors for the polarizable AMOEBA force field in both the ferric and ferrous forms. We consider two types of complexes, one with two histidine side chains as axial ligands and one with a histidine and a methionine side chain as ligands. We have derived permanent multipoles from second-order Møller−Plesset perturbation theory (MP2). The sets of parameters have been validated in a first step by comparison of AMOEBA interaction energies of heme and a collection of biologically relevant molecules with MP2 and Density Functional Theory (DFT) calculations. In a second validation step, we consider interaction energies with large aggregates comprising around 80 H_2O molecules. These calculations are repeated for 30 structures extracted from semiempirical PM7 DM simulations. Very encouraging agreement is found between DFT and the AMOEBA force field, which results from an accurate treatment of electrostatic interactions. We finally report long (10 ns) MD simulations of cytochromes in two redox states with AMOEBA testing both the 2003 and 2014 AMOEBA water models. These simulations have been carried out with the TINKER-HP (High Performance) program. In conclusion, owing to their ubiquity in biology, we think the present work opens a wide array of applications of the polarizable AMOEBA force field on hemeproteins.

I. INTRODUCTION

Hemeproteins play important roles in diverse biological functions including transportation or storage of dioxygen (e.g., hemoglobin, myoglobin, neuroglobin), electron transport (e.g., cytochromes), or in enzymatic reactions (e.g., cytochromes P450). This remarkable diversity of functions results from the specific chemical structure of the heme motif, which consists of an iron cation chelated by the four nitrogen atoms of a deprotonated porphyrin ligand. The iron cation arranges in an almost planar geometry with one or two ligands completing the coordination sphere in axial position. The axial ligand may either be amino acid residues, typically histidine or methionine side chains, or exogenous ligands (O_2, NO, H_2O...). Both the ferrous and ferric forms of the iron cation can be stabilized by the complex. This ensemble of structural and redox properties confers distinct biological functions to hemeproteins. Intensive research in both experiments[1−4] and theoretical simulations[5−14] have been done to unravel the molecular mechanisms associated with the biological functions of hemeproteins.

Focusing on the modeling of the redox properties of hemeproteins, much progress has been realized over the past two decades.[7,8,10,15,16] Computational modeling is essential in this research field to help understand the variability of redox

Received: November 9, 2017
Published: April 9, 2018

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

properties of different hemeproteins. For example, one can rationalize variations of redox potentials or of reorganization energies of hemeproteins from the secondary structure of the protein. Indeed, the presence of charged or polar chemical groups is important in determining the free energies of oxidation of hemes. It is now well documented that the redox properties of hemeproteins depend not only on the structure of the heme cofactor but also on the structure and on the dynamics of the protein matrix and on the hydration level of the proteins.[17] When interested in evaluating redox potentials or electron transfer rates, a particularly powerful approach consists of combining quantum mechanical (QM) methodologies to molecular dynamics (MD) simulations with molecular mechanics (MM) methodologies (i.e., classical force fields). We refer the reader to recent review papers detailing the different strategies employed so far in the literature to evaluate redox properties of proteins.[15,18,19] There are several parameters that impact the accuracy of computed redox properties: (i) the choice of the QM method to evaluate the intrinsic propensity of heme to lose an electron (ionization potential), (ii) the accuracy of the force field (FF) to describe the environment and its interaction with the heme, and (iii) the extensiveness of the conformational sampling of the proteins in the different redox states. Even though much progress has been realized in past decades in terms of the accuracy of redox property computations, numerical approaches are still far from reaching experimental precision. For instance, uncertainties of a few millivolts on the redox potentials which are reachable by electrochemical means correspond to an accuracy of less than 1 kcal/mol on free energies of oxidation. A key challenge is to accurately describe electrostatic interactions between the heme cofactor and its environment (protein, water, counterions...). Indeed the change from the +II to +III (or *vice versa*) induces significant conformational rearrangements of the environment. Previous studies showed that inclusion of electrostatic induction in classical FFs is mandatory to capture reorganization of surrounding atoms.[17,20−22] Indeed, the associated reorganization energy is related to the optical dielectric constant of the medium and can only be reproduced by polarizable FFs.[8,15,18] Another fundamental limitation of standard FFs is the monopolar representation of the permanent electron cloud of molecules.

Force field parameters for heme have been proposed for various nonpolarizable FFs.[23−30] In this work, we are interested in the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field implemented in the Tinker program package (version 7.1.2).[31] AMOEBA stands as a highly accurate FF for water[32,33] and for ion hydration[34−38] and to reproduce the structures and thermodynamics of organic molecules and biomolecules.[39−43] The AMOEBA water model was also compared to the Drude polarizable force field[44] to evaluate reorganization free energy for electron self-exchange in aqueous Ru(II)−Ru(III).[45] Electrostatic interactions between permanent charge distributions are accounted for by sets of permanent monopolar, dipolar, and quadrupolar moments centered on atoms. Moreover, AMOEBA implements an induction model which allows reproduction of the anisotropy and the nonadditivity of the molecular polarization response. Furthermore, recent developments have introduced short-range penetration corrections between molecules.[46−48] We are not aware of any MD simulation of hemeproteins carried out with AMOEBA, a situation undoubtedly related on one hand to the lack of parameters for the heme cofactor and on the other hand to the computational cost of such simulations with common computer codes. Indeed, as one can expect, such a sophisticated FF is far more expensive than standard force fields developed to study biomolecules. In that regard, noticeable algorithmic developments in the Tinker-HP (High-Performance) software[49] have been carried out by some of us which enable long (>100 ns) MD simulations on biological systems with AMOEBA.[50,51] The development of AMOEBA FF parameters of heme is therefore timely. We report here to the best of our knowledge the first sets of AMOEBA parameters for the heme cofactor in both the ferric and ferrous states.

The details of the parametrization procedure will be given in section II. The sets of parameters will be validated against quantum chemistry calculations in section III. The results show a clear improvement of interaction energies between heme and surrounding molecules compared to point-charge FF. Finally, we will report in section IV, to the best of our knowledge, the first nanosecond MD simulations of a cytochrome with AMOEBA as implemented in TINKER-HP.[49]

II. METHODOLOGY

II.A. The Amoeba Potential Energy Model. The functional form of the potential energy computed by AMOEBA is given by eq 1:

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{cross}} + U_{\text{opp}} + U_{\text{torsion}} + U_{\text{vdW}} + U_{\text{ele}}^{\text{perm}} + U_{\text{ele}}^{\text{ind}} \quad (1)$$

The first five terms are the valence interactions including bonds, angles, bond-angle cross coupling terms, out-of-plane deformations, and torsional rotations. AMOEBA uses mathematically flexible expressions that go beyond the harmonic approximation for the intramolecular terms. These terms have the same functional forms as those used by the MM3 force field.[52] Analytical expressions for each of these terms can be found in ref 39. The last three terms of eq 1 gather intermolecular interactions. U_{vdw} refers to van der Waals interactions and is calculated with Halgren's buffered 14−7 function. This function yields a repulsive region softer than the Lennard-Jones 6−12 function but steeper than typical Buckingham exp-6 formulations. $U_{\text{ele}}^{\text{perm}}$ collects the electrostatic interactions between permanent multipoles. Monopoles, dipole vectors, and quadrupole tensors are positioned on each atom site in order to reproduce the molecular electrostatic potentials accurately. These permanent atomic multipoles are defined with respect to local frame and maintain a constant orientation during simulations. Finally, $U_{\text{ele}}^{\text{ind}}$ refers to interactions between permanent charges and induced dipoles. Atomic polarizabilities are included to determine induced dipole moments on each polarizable site. Electronic induction is achieved via an interactive atomic dipole induction scheme.[51] The induced dipole at site *i* further polarizes all other sites until the induced dipoles at each site reach convergence. The simulations use a preconditioned conjugate gradient solver coupled to Kolafa's Always Stable Predictor-Corrector (ASPC) algorithm.[53] To avoid the polarization catastrophe phenomenon, AMOEBA employs Thole's short-range damping method.[54] In the present work, we do not include short-range penetration corrections.[46,48]

II.B. Parametrization Strategy. We have followed the standardized parametrization protocol of AMOEBA FF.[55] As recalled in the Introduction, the coordination sphere of Fe in heme can encompass various chemical groups depending on

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

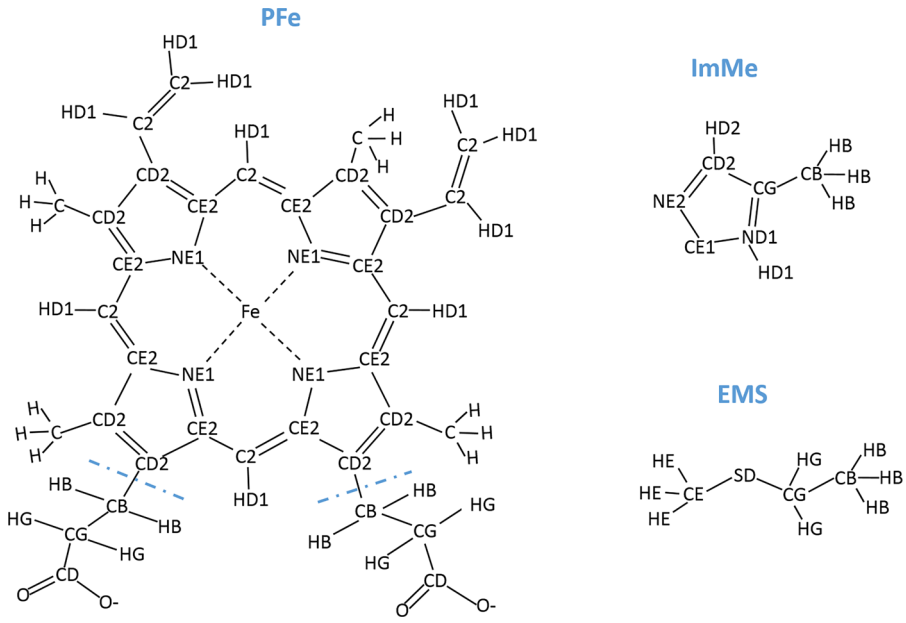


Figure 1. Heme structure, ligand models, and atom definitions for the parametrization procedure.

the protein of interest. For example, the Fe cation can be coordinated by one or two histidine residues (*e.g.*, cytochrome P450 or cytochrome b_5), by one histidine and one methionine residue (cytochrome c551), or by histidine and small ligands (like O_2, NO...). These different coordination spheres would necessitate distinct sets of parameters. In this work, we are primarily interested in six-coordinated hemes with either two histidines or one histidine and one methionine axial ligand. Such coordination patterns are, for instance, frequently encountered in cytochromes or globins. In our first parametrization attempts, we considered the full six-coordinated complexes; however, we experienced difficulties in obtaining reliable multipoles, and we therefore decided to adopt another strategy. Our approach consists of parametrizing the four-coordinated iron porphyrin ($PFe^{II/III}$), on one hand, and the axial ligands separately. This approach is justified when the axial ligands are either histidine or methionine residues by the fact that charge transfer from the iron cation toward the ligand is rather small. Indeed, charge transfer from the axial ligands to the iron-porphyrin complex amounts to 0.05 e^- and 0.15 e^- in the ferrous and ferric states respectively, based on Density Functional Theory (DFT) calculations combined with the iterative Hirshfeld[56] electronic population scheme (Table S1). We expect the multipolar description of the FF to describe sufficiently well the interaction between PFe and the axial ligands. The chemical structure of the PFe group used during the parametrization protocol is depicted in Figure 1. This PFe group has been further truncated into three fragments as shown in the figure. The amino acid backbone atoms of histidine and methionine have been modeled as methyl-imidazole (ImMe) and ethyl-methyl-sulfide (EMS), respectively. Note that this strategy also assumes that no spin density is transferred from the iron porphyrin core to the axial ligands (see Table S1).

Figure 1 depicts the atom types for all atoms. For the intramolecular terms, the atom types and classes of PFe atoms and the two ligands have been defined by analogy with parameters from the AMOEBA-2013 FF for proteins.[43] For example, the AMOEBA classes and types of atoms CD2, CE2, and NE1 within the porphyrin ligand were transposed from the analogous atoms of the five-membered ring of the tryptophan residue. NE1 is nitrogen atoms pertaining to the azole ring, CE2 is sp^2's of the azole cycle linked to NE1 and to one sp^3 carbon and one sp^2 carbon. Finally, CD2 is sp^2's of the azole cycle linked to two other sp^2 carbon atoms. The valence parameters (bond stretching, angle bending, torsions, van der Waals, and polarizabilities) were taken from the set without modification. The van der Waals and atomic polarizability for Fe were taken from the work of Semrouni et al. for the ferrous state.[34] The same van der Waals parameters are here for both the ferrous and ferric state. This has been common practice for simulations of hemeproteins using more standard force fields.[18,27] We actually think that for iron cations nested at the heart of the porphyrin ligand in well-defined coordination spheres, the differences of nonbonding interactions between the two redox states can be captured by adequately tuned sets of multipoles on the metal ion and its coordination sphere. This methodological strategy will be validated extensively in sections II and III.

A central aspect of the parametrization procedure is the fitting of multipoles for PFe. The sets of multipoles have to be different in the ferrous and ferric redox states. To that end, we have optimized the geometry of the [PFe-ImMe-EMS] complex by DFT calculations with the OPTX-PBE functional[57,58] and the DZVP-GGA basis set.[59] These calculations have been carried out with deMon2k.[60] To avoid spurious electronic delocalization between the iron cation and the carboxylate functions, the latter have been protonated during geometry optimizations. The OPTX-PBE functional has been chosen for its good performance to reproduce the electronic energies of different spin states in transition metal complexes.[61,62] deMon2k relies on auxiliary fitted densities to calculate the Coulomb and exchange correlation energies and potentials.[63] The auxiliary basis sets are automatically generated by the program.[59] We chose the GEN-A2 for H and C, and the more flexible, therefore more accurate, GEN-A2* for Fe, N, and S. Geometry optimizations have been conducted in the singlet, triplet, and quintet spin states for the ferrous state and in the doublet, quartet, and sextet spin states for the ferric state. We

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

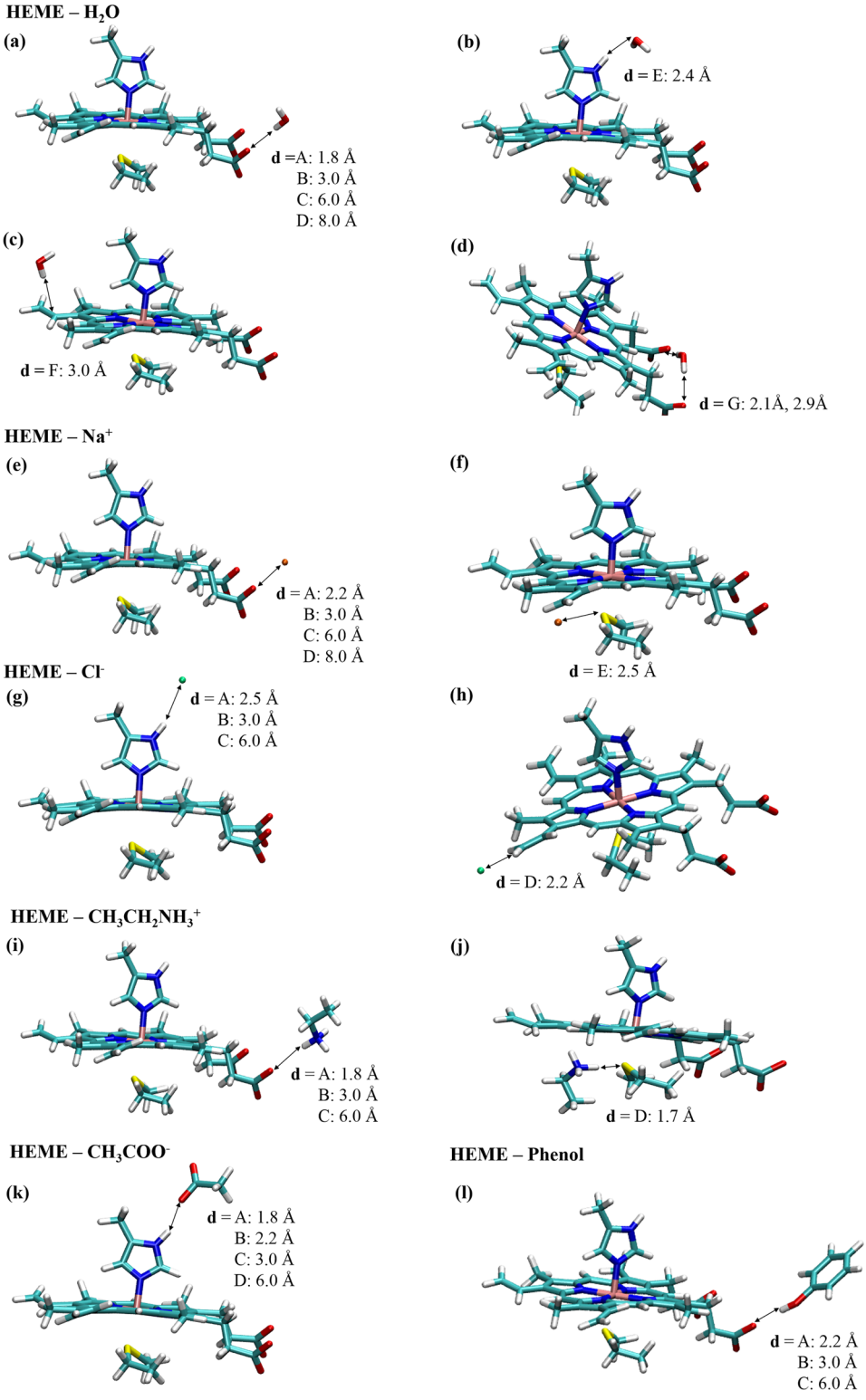


Figure 2. Geometries of heme interacting with molecules used in the validation of the AMOEBA FF parameters. Picture made with VMD.[78]

have found the singlet and doublet spin states to be the most stable for ferrous and ferric redox states, respectively, by 0.72 and 0.57 eV. We have restricted our parametrization procedure to these two spin states.

After geometry optimization, the structures have been fragmented according to the partition of Figure 1. Hydrogen atoms have been added at the cutting positions between the porphyrin ring and the propionate groups, and we have reoptimized the hydrogen atom positions. Following the AMOEBA parametrization procedure, single point calculations of each fragment have been performed at the MP2/cc-pVDZ level of theory with Gaussian 09.[64] The ground states of four-coordinated iron-porphyrins are of intermediate spin. MP2 calculations on the fragments have been carried out in the lowest spin state to be consistent with the lowest spin state of the six-coordinated iron-porphyrin. The Distributed Multipole

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

Analysis (DMA) has been carried out from the MP2 electron density using the GDMA program (version 2.2.11) and the original DMA algorithm.[65−68] The default relative atomic radii used in the DMA algorithm have been chosen except for hydrogen, for which a value of 0.31 has been chosen. This value was previously shown to be appropriate to avoid erroneous charges during the DMA procedure.[69]

The POLEDIT program available in Tinker[31] has been subsequently run on the GDMA outputs with the suggested polarizability values. All atoms are placed into a single polarization group by default. This resulting version of multipoles is obtained directly from the DMA procedure. This set of parameters will be referred to as "AMOEBA DMA." Another approach to derive atomic multipoles involves an optimization against the QM electrostatic potential around the system. The AMOEBA DMA multipole values have been used as starting values to the fitting against the MP2/cc-pVDZ electrostatic potential. The DMA partial charges are held fixed during the potential fitting process, while dipoles and quadrupoles are readjusted. The final gradient convergence value has been set to 0.5 kcal mol^{-1} $electron^{-2}$ to avoid overfitting. Indeed, a tighter convergence criterion can lead to multipoles that depend strongly on the geometry used for the extraction, and therefore that are less transferable. We will refer to this second set of parameters as "AMOEBA FIT." Finally, according to the defined atom types, the multipole values have been averaged over the equivalent atoms. The procedure was carried out for each fragment. Two versions of each multipole set have been created for both ferrous and ferric states. At the end, all the fragments have been combined by removing excess hydrogens, and the global charge has been adjusted to be −2 and −1 for ferrous and ferric states, respectively. In summary, four parameter sets have been generated and named Fe^{II} AMOEBA DMA, Fe^{II} AMOEBA FIT, Fe^{III} AMOEBA DMA, and Fe^{III} AMOEBA FIT. They will be now tested against quantum mechanical calculations. The parameters are available in the SI.

III. VALIDATION AND SIMULATIONS

In section III, we validate the different AMOEBA parameter sets on the calculation of interaction energies between the heme and its environment. We start by gas phase interaction energies (IE) between the [PFe-ImMe-EMS] complex and different molecules relevant to biological systems. We continue by calculating IE with large clusters of water molecules. We also address the transferability of our parameters for the [PFe-$(ImMe)_2$] complex. We finally conclude the section by the adjustment of the internal bond and valence angle terms to tune the geometry of the iron first coordination shell.

III.A. Computational Details. As a first test of the sets of AMOEBA parameters, we here report interaction energies between the [PFe-ImMe-EMS] and various molecules in the gas phase. The IE between the iron complex and a molecule (M) is defined as

$$\Delta E_{int} = E([PFe, M]) - E([PFe]) - E(M) \quad (2)$$

where $E([PFe,M])$ is the energy of the supramolecular complex and $E([PFe])$ and $E(M)$ are the energies of the two fragments. For each supramolecular system, the geometry has been constructed by associating the [PFe^{II}−ImMe−EMS] complex, geometrically optimized in the absence of the partner M, with the geometry of the fragment. A restrained geometry optimization of the supramolecular ensemble has been carried out at the DFT level, freezing the geometry of the heme partner, and optimizing only the internal geometry of partner M. In these restrained optimizations, we have eventually further imposed the distances between the two partners to a given value. Basis set superposition error corrections have not been computed. The list of molecules includes H_2O, $CH_3CH_2NH_3^+$ (a model of lysine side chain), CH_3COO^- (a model of aspartate and glutamate side chains), phenol (a model of tyrosine side chain), and the chlorine and sodium ions. This is a list of representative types of interactions one can find in hemeproteins between the heme cofactors and its environment. In particular, charged or polar residues are known to play a special role in determining redox properties. For example, in flavohemoglobin, a lysine residue is hydrogen bonded to one propionate group of the heme,[70,71] and a glutamate residue interacts with axial histidine residues.[70] Tyrosine residues have also been found to interact with heme propionates, for instance, in cytochrome *c* oxidase.[72] It is therefore important to assess whether the parameters developed in this work are able to accurately reproduce the associated interaction energies. The various supramolecular complexes are depicted in Figure 2.

Two types of QM calculations have been carried out, relying either on DFT (B3LYP[73]) or on MP2. Note that we have tested B3LYP-D3 calculations for some complexes and found negligible effects of dispersion on the computed interaction energies. We mention that convergence of Hartree−Fock or the DFT self-consistent field of deprotonated heme is tedious in the gas phase, especially for the ferric state. Indeed, the terminal propionate groups tend to be oxidized instead of the iron cation, eventually causing severe self-consistent field (SCF) convergence issues and unexpected electronic structures. It was therefore not always possible to obtain QM results for some geometries. MP2 single-point calculations have been performed within the Resolution of the Identity (RI) approach using the TZVP basis set for all atoms with the Turbomole 7.1 program.[74] DFT single-point energies have been calculated with deMon2k at the B3LYP/TZVP level (DZVP-GGA for Fe) and with the auxiliary function set GEN-A2* for all atoms. An adaptive grid of accuracy 10^{-8} Ha was defined to integrate the exchange correlation energy and potential.[75] Exact exchange was computed via a variational fitting of the Fock potential.[76]

For the FF calculations, the interaction energies with AMOEBA have been computed with the ANALYZE program from the TINKER package. The parameters of the small molecules and ions are taken from the AMOEBA-2013 parameter set for proteins. For H_2O, two versions of the AMOEBA water model, water03[32] and water14,[33] are tested. For $CH_3CH_2NH_3^+$ and CH_3COO^-, sets of multipoles have been determined with the previously described procedure for the heme. The energies entering eq 2 were calculated with eq 1 (AMOEBA). The AMOEBA interaction energies therefore include energy differences from the van der Waals, permanent electrostatic, and induction contributions to the energy. We have also tested a nonpolarizable FF based on the permanent point charge model. For this FF, the reported interaction energies include van der Waals contributions calculated with a 12−6 Lennard-Jones potential and a charge−charge Coulomb interaction. The parameters of vdW are taken from the CHARMM for ferrous heme.[27] The punctual charges have been derived by an iterative Hirshfeld (IH) population analysis[56] as implemented in deMon2k.[77] For water, the IH charges are close to those given by the TIP3P water model. All the computed interaction energies can be found in Table S2.

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

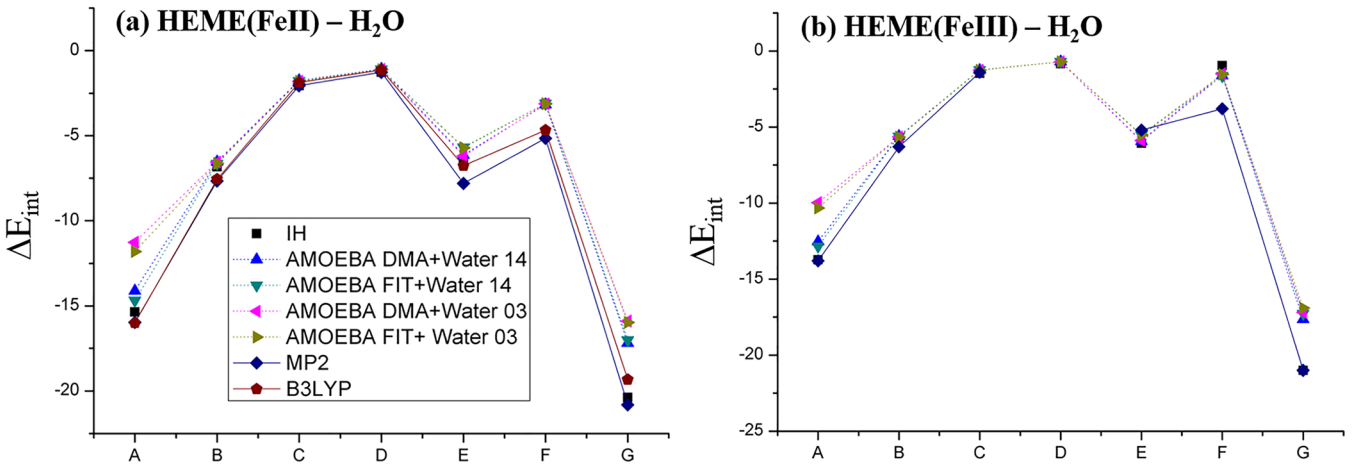


Figure 3. Heme−water interaction energies (kcal/mol) of the different geometries represented in Figure 2. Left, results of ferrous state; right, results of ferric state.

Table 1. Interaction Energy Differences in kcal/mol Obtained at the Various Computational Levels with Respect to the MP2/TZVP Reference

geometry	B3LYP	IH	AMOEBA DMA + water14	AMOEB FIT + water14	AMOEBA DMA + water03	AMOEBA FIT + water03
			HEME(FeII) − H_2O			
A	0.03	0.61	1.86	1.28	4.72	4.18
B	0.11	0.85	1.13	1.02	1.10	1.03
C	0.19	0.00	0.30	0.28	0.26	0.25
D	0.14	0.01	0.20	0.19	0.17	0.16
E	1.03	1.14	1.67	2.13	1.61	2.10
F	0.49	2.02	1.99	2.04	2.03	2.04
G	1.48	0.42	3.62	3.80	4.92	4.83
			HEME(FeIII) − H_2O			
A	n.c.[b]	0.07	1.28	0.95	3.83	3.48
B	n.c.	0.60	0.70	0.69	0.65	0.64
C	n.c.	0.06	0.14	0.15	0.13	0.14
D[a]	n.c.	0.00	0.12	0.12	0.11	0.11
E	n.c.	0.86	0.74	0.34	0.69	0.30
F	n.c.	2.85	2.19	2.16	2.30	2.26
G	n.c.	0.00	3.36	3.73	3.81	4.10

[a]Neither MP2 nor B3LYP calculations converged for geometry D in the ferric state. For this series, IH is taken arbitrarily as a reference. [b]The B3LYP SCF could not converge.

III.B. Interaction between Heme and Single Molecules. *Heme−Water.* We start our discussion with results on IE between heme and a water molecule (see Figure 2a−d). Geometries a correspond to a water molecule engaged in hydrogen bonding with one propionate side chain with different lengths (1.8 Å for A, 3.0 Å for B, 6.0 Å for C, 8.0 Å for D). Geometry b,E corresponds to a hydrogen-bond interaction from the NH group of ImMe toward the oxygen atom of water. The equilibrium distance resulting from geometry optimization with DFT amounts to 2.4 Å. Geometry c,F corresponds to a weak OH−π interaction between one C=C double bond of heme and one hydrogen atom of water, with an interaction distance of 3.0 Å. Finally, geometry d,G involves a double hydrogen-bond interaction with the water bridging the two propionate side chains with the O−H distances amounting to 2.1 and 2.9 Å. Results are plotted in Figure 3. For geometries A to D, all the methods give the same trend. As expected, the interaction becomes less favorable as the distance increases.

The interaction energy differences for each method with respect to MP2 are provided in Table 1. For geometries C and D, in which the hydrogen bond is weak (>3 Å), all the methods give an IE within 0.3 kcal/mol to that given by MP2. For such long-distance interactions, a monopolar description of the electron density is likely to be valid and therefore accounts for the convergence of nonpolarizable and AMOEBA FF results. For geometries A and B, the differences are more significant between the methods. Among them, the IH results are the closest to MP2 with an error of less than 1.0 kcal/mol. For AMOEBA calculations on geometry B, all parameter sets give similar results. For geometry A, a larger error is found with the water03 model. A plausible explanation for this discrepancy may be the lack of correction for penetration effects.[46] These effects are not taken into account by the version of the AMOEBA FF used in this work. On the other hand, the water14 model is able to significantly reduce the error with respect to MP2 due to the reparameterization of the electrostatic term.

For geometry E, the results for IH and AMOEBA are almost similar. The differences of IE with respect to MP2 are ca. 2 kcal/mol for the ferrous state, while they are only ca. 0.7 kcal/mol for the ferric state. The results with the different FFS for geometry F are similar for the ferrous state with a difference at

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

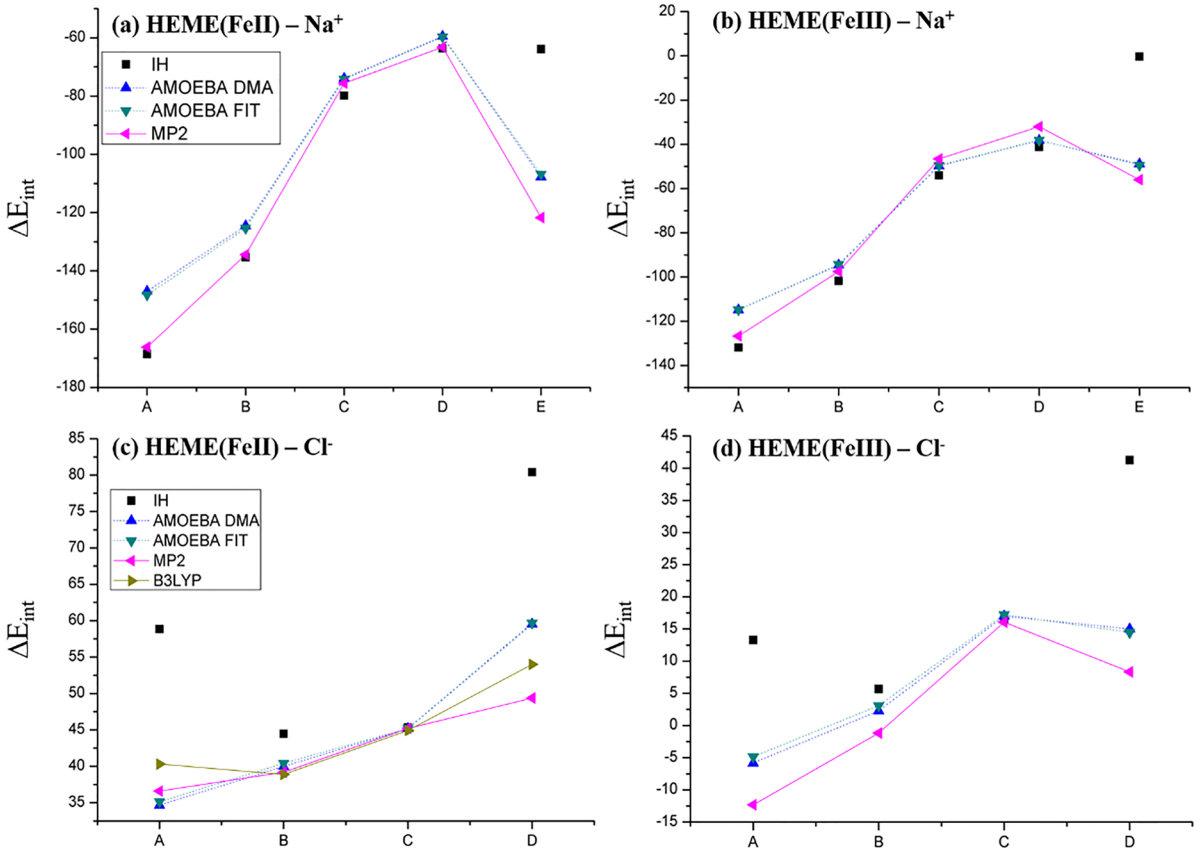


Figure 4. Heme−ions (Na^+ and Cl^-) interaction energies (kcal/mol) of the different geometries represented in Figure 2. Left: results for the ferrous state. Right: results for the ferric state. B3LYP calculations for the ferric states are not reported because of self-consistent-field convergence issues.

ca. 1.6 kcal/mol. However, regarding the ferric state, AMOEBA gives a difference of 2.1 kcal/mol, while IH has a difference of 2.9 kcal/mol. For geometry G, the error for both states is ca. 3.5 and 5 kcal/mol for the water14 and water03 parameters, respectively. As for geometry A, we find that the water14 model seems to perform better, and the nonpolarizable model gives as accurate results.

Heme−Atomic Ions. Five geometries were created to test the accuracy of the interaction with the sodium cation. As shown in Figure 2e, we have investigated interactions with either the propionates groups (A, B, C, or D) or the sulfur atom of the EMS ligand (E). The results obtained with the different computational methods are plotted in Figure 4 for the ferrous and ferric states (panels a and b, respectively). For the A, B, C, and D series, the interaction becomes less favorable with increasing the distance, while the difference between the various methods decreases. This is similar to what was found for the interaction with H_2O. For geometries C and D in the ferrous state, all methods are within 4 kcal/mol of the MP2 values. The energy differences are larger (ca.7 kcal/mol) for the ferric state. For geometries A and B, IH gives good results with errors lower than 5 kcal/mol. With the AMOEBA FF, we obtain differences around 19 and 9 kcal/mol for geometries A and B, respectively. For geometry E, AMOEBA gives a difference of 14 and 7 kcal/mol with respect to MP2 for the ferrous and ferric states, respectively. For this type of interaction involving a large and polarizable group (thioether), AMOEBA represents a clear improvement over nonpolarizable FF for which differences larger than 55 kcal/mol are found.

We now turn to interaction with chlorine (Figure 2). Geometries A, B, and C correspond to interaction with the ImMe, while geometry D corresponds to a weak dispersive interaction between Cl^- and one C=C bond or the porphyrin ring. Results are shown in Figure 4 (panels c and d). We remark that as a consequence of the overall −2 or −1 charge of the heme complex in the ferrous and ferric states, the interaction energies with the chlorine anion are almost always positive. Nevertheless, in actual biosystems, other interactions with positively charged residues may counterbalance the repulsive interaction with heme so that anions may still approach them. It is therefore important to test the capability of the AMOEBA parameters to correctly describe the electrostatic interactions between heme and anions. For geometries A, B, and C, AMOEBA reproduces the increase of IE with the distance between heme and Cl^-. The difference between AMOEBA and MP2 amounts to 2 kcal/mol in the ferrous states and is a little bit larger in the ferric state. The IH curves exhibit, on the other hand, opposite trends, failing to reproduce even qualitatively the evolution of IE provided by QM methods. We could trace back this discrepancy in the unbalanced treatment of electrostatic and Lennard-Jones interactions. AMOEBA provides a better description of the noncovalent interaction between atomic anions and heme. For geometry D also, AMOEBA gives better results for the ferrous state than IH by ca. 21 kcal/mol. This large difference can be mainly due to the interaction with the C=C bond. Here again, we see the advantage of using a polarizable force field.

Heme−$CH_3CH_2NH_3^+$. As shown in Figure 2, geometries (i) A, B, C, and D represent hydrogen-bonded interactions

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

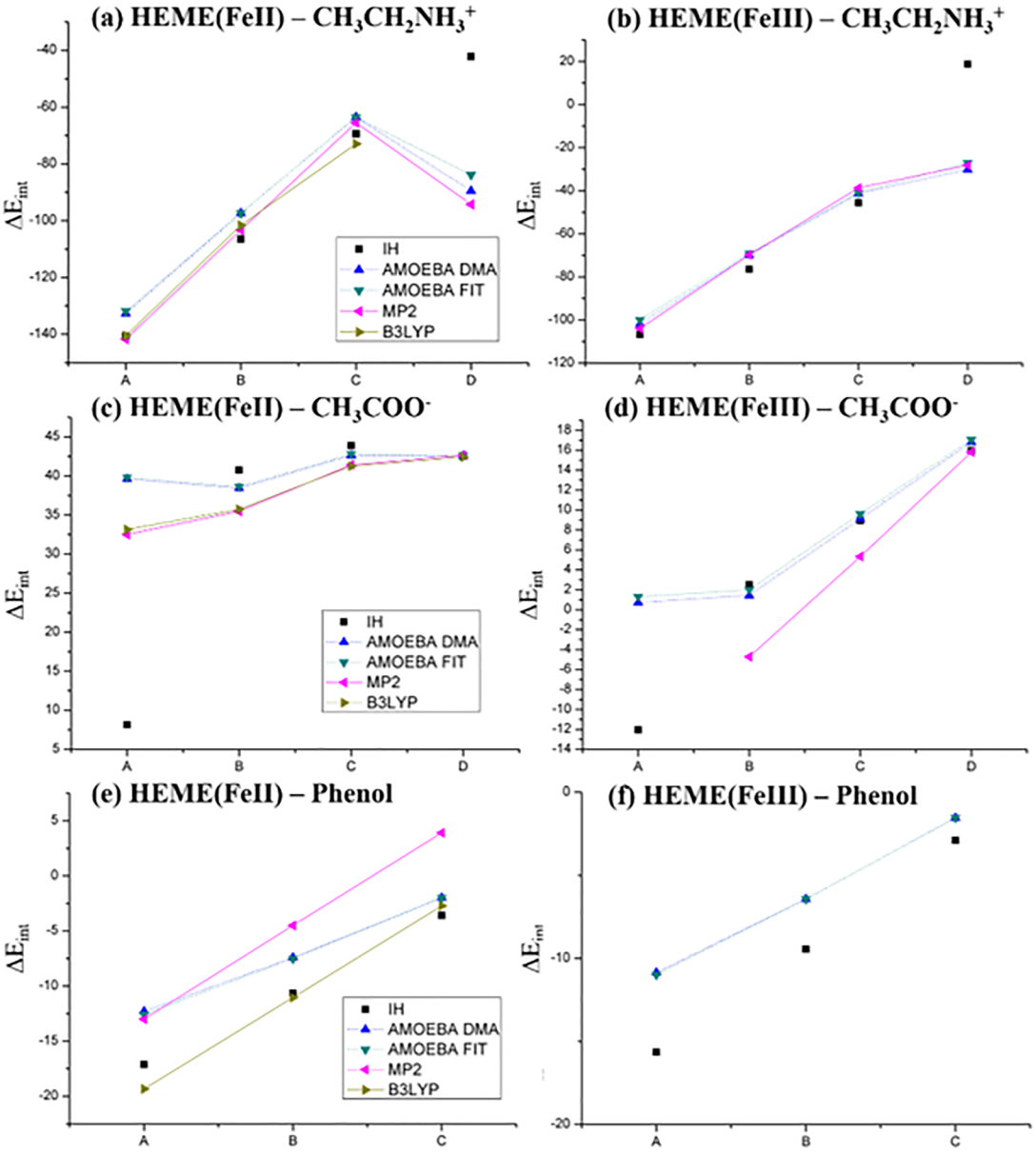


Figure 5. Heme−molecule interaction energies (kcal/mol) of the different geometries represented in Figure 2. Left: results for the ferrous state. Right: results for the ferric state. Missing points for quantum mechanical methods (MP2 and/or B3LYP) are due to self-consistent-field convergence issues.

between the propionate side chain of heme and $CH_3CH_2NH_3^+$ with increasing length. Geometry D involves an interaction with the thioether. The values of the interaction energies are plotted in Figure 5a for the ferrous state and 5b for the ferric state. Overall, we find that AMOEBA gives a satisfactory treatment of the interaction between this cationic organic species and the heme. IH also gives good results for geometries A, B, and C; however, it fails for geometry D with a difference ca. 50 kcal/mol with MP2. This finding is reminiscent of the difficulty of treating the interaction between Na^+ and the thioether ligand, illustrating again the advantage of FF relying on a balanced treatment of electrostatic interactions involving permanent multipoles and induced dipoles. That said, like for geometry D, it is probably not likely to be encountered often in the course of MD simulations owing to the short distance between the two interacting partners.

Heme−CH_3COO^-. Geometries A, B, C, and D involve a hydrogen bond between ImMe and CH_3COO^- with increasing length (Figure 2k). The values of the interaction energies are plotted in Figure 5. At long distances, all the FF results are very close to the QM based method (<0.1 kcal/mol), for both redox states. When decreasing the hydrogen bond length, the interaction between the two anionic partners becomes less unfavorable. AMOEBA as well as IH adequately reproduce the trend provided by MP2 or B3LYP. One noticeable exception is geometry A, which is overstabilized with a simple point-charge model (IH).

Heme-Phenol. As shown in Figure 2l, geometries A, B, and C represent the hydrogen-bond interaction between the phenol and the propionate side chain of heme. Interaction energies are given in Figure 5. We first remark on a systematic difference of 6.5 kcal/mol between MP2 and B3LYP for the ferrous state, which is difficult to interpret. The IH results are close to the

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

B3LYP ones, while the AMOEBA results lie in between the B3LYP and MP2. For the ferric state, we could not converge either MP2 or DFT calculations of the heme cofactor. A similar trend than the ferrous state is obtained with the three FFs.

Summary. In summary, AMOEBA is globally in good agreement with QM methods (MP2 and DFT). This is especially true for interaction with highly polarizable groups such as S, C=C double bonds, or anions like Cl^-. Finally, no large difference has been observed between the two sets of multipoles (DMA or FIT). We can observe that the FIT version gives slightly better results than the DMA one by ca. 1 kcal/mol. For short distances, larger deviations between AMOEBA and QM have been found.

III.C. Heme within Droplets of Water. In the previous section, we validated our AMOEBA parameters looking at interaction energies between the [$PFe^{II/III}$−ImMe−EMS] complex and various molecules. We now address collective effects by computing interaction energies between the [$PFe^{II/III}$−ImMe−EMS−$(H_2O)_6$] complexes and a large ensemble of water molecules. Note that we have included six water molecules around the propionate moieties of the heme to avoid spurious oxidation by the Fe^{III} cation. We also address in this section the transferability of the parameters to the description of the [PFe^{II}−$(ImMe)_2$−$(H_2O)_6$] complex.

In a preliminary step, MD simulations have been carried out with the PM7 semiempirical method and the CUBY4 environment.[79] Details of the PM7MD simulations can be found in the SI. The heme complexes ([PFe^{II}−ImMe−EMS] and [PFe^{II}−$(ImMe)_2$]) are embedded in droplets of water comprising around 1850 water molecules. After geometry optimization, MD simulations have been carried out for 80 ps. Interaction energies have been computed according to eq 2 with energies calculated using deMon2k (DFT) or Tinker (AMOEBA). The normalized autocorrelation functions of the IE obtained with AMOEBA approach zero in periods of a few hundreds of femtoseconds to a few picoseconds (Figure S1). Consequently, we have extracted snapshots for subsequent interaction energy calculations every 2 ps during the last 60 ps, leading to a total of 30 geometries for each complex that can be considered as uncorrelated. To make the DFT level tractable, the geometries have been pruned to retain only the water molecules within less than 8 Å from the iron cation. The pruned geometries typically comprise 75 to 80 H_2O (for a total of 330 atoms on average). To carry out these computationally very intensive calculations, the Coulomb, local exchange-correlation, and nonlocal exact exchange contribution to the Kohan−Sham potential have been computed with auxiliary fitted quantities.[63,76] We have used the DZVP-GGA/GEN-A2* combination of atomic orbital and auxiliary basis sets. The grid accuracy to integrate the exchange correlation energy and potential has been set to 10^{-8} Ha. Calculations have been carried out with B3LYP,[73] PBE,[58] and PBE0.[80] For complexes in the ferrous states, SCF convergence has been reached with a tolerance of 10^{-7} Ha on the SCF energy and 10^{-6} on density fitting coefficients. For the ferric state, SCF convergence is much more difficult to obtain on all the complexes, and tolerance thresholds have been set to 10^{-4} Ha and 4.10^{-4} for some of them. However, we have verified on nine geometries for which tight convergence could be reached up to 10^{-7} Ha and 10^{-6} respectively, that the computed interaction energies are within 0.05 kcal/mol of those obtained with the looser 10^{-4} Ha and 4.10^{-4} convergence thresholds (Table S3).

Results are summarized in Table 2 for PBE0 and in Table S4 for PBE and B3LYP. The full list of IE is given in Table S5. For each complex, we report the average difference between the DFT and force field interaction energies ($\langle\Delta IE\rangle$) over the series and the associated standard deviation ($\sigma_{(\Delta IE)}$). We also report the Root-Mean-Square Error (RMSE) and the linear correlation coefficient between DFT results and FF results (R^2). We start by remarking noticeable differences among DFT methods for the [PFeII−ImMe−EMS−$(H_2O)_6$] complex. On one hand, PBE and PBE0 give similar IEs, with an average difference of *ca.* 3 kcal/mol. On the other hand, interaction energies calculated with B3LYP are shifted by 30 kcal/mol with respect to both PBE and PBE0 (Table S4). This discrepancy among various DFT XC functionals is large and illustrates the difficulty of obtaining reference values for validating our FF parameters. Recent benchmark calculations of relative energies of large water clusters against CCSD(T)/CBS (Complete Basis Set Limit) showed the sensitivity of such computations with XC functionals.[81] The authors recommended the use of range-separated hybrids (ωB97XD[82] or LC-ωPBE-D3[83,84]) or meta-GGA global hybrids (M05-2X[85]) for this type of calculation. These classes of DFT functionals are not currently available in the version of deMon2k we are using. PBE0 was ranked before PBE and B3LYP in this study, and we will mainly base our discussion on the PBE0 results.

Table 2. Average Interaction Energy (kcal/mol), Standard Deviation, RMSE, and Linear Correlation Coefficient between DFT (PBE0) Results and Force Field Results

		AMOEBA			
		DMA	FIT	DMA	FIT
	IH	water14	water14	water03	water 03
		[PFe^{II}ImMe−EMS−$(H_2O)_6$]			
$\langle\Delta IE\rangle$	−26.36	0.95	2.27	1.91	3.29
$\sigma_{(\Delta IE)}$	14.72	8.35	8.06	11.82	11.55
RMSE	30.07	8.26	8.25	11.77	11.83
R^2	0.70	0.86	0.87	0.73	0.74
		[PFe^{III}ImMe−EMS−$(H_2O)_6$]			
$\langle\Delta IE\rangle$	−25.84	−3.74	−2.46	−4.37	−2.93
$\sigma_{(\Delta IE)}$	13.26	7.82	7.57	10.02	10.06
RMSE	28.91	8.52	7.80	10.73	10.26
R^2	0.71	0.87	0.88	0.78	0.78
		[PFe^{II}−$(ImMe)_2$−$(H_2O)_6$]			
$\langle\Delta IE\rangle$	39.27	13.04	14.84	14.02	15.91
$\sigma_{(\Delta IE)}$	17.68	7.18	7.22	8.41	8.57
RMSE	42.94	14.82	16.44	16.27	18.00
R^2	0.62	0.94	0.94	0.93	0.93

We start our discussion with the [$PFe^{II/III}$−ImMe−EMS−$(H_2O)_6$] complex. The interaction energies computed with the nonpolarizable FF (using IH charges) lead to an average shift larger than 23 kcal/mol with respect to the PBE0 results. The scattering of the computed data is also large. As can be seen in Figure 6 (top, black points) the agreement with DFT is poor. This is reflected by a correlation coefficient of only 0.7. IEs obtained with AMOEBA are much more satisfactory, as seen from the figure and from the values of R^2. The agreement between AMOEBA and DFT (either PBE or PBE0) is especially good when we use the 2014 AMOEBA water model with R^2 approaching 0.9. The RMSE drops from 30 to 8.2 kcal/mol between the nonpolarizable FF and AMOEBA. These are very satisfactory results, indicating that moving from

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

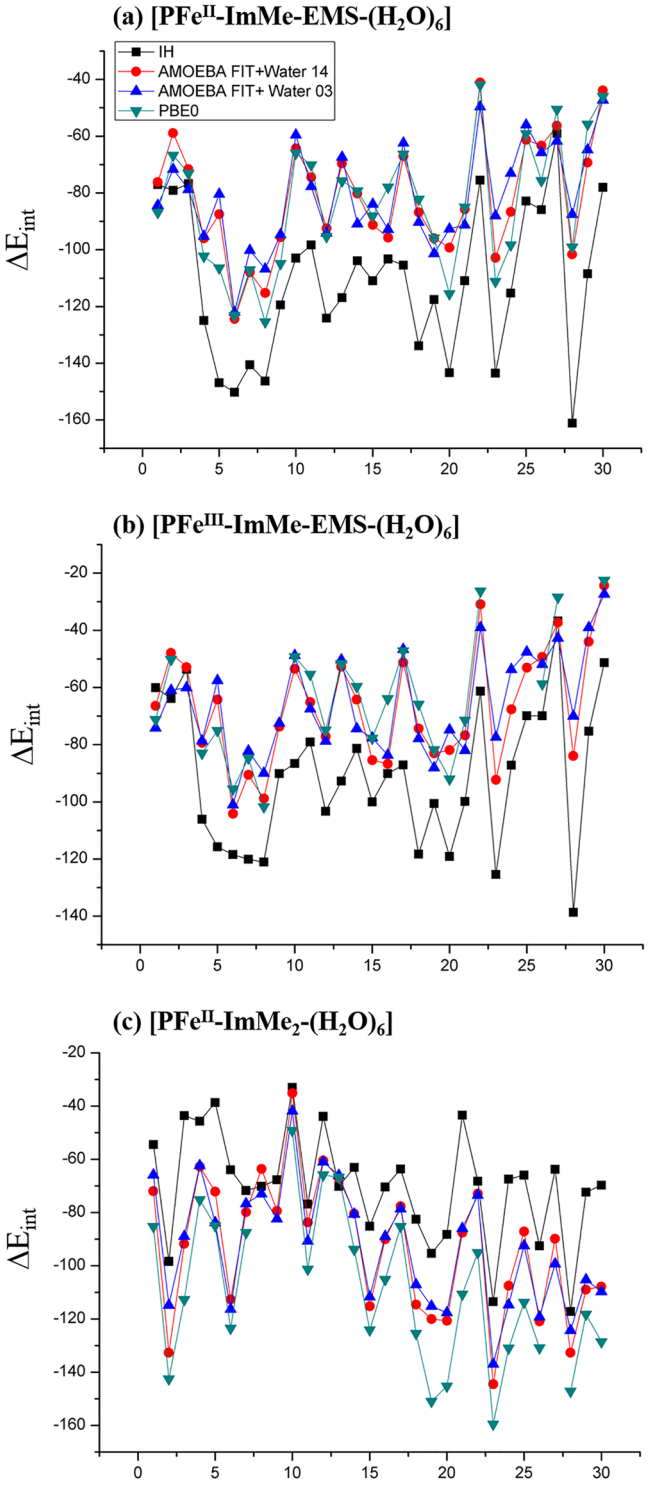


Figure 6. Interaction energies (kcal/mol) of the [PFe^{II}−ImMe−EMS−$(H_2O)_6$] (top), [PFe^{III}−ImMe−EMS−$(H_2O)_6$] (middle), and [PFe^{II}−$(ImMe)_2$−$(H_2O)_6$] (bottom) complexes with droplets of water molecules. For each complex, a collection of 30 geometries is considered.

a monopolar to a multipolar and polarizable description of the electrostatic potential created by the heme is beneficial for the accuracy of the force field. With the 2003 AMOEBA water model, the agreement is a little bit less satisfactory, with some points of the series departing more from the DFT results, but results are similar in nature and the set of investigated points limited. We emphasize that the same heme parameters are used in both series and that only the water model is changed. For the energies of present interest, the accuracy of the 2014 water model is superior. We note small differences between results obtained with either DMA or FIT sets of multipoles. Very similar conclusions can be drawn for the [PFe^{III}−ImMe−EMS−$(H_2O)_6$] complex with a better agreement of computed interaction energies with AMOEBA using the 2014 water model. Finally, for the [PFe^{II}−$(ImMe)_2$−$(H_2O)_6$] complex, we find that correlation between AMOEBA and DFT is much better than with nonpolarizable FF (R^2 = 0.94 vs 0.62). In summary, these calculations validate our parametrization approach for the heme complex. The computed interaction energies clearly indicate that the sets of multipoles are able to accurately reproduce the electrostatic potential created by the heme complex in both ferrous and ferric states.

III.D. Heme Coordination Patterns. Before moving in section IV to the simulation of hemeproteins with AMOEBA, we focus here on the inner-sphere coordination of the iron cation. We observed that in AMOEBA MD simulations of the [$PFe^{II/III}$−ImMe−EMS] complex in water, the (bi)squared pyramidal geometry around Fe is lost. This means that the sole presence of electrostatic multipoles on the heme atoms is not sufficient to define a proper coordination pattern for iron. This is somehow not surprising because the Fe−N or Fe−S bonds are more complex than resulting from a mere electrostatic interaction. We thus introduce supplementary terms in the potential energy between the iron atom and the two coordinating atoms in axial positions. To avoid the introduction of new atom classes in TINKER or TINKER-HP that would multiply the number of parameters, we have introduced the bonding and angle terms with the "Restrain" option of the program. A restraint term takes the form of a flat-welled harmonic potential. When we use the restrain option, we provide a force constant in kcal/Å^2 and two distances to define a distance range. If the distance between the two atoms under consideration is in this range, the energy from the restraint potential is zero. Outside the distance bounds, a standard harmonic term is used using the force constant and the interatomic distance. We have determined the parameters for the restraints by trial and error. We finally decided to restrain bonds between Fe and the coordinating atoms of the axial ligands (NE2 for histidine and SD for methionine) as well as the angles NE1−Fe−NE2 and NE1−Fe−SD where NE1 and NE2 are the porphyrin nitrogen atoms. The restraint parameters are given in Table S6.

Polarizable MD simulations have been performed with TINKER-HP (version 1.0).[49] The [PFe−ImMe−EMS] complex is immersed into a water box of edge length 24.875 Å. The Particle Mesh Ewald (PME) summation technique has been applied to treat electrostatic interactions, including polarization, with a real-space cutoff distance of 7.0 Å, a grid of 64 × 64 × 64, and fifth-order B-splines. A cutoff with a switching window at 10 Å has been applied to the van der Waals interactions. Induced dipoles have been iterated to convergence, until the Root-Mean-Square (RMS) change between interactions fell below 10^{-5} D per atom using the ASPC approach.[53] The trajectories have been propagated using a velocity form of Bernie Brook's "Better Beeman" integration algorithm with a 1.0 fs time step. The system temperature has been controlled with a Berendsen thermostat at 300 K.[86]

The average values and standard deviations of a few geometrical parameters of iron coordination are gathered in

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

Table 3. Bond Lengths (in Å), Angles (in degree), and Doming Effect Dihedrals for the reduced and oxidized (Italic) eme Structure[a]

parameters	451C/351C[b]	mean X-ray[c]	DFT	PM7[d]	AMOEBA DMA	AMOEBA FIT
$Fe-N_{his}$	1.97	2.03 ± 0.15	1.97	1.97 (0.04)	2.07 (0.03)	2.06 (0.03)
	1.99		1.99		*2.05 (0.03)*	*2.03 (0.03)*
$Fe-S_{met}$	2.35	2.31 ± 0.18	2.28	2.33 (0.05)	2.39 (0.03)	2.39 (0.03)
	2.36		2.31		*2.38 (0.03)*	*2.38 (0.03)*
$Fe-N_{por}$	2.03	2.00 ± 0.05	2.00	1.95 (0.04)	2.10 (0.03)	2.10 (0.03)
	2.03		1.99		*2.10 (0.03)*	*2.10 (0.03)*
$N_{his}-Fe-N_{por}$	91.19	90 ± 2	89.89	89.88 (3.23)	90.10 (2.60)	90.10 (2.61)
	90.89		90.24		*89.72 (2.61)*	*89.88 (2.57)*
$S_{met}-Fe-N_{por}$	88.80	90 ± 2	90.10	90.12 (3.78)	89.99 (2.76)	89.87 (2.79)
	89.08		89.75		*90.39 (2.81)*	*90.12 (2.85)*
C2−Fe−C2	174.58	174 ± 3	174.71	177.81 (2.99)	177.65 (2.21)	176.26 (2.17)
	174.76		167.20		*176.09 (2.29)*	*176.17 (2.16)*
doming	0.04	0.04 ± 0.02[e]	0.01	0.051 (0.025)	0.04 (0.03)	0.04 (0.03)
	0.03		0.00		*0.04 (0.03)*	*0.04 (0.03)*

[a]For the PM7 and AMOEBA MD simulations, we provide average values and standard deviations (in parentheses). [b]Analysis of cyt C551 in both oxidized form (PDB 351) and reduced form (PDB 451C) were performed with VMD program. [c]Average variation observed in seven high resolution X-ray structures from ref FELIX (PDB: 5cyt, 1qn2, 3c2c, lio3, 1hro, 1c2r, 1cxc). [d]MD simulations showed that the coordination between Fe and the apical ligand is unphysical with the PM7 method in the ferric state. We thus do not report PM7 data for this redox state. Average variation observed in seven high resolution X-ray structures from 27. [e]Analysis using the seven structures used in ref 27.

Table 3. The upper part of the table refers to the bond distances and angles that were specifically restrained to a target value, while the lower part refers to other relevant parameters. Data for both redox states are provided. For comparison, we also provide similar values from an X-ray structure of cytochrome *c*551. The coordination around heme in this protein is representative and agrees well with the mean value of seven high resolution X-ray structures of other cytochromes containing the same heme group.[27] We also provide values extracted from the DFT optimized structure on which we fitted the sets of multipoles (Table 3) and average values obtained from MD simulations with PM7 (see previous subsection and SI).

The $Fe-N_{his}$, $Fe-N_{por}$, and $Fe-S_{met}$ average bond lengths obtained with AMOEBA DMA amount to 2.07, 2.10, and 2.39 Å, respectively, in the ferrous state. Similar values are obtained in the ferric state. These values are slightly larger than those obtained from PM7MD simulations or DFT geometry optimizations but within the experimental uncertainty of X-ray structures. The fluctuations of bond length are found to be on the order of 0.03 Å, in close agreement with PM7 simulations and on the order of the expected values for this type of chemical bonding. It could be possible to further increase the restraint force constants to further reduce the gap with DFT values, but we found that very strong force constants were needed to achieve this goal, probably because of repulsion between the axial ligand and the heme fragment originating from the van der Waals and permanent multipoles. It is still also possible that the distances are too long due to the lack of charge transfer between the metal and the porphyrin in the AMOEBA potential. Moreover, increasing the force constants would reduce too much the amplitude of the bond length fluctuations. We therefore decided to retain the set of restraints given in Table S6. The angles around the iron cation are also well reproduced by the AMOEBA simulations when comparing to the experimental and theoretical references. We find few differences for all these parameters when using either the AMOEBA DMA or AMOEBA FIT sets of multipoles. Regarding the C2−Fe−C2 angle (Figure 1) and the out-of-plane deformation of Fe (doming), we find slight displacement of the cation from the average plane of the porphyrin nitrogen atoms. We investigated also the energy profile corresponding to the rotation of the $C_{his}-N_{his}-Fe-N_{por}$ dihedral angle (Figure S2). The global minimum of the profile is located at 0° both at DFT and at AMOEBA levels of theory. The secondary minima are at the same positions with AMOEBA and PBE. The global potential well depth is more pronounced with AMOEBA than in DFT. However, we expect the dihedral distribution during a MD simulation to have similar behavior in AMOEBA and in DFT, centered around 0°, but slightly broader in DFT. Finally, we have inspected an extensive list of complementary geometrical parameters and found excellent agreement between AMOEBA and DFT or X-ray data (Tables S7).

IV. APPLICATION TO MD SIMULATIONS OF CYTOCHROME

In the last section of this article, we report simulations of cytochromes in the ferrous and ferric states with the newly developed AMOEBA FF parameters using the Tinker-HP program.[49] The FIT parameter sets are used. We consider cytochrome b_5 of *Rattus norvegicus* (PDB code: 1B5A[87]). The prosthetic group of cytochrome b_5 consists of a heme core with the two axial ligands being histidine residue side chains (His39 and His63). The initial structure has been extracted from NMR data and has been solvated in a cubic water box (edge length of 100 Å) using the CHARMM package.[88] The simulated system is comprised of 97 858 atoms. The protonation state of the residues has been evaluated with the PROPKa server. Aspartate and glutamate residues are deprotonated, while all histidine residues are protonated on δ positions. The system has been neutralized by adding 10 sodium cations. We have subsequently added nine (Na^+,Cl^-) ion pairs to reach a salt concentration of 0.15 M. The NMR structure has been preliminarily prepared using the nonpolarizable c31 CHARMM.[88] To this end, we have first carried out 10 000 steps of energy minimization with NAMD.[89] Equilibration has then been reached running successive 20 ps MD with a progressive increase of the temperature from 50 to 298 K by steps of 25 K. During the

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

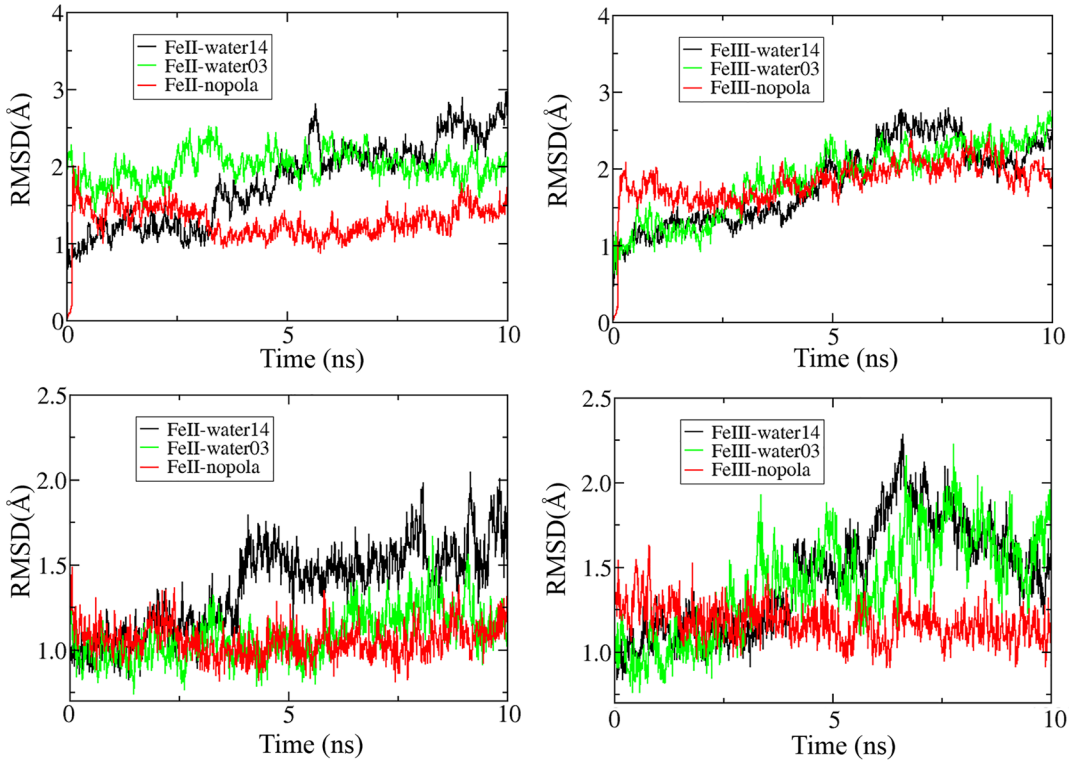


Figure 7. Upper panel, RMSD of protein backbone atoms. Lower panel, RMSD of protein backbone atoms belonging to alpha helices only. In all cases, the last seven residues that belong to a flexible loop were not included in the RMSD calculation.

Table 4. Bond Distance (Å) and Angle (degree) Values for Heme Structure[a]

			AMOEBA			
parameters	1B5A	DFT	Fe^{II} Water14	Fe^{III} Water14	Fe^{II} Water03	Fe^{III} Water03
$Fe-N_{his}$	2.00	1.97	2.07 (0.03)	2.05 (0.03)	2.07 (0.03)	2.05 (0.03)
$Fe-N_{por}$	2.00	2.01	2.10 (0.03)	2.11 (0.03)	2.10 (0.03)	2.11 (0.03)
$N_{his}-Fe-N_{por}$	90.03	90.00	90.00 (2.51)	89.99 (2.51)	90.00 (2.50)	89.86 (2.50)

[a]RMN structure. DFT optimized geometry and averaged over the trajectories of different MD simulations with the standard deviations (within parentheses).

heating, the protein heavy atoms have been kept fixed by adding harmonic restraints on their positions with force constants of 10 kcal/mol/Å^2. Subsequently, four successive MD simulations in the NPT ensemble have been conducted over 3 × 100 and 500 ps with restraint force constants of 10, 5, 1, and 0 kcal/mol/Å^2, respectively. A time step of 2 fs has been used. The final equilibrated structure is used as the starting geometry for AMOEBA polarizable simulations.

Simulations have been conducted in the two redox states and with either the water03 or water14 models. The same parameters as in the previous section are used, except that a RESPA integration algorithm allowing a 2.0 fs time step is used here. The temperature has been controlled by a Bussi thermostat.[90] The Particle Mesh Ewald (PME) summation technique has been applied to treat electrostatic interactions, including polarization, with a real-space cutoff distance of 7.0 Å, a grid of 64 × 64 × 64, and fifth-order B-splines. A cutoff with a switching window at 10 Å has been applied to the van der Waals interactions. Induced dipoles have been iterated to convergence, until the Root-Mean-Square (RMS) change between interactions fell below 10^{-5} D per atom and simulations used the ASPC approach. With the equilibrated structure obtained with the nonpolarizable FF, we have performed another equilibration procedure with AMOEBA parameters. For this, we have first fixed the hemeprotein, and have done energy minimization with a convergence value of 0.5 kcal mol^{-1} Å^{-1}. Equilibration has been obtained running successive 50 ps NVT MD with a progressive increase of the temperature from 50 to 300 K by steps of 50 K keeping hemeprotein fixed. Then, we have carried out another equilibration with the hemeprotein atom free to move. Finally, we have conducted an 80 ps NPT simulation at 300 K and 1 atm to obtain a stable density of the system (higher than 1 gr/cm^3). A 10 ns MD simulation in the NVT ensemble has been accumulated for analysis. Restraints on angles and bonds involving the iron cation have been set as explained in the previous section. For comparisons, we also carried out MD simulations with a nonpolarizable force field (CHARMM) using the NAMD program.[89]

The RMSDs of the protein backbone (focusing on residues 1 to 87) during MD simulation dynamics are shown in Figure 7. For each simulation, the RMSD value is below 3 Å. When focusing on amino acid residues pertaining to structured alpha helices, the RMSDs are, as expected, smaller. We have not noticed any unfolding of tertiary secondary structures during the simulations. Compared to simulations with nonpolarizable

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

FF, the RMSDs show more slight fluctuations but remain on the same order. In regard to the heme coordination sphere, AMOEBA gives good agreement for bond lengths and angles with NMR and DFT data (Table 4, Table S8). The force constants adjusted to define the iron cation inner-sphere coordination are transferable to the heme embedded into the protein matrix. Overall, these data indicate stable dynamics with the here developed parameters in both ferrous and ferric forms. The atom RMSDs for amino acid residues localized within 10 Å are shown in Figure S3. They are examined by types of residues (polar, charged, apolar...). Again, we have not found any conformational changes of theses residues that would be suspicious. However, conformation fluctuations are often more pronounced for the AMOEBA force field, probably because of the more flexible form of the underlying potential energy function. Future works will have to consider much longer simulations to assess the stability of hemoproteins and to investigate whether the herein developed AMOEBA force field for heme leads to more accurate redox properties than standard force fields.

These MD simulations have been run using 1440 2.6-GHz processors connected via infinite band technology (IB 4x FDR). The total wall clock time was approximately 12 h/ns of simulation on the Occigen machine at CINES (Montpellier, France). We mention that recent algorithmic developments of Tinker-HP should further improve efficiency. MD simulation of hemeproteins with AMOEBA is now computationally feasible in reasonable times.

V. CONCLUSION

In this article, we have reported to our knowledge the first parameters of heme for the polarizable AMOEBA force field. They have been derived both for the ferrous and ferric forms. Extensive validation has been obtained from calculations of interaction energies with series of small molecules of biological interest and on large water aggregates comprising around 80 water molecules. We have found that the description of the electrostatic interaction with the heme is greatly improved with AMOEBA. This is especially noticeable regarding interaction energies with water aggregates for which the correlation with DFT results outperforms that of a nonpolarizable model. That said, we have also found the model for water to have strong influence on the accuracy of the computed interaction energies. The 2014 water model has been found to be in better agreement with the reported DFT reference calculations, although results obtained with the 2003 potential are similar in nature. We have finally reported MD simulation of a cytochrome using the AMOEBA force field.

The availability of parameters for heme combined with the advent of highly efficient implementation of AMOEBA in the TINKER-HP software opens the door to wide applications of MD simulations of hemeproteins with AMOEBA. The herein devised parameters are freely available in the SI or upon request to the authors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b01128.

Computational protocol for PM7 simulations, tables of interaction energies for all the structures, restraint parameters for the inner-sphere coordination of the iron cation, AMOEBA multipoles for heme, autocorrelation function of AMOEBA interaction energies of heme with water droplets, energy profile associated to the rotation around the Nhis-Fe bond, RMSD of aminoacid residues during MD simulations of cytochrome (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: carine.clavaguera@u-psud.fr.
*E-mail: aurelien.de-la-lande@u-psud.fr.

ORCID

Carine Clavaguera: 0000-0001-5531-2333
Jean-Philip Piquemal: 0000-0001-6615-9426
Aurélien de la Lande: 0000-0003-0745-4171

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Calculations were carried out at the French computer center CINES, France. We greatly acknowledge GENCI for providing us with generous computational resources (project number A0020706913).

■ REFERENCES

(1) Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **2014**, *114* (7), 3919−3962.
(2) Smith, L. J.; Kahraman, A.; Thornton, J. M. Heme proteins—Diversity in structural characteristics, function, and folding. *Proteins: Struct., Funct., Genet.* **2010**, *78* (10), 2349−2368.
(3) Loew, G. H.; Harris, D. L. Role of the Heme Active Site and Protein Environment in Structure, Spectra, and Function of the Cytochrome P450s. *Chem. Rev.* **2000**, *100* (2), 407−420.
(4) Van Doorslaer, S. Understanding heme proteins with hyperfine spectroscopy. *J. Magn. Reson.* **2017**, *280* (Supplement C), 79−88.
(5) Prytkova, T. R.; Kurnikov, I. V.; Beratan, D. N. Coupling Coherence Distinguishes Structure Sensitivity in Protein Electron Transfer. *Science* **2007**, *315* (5812), 622−625.
(6) Keinan, S.; Nocek, J. M.; Hoffman, B. M.; Beratan, D. N. Interfacial hydration, dynamics and electron transfer: multi-scale ET modeling of the transient [myoglobin, cytochrome b5] complex. *Phys. Chem. Chem. Phys.* **2012**, *14* (40), 13881−13889.
(7) Blumberger, J.; Klein, M. L. Reorganization Free Energies for Long-Range Electron Transfer in a Porphyrin-Binding Four-Helix Bundle Protein. *J. Am. Chem. Soc.* **2006**, *128* (42), 13854−13867.
(8) Tipmanee, V.; Oberhofer, H.; Park, M.; Kim, K. S.; Blumberger, J. Prediction of Reorganization Free Energies for Biological Electron Transfer: A Comparative Study of Ru-Modified Cytochromes and a 4-Helix Bundle Protein. *J. Am. Chem. Soc.* **2010**, *132* (47), 17032−17040.
(9) Breuer, M.; Zarzycki, P.; Blumberger, J.; Rosso, K. M. Thermodynamics of Electron Flow in the Bacterial Deca-heme Cytochrome MtrF. *J. Am. Chem. Soc.* **2012**, *134* (24), 9868−9871.
(10) Breuer, M.; Rosso, K. M.; Blumberger, J. Electron flow in multiheme bacterial cytochromes is a balancing act between heme electronic interaction and redox potentials. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (2), 611−616.
(11) Derat, E.; Shaik, S.; Rovira, C.; Vidossich, P.; Alfonso-Prieto, M. The Effect of a Water Molecule on the Mechanism of Formation of Compound 0 in Horseradish Peroxidase. *J. Am. Chem. Soc.* **2007**, *129* (20), 6346−6347.
(12) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes. *Chem. Rev.* **2005**, *105* (6), 2279−2328.

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

(13) Mao, J.; Hauser, K.; Gunner, M. R. How Cytochromes with Different Folds Control Heme Redox Potentials. *Biochemistry* **2003**, *42* (33), 9829−9840.
(14) Song, Y.; Mao, J.; Gunner, M. R. Electrostatic Environment of Hemes in Proteins: pKas of Hydroxyl Ligands. *Biochemistry* **2006**, *45* (26), 7949−7958.
(15) Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **2015**, *115* (20), 11191−11238.
(16) Olsson, M. H. M.; Hong, G.; Warshel, A. Frozen Density Functional Free Energy Simulations of Redox Proteins: Computational Studies of the Reduction Potential of Plastocyanin and Rusticyanin. *J. Am. Chem. Soc.* **2003**, *125* (17), 5025−5039.
(17) Parson, W. W.; Chu, Z. T.; Warshel, A. Reorganization Energy of the Initial Electron-Transfer Step in Photosynthetic Bacterial Reaction Centers. *Biophys. J.* **1998**, *74* (1), 182−191.
(18) Blumberger, J. Free energies for biological electron transfer from QM/MM calculation: method, application and critical assessment. *Phys. Chem. Chem. Phys.* **2008**, *10* (37), 5651−5667.
(19) de la Lande, A.; Gillet, N.; Chen, S.; Salahub, D. R. Progress and challenges in simulating and understanding electron transfer in proteins. *Arch. Biochem. Biophys.* **2015**, *582*, 28−41.
(20) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3* (6), 2034−2045.
(21) Warshel, A.; Parson, W. W. Computer Simulations of Electron-Transfer Reactions in Solution and in Photosynthetic Reaction Centers. *Annu. Rev. Phys. Chem.* **1991**, *42* (1), 279−309.
(22) Tipmanee, V.; Blumberger, J. Kinetics of the Terminal Electron Transfer Step in Cytochrome c Oxidase. *J. Phys. Chem. B* **2012**, *116* (6), 1876−1883.
(23) Oda, A.; Yamaotsu, N.; Hirono, S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **2005**, *26* (8), 818−826.
(24) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012**, *33* (2), 119−133.
(25) Kaszuba, K.; Postila, P. A.; Cramariuc, O.; Sarewicz, M.; Osyczka, A.; Vattulainen, I.; Róg, T. Parameterization of the prosthetic redox centers of the bacterial cytochrome bc 1 complex for atomistic molecular dynamics simulations. *Theor. Chem. Acc.* **2013**, *132* (6), 1370.
(26) Favia, A. D.; Cavalli, A.; Masetti, M.; Carotti, A.; Recanatini, M. Three-dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protoporphyrin IX for a molecular dynamics study of heme-cysteinato cytochromes. *Proteins: Struct., Funct., Genet.* **2006**, *62* (4), 1074−1087.
(27) Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.* **2004**, *25* (13), 1613−1622.
(28) Johansson, M. P.; Kaila, V. R. I.; Laakkonen, L. Charge parameterization of the metal centers in cytochrome c oxidase. *J. Comput. Chem.* **2008**, *29* (5), 753−767.
(29) Adam, S.; Knapp-Mohammady, M.; Yi, J.; Bondar, A.-N. Revised CHARMM force field parameters for iron-containing cofactors of photosystem II. *J. Comput. Chem.* **2018**, *39*, 7.
(30) Muegge, I.; Qi, P. X.; Wand, A. J.; Chu, Z. T.; Warshel, A. The Reorganization Energy of Cytochrome c Revisited. *J. Phys. Chem. B* **1997**, *101* (5), 825−836.
(31) Ponder, J. *TINKER - Software Tools for Molecular Design*, version 7.1.2.
(32) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933−5947.
(33) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119* (29), 9423−9437.
(34) Semrouni, D.; Isley, W. C.; Clavaguéra, C.; Dognon, J.-P.; Cramer, C. J.; Gagliardi, L. Ab Initio Extension of the AMOEBA Polarizable Force Field to Fe2+. *J. Chem. Theory Comput.* **2013**, *9* (7), 3062−3071.
(35) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125* (50), 15671−15682.
(36) Clavaguéra, C.; Calvo, F.; Dognon, J.-P. Theoretical study of the hydrated Gd3+ ion: Structure, dynamics, and charge transfer. *J. Chem. Phys.* **2006**, *124* (7), 074505.
(37) Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J. Chem. Phys.* **2006**, *125* (5), 054511.
(38) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6* (7), 2059−2070.
(39) 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. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549−2564.
(40) Satpati, P.; Clavaguéra, C.; Ohanessian, G.; Simonson, T. Free Energy Simulations of a GTPase: GTP and GDP Binding to Archaeal Initiation Factor 2. *J. Phys. Chem. B* **2011**, *115* (20), 6749−6763.
(41) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8* (4), 1314−1324.
(42) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143−3161.
(43) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046−4063.
(44) Lamoureux, G.; Roux, B. t. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119* (6), 3025−3039.
(45) Blumberger, J.; Lamoureux, G. Reorganization free energies and quantum corrections for a model electron self-exchange reaction: comparison of polarizable and non-polarizable solvent models. *Mol. Phys.* **2008**, *106* (12−13), 1597−1611.
(46) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11* (6), 2609−2618.
(47) Narth, C.; Lagardère, L.; Polack, É.; Gresh, N.; Wang, Q.; Bell, D. R.; Rackers, J. A.; Ponder, J. W.; Ren, P. Y.; Piquemal, J.-P. Scalable improvement of SPME multipolar electrostatics in anisotropic polarizable molecular mechanics using a general short-range penetration correction up to quadrupoles. *J. Comput. Chem.* **2016**, *37* (5), 494−506.
(48) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An optimized charge penetration model for use with the AMOEBA force field. *Phys. Chem. Chem. Phys.* **2017**, *19* (1), 276−291.
(49) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9* (4), 956−972.
(50) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014**, *10* (4), 1638−1651.

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

(51) Lagardère, L.; Lipparini, F.; Polack, É.; Stamm, B.; Cancès, É.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II. Toward Massively Parallel Computations Using Smooth Particle Mesh Ewald. *J. Chem. Theory Comput.* **2015**, *11* (6), 2589−2599.
(52) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111* (23), 8551−8566.
(53) Kolafa, J. Time-reversible always stable predictor−corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.* **2004**, *25* (3), 335−342.
(54) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59* (3), 341−350.
(55) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1138.
(56) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126* (14), 144111.
(57) Hoe, W.-M.; Cohen, A. J.; Handy, N. C. Assessment of a new local exchange functional OPTX. *Chem. Phys. Lett.* **2001**, *341* (3), 319−328.
(58) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865−3868.
(59) Calaminici, P.; Janetzko, F.; Köster, A. M.; Mejia-Olvera, R.; Zuniga-Gutierrez, B. Density functional theory optimized basis sets for gradient corrected functionals: 3d transition metal systems. *J. Chem. Phys.* **2007**, *126* (4), 044108.
(60) Koster, A. M.; Calaminici, P.; Casida, M. E.; Dominguez, V. D.; Flores-Moreno, R.; Geudtner, G.; Goursot, A.; Heine, T.; Ipatov, A.; Janetzko, F.; del Campo, J. M.; Reveles, J. U.; Vela, A.; Zuniga-Gutierrez, B.; Salahub, D. R. *deMon2k*, version 2; The deMon Developers: Mexico City, 2011.
(61) Abdurahman, A.; Renger, T. Density Functional Studies of Iron-Porphyrin Cation with Small Ligands X (X: O, CO, NO, O2, N2, H2O, N2O, CO2). *J. Phys. Chem. A* **2009**, *113* (32), 9202−9206.
(62) Conradie, J.; Ghosh, A. Electronic Structure of Trigonal-Planar Transition-Metal−Imido Complexes: Spin-State Energetics, Spin-Density Profiles, and the Remarkable Performance of the OLYP Functional. *J. Chem. Theory Comput.* **2007**, *3* (3), 689−702.
(63) Köster, A. M.; Reveles, J. U.; del Campo, J. M. Calculation of exchange-correlation potentials with auxiliary function densities. *J. Chem. Phys.* **2004**, *121* (8), 3417−3424.
(64) 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.; Zheng, G.; Sonnenberg, J. L.; 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. O.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Ayengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(65) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005**, *1* (6), 1128−1132.
(66) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. *Orient: a program for studying interactions between molecules*, version 4.5; University of Cambridge, 2002.
(67) Stone, A. J. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981**, *83* (2), 233−239.
(68) Stone, A. J.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56* (5), 1047−1064.
(69) Semrouni, D.; Sharma, A.; Dognon, J.-P.; Ohanessian, G.; Clavaguéra, C. Finite Temperature Infrared Spectra from Polarizable Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10* (8), 3190−3199.
(70) Ermler, U.; Siddiqui, R. A.; Cramm, R.; Friedrich, B. Crystal structure of the flavohemoglobin from Alcaligenes eutrophus at 1.75 A resolution. *EMBO J.* **1995**, *14* (24), 6067−6077.
(71) El Hammi, E.; Houee-Levin, C.; Rezac, J.; Levy, B.; Demachy, I.; Baciou, L.; de la Lande, A. New insights into the mechanism of electron transfer within flavohemoglobins: tunnelling pathways, packing density, thermodynamic and kinetic analyses. *Phys. Chem. Chem. Phys.* **2012**, *14* (40), 13872−13880.
(72) Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. Structure at 2.8 Å resolution of cytochrome c oxidase from Paracoccus denitrificans. *Nature* **1995**, *376*, 660.
(73) Becke, A. D. A new mixing of Hartree−Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98* (2), 1372−1377.
(74) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4* (2), 91−100.
(75) Köster, A. M.; Flores-Moreno, R.; Reveles, J. U. Efficient and reliable numerical integration of exchange-correlation energies and potentials. *J. Chem. Phys.* **2004**, *121* (2), 681−690.
(76) Mejía-Rodríguez, D.; Köster, A. M. Robust and efficient variational fitting of Fock exchange. *J. Chem. Phys.* **2014**, *141* (12), 124114.
(77) de la Lande, A.; Clavaguéra, C.; Köster, A. On the accuracy of population analyses based on fitted densities. *J. Mol. Model.* **2017**, *23* (4), 99.
(78) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(79) Řezáč, J. Cuby: An integrative framework for computational chemistry. *J. Comput. Chem.* **2016**, *37* (13), 1230−1237.
(80) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110* (13), 6158−6170.
(81) Yuan, D.; Li, Y.; Ni, Z.; Pulay, P.; Li, W.; Li, S. Benchmark Relative Energies for Large Water Clusters with the Generalized Energy-Based Fragmentation Method. *J. Chem. Theory Comput.* **2017**, *13* (6), 2696−2704.
(82) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128* (8), 084106.
(83) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125* (23), 234109.
(84) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
(85) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2* (2), 364−382.
(86) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(87) Dangi, B.; Sarma, S.; Yan, C.; Banville, D. L.; Guiles, R. D. The Origin of Differences in the Physical Properties of the Equilibrium Forms of Cytochrome b5 Revealed through High-Resolution NMR Structures and Backbone Dynamic Analyses. *Biochemistry* **1998**, *37* (23), 8289−8302.
(88) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch,

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.

(89) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.

(90) Bussi, G.; Parrinello, M. Stochastic thermostats: comparison of local and global schemes. *Comput. Phys. Commun.* **2008**, *179* (1), 26−29.

DOI: 10.1021/acs.jctc.7b01128
J. Chem. Theory Comput. 2018, 14, 2705−2720

AMOEBA POLARIZABLE FORCE FIELD PARAMETERS OF THE HEME COFACTOR IN ITS FERROUS AND FERRIC FORMS

Xiaojing Wu[1], Carine Clavaguera[1]*, Louis Lagardère[2,3], Jean-Philip Piquemal [4,5], Aurélien de la Lande[1]*

1: Laboratoire de Chimie Physique, Université Paris Sud - CNRS. Université Paris Saclay, 15, avenue Jean Perrin, 91405 Orsay Cedex, France.

2: Sorbonne Université, CNRS, Institut Parisien de Chimie Physique et Théorique (IP2CT), 4 place Jussieu, F-75005, Paris, France.

3. Sorbonne Université, Institut des Sciences du Calcul et des Données (ISCD), 4 place Jussieu, F-75005, Paris, France.

4. Sorbonne Université, CNRS, Laboratoire de Chimie Théorique (LCT), 4 place Jussieu, F-75005, Paris, France.

5 : Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States.

Supporting information

Table S1: Iterative Hirshfeld charges summed over fragments for the [PFe$^{II/III}$-ImMe-EMS] and [PFe$^{II/III}$-ImMe2] complexes (OPTX functionam)

[PFe$^{II/III}$-ImMe-EMS]	ferrous	ferric
Porphyrin	-1.72	-1.05
Fe	1.60	1.71
EMS	0.07	0.14
ImMe	0.05	0.20
[PFe$^{II/III}$-ImMe$_2$]		
Porphyrin	-1.73	-1.05
Fe	1.69	1.79
ImMe	0.02	0.13
ImMe	0.02	0.13

Table S2: Interaction energy of heme and single molecules in kcal/mol obtained at the various computational levels.

	IH	AMOEBA				MP2	B3LYP
		DMA Water 14	FIT Water 14	DMA Water 03	FIT Water 03		
HEME(FeII) – H_2O							
A	-15.36	-14.11	-14.69	-11.26	-11.80	-15.98	-16.01
B	-6.82	-6.53	-6.64	-6.56	-6.63	-7.66	-7.55
C	-2.06	-1.76	-1.78	-1.80	-1.81	-2.06	-1.87
D	-1.27	-1.08	-1.09	-1.11	-1.11	-1.28	-1.14
E	-6.66	-6.13	-5.68	-6.20	-5.70	-7.80	-6.77
F	-3.13	-3.16	-3.11	-3.12	-3.12	-5.15	-4.67
G	-20.39	-17.19	-17.00	-15.89	-15.98	-20.81	-19.33
HEME(FeIII) – H_2O							
A	-13.73	-12.52	-12.85	-9.97	-10.32	-13.80	nc.[a]
B	-5.70	-5.60	-5.61	-5.65	-5.66	-6.30	nc.
C	-1.46	-1.26	-1.25	-1.27	-1.26	-1.40	nc.
D	-0.84	-0.70	-0.70	-0.71	-0.71	nc.	nc.
E	-6.06	-5.94	-5.54	-5.89	-5.50	-5.20	nc.
F	-0.95	-1.61	-1.64	-1.50	-1.54	-3.80	nc.
G	-21.00	-17.64	-17.27	-17.19	-16.90	-21.00	nc.
HEME(FeII) – Na^+							
A	-168.54	-146.98	-148.28	-	-	-166.15	nc.
B	-135.29	-124.56	-125.33	-	-	-134.42	nc.
C	-79.85	-73.92	-74.24	-	-	-75.61	nc.
D	-63.69	-59.49	-59.72	-	-	-63.25	nc.
E	-63.88	-107.79	-106.87	-	-	-121.70	nc.
HEME(FeIII) – Na^+							
A	-131.89	-114.89	-114.83	-	-	-126.70	nc.
B	-101.79	-94.60	-94.31	-	-	-97.50	nc.
C	-54.04	-49.80	-49.61	-	-	-46.60	nc.
D	-41.24	-38.22	-38.09	-	-	-32.00	nc.
E	-0.39	-48.93	-49.47	-	-	-56.00	nc.
HEME(FeII) – Cl^-							
A	58.84	34.67	35.13	-	-	36.60	40.30
B	44.44	39.99	40.41	-	-	39.20	38.89
C	45.33	45.15	45.19	-	-	45.20	44.93
D	80.40	59.53	59.67	-	-	49.38	54.00
HEME(FeIII) – Cl^-							
A	13.29	-5.83	-4.87	-	-	-12.31	nc.
B	5.68	2.26	3.07	-	-	-1.19	nc.
C	16.42	16.99	17.23	-	-	16.05	nc.
D	41.26	14.99	14.46	-	-	8.33	nc.
HEME(FeII) – $CH_3CH_2NH_3^+$							

A	-140.64	-132.66	-131.89	-	-	-141.70	-140.45
B	-106.48	-97.32	-97.30	-	-	-103.30	-101.58
C	-69.31	-63.48	-63.58	-	-	-65.50	-72.90
D	-42.11	-89.48	-83.82	-	-	-94.20	nc.
HEME(FeII) – $CH_3CH_2NH_3^+$							
A	-106.75	-101.82	-99.90	-	-	-103.90	nc.
B	-76.40	-69.89	-69.10	-	-	-69.70	nc.
C	-45.54	-41.29	-41.00	-	-	-38.70	nc.
D	18.82	-30.12	-27.06	-	-	-28.10	nc.
HEME(FeII) – CH_3COO^-							
A	-140.64	-132.66	-131.89	-	-	-141.70	-140.45
B	-106.48	-97.32	-97.30	-	-	-103.30	-101.58
C	-69.31	-63.48	-63.58	-	-	-65.50	-72.90
D	-42.11	-89.48	-83.82	-	-	-94.20	nc.
HEME(FeIII) – CH_3COO^-							
A	-106.75	-101.82	-99.90	-	-	-103.90	nc.
B	-76.40	-69.89	-69.10	-	-	-69.70	nc.
C	-45.54	-41.29	-41.00	-	-	-38.70	nc.
D	18.82	-30.12	-27.06	-	-	-28.10	nc.
HEME(FeII) – Phenol							
A	-17.11	-12.28	-12.59	-	-	-13.01	-19.32
B	-10.64	-7.41	-7.51	-	-	-4.50	-11.04
C	-3.58	-1.98	-1.99	-	-	3.90	-2.73
HEME(FeIII) – Phenol							
A	-15.64	-10.85	-10.98	-	-	nc.	nc.
B	-9.44	-6.43	-6.43	-	-	nc.	nc.
C	-2.89	-1.54	-1.53	-	-	nc.	nc.

a: not computed (SCF convergence issues or obviously unreliable values)

Table S3: Average interaction energies (kcal/mol) for the [PFe^{III}ImMe-EMS-$(H_2O)_6$] complex with two convergence criteria.

	IH	AMOEBA			
[PFe^{III}ImMe-EMS-$(H_2O)_6$]		DMA Water14	FIT Water14	DMA Water03	FIT Water 03
SCF converge 10^{-4} Ha					
$\langle \Delta IE \rangle$	-22.84	-6.91	-5.44	-5.93	-4.25
SCF converge 10^{-6} Ha					
$\langle \Delta IE \rangle$	-22.80	-6.88	-5.40	-5.89	-4.21
difference					
$\langle \Delta IE \rangle$	-0.04	-0.04	-0.04	-0.04	-0.04

Table S4: Average interaction energies (kcal/mol), standard deviations, RMSE and linear correlation coefficient between DFT (PBE and B3LYP) results and force field results.

	IH	AMOEBA			
[PFe^{II}ImMe-EMS-$(H_2O)_6$]		DMA Water14	FIT Water14	DMA Water03	FIT Water 03
PBE					
$\langle \Delta IE \rangle$	-23.65	3.66	4.98	4.61	6.00
$\sigma_{(\Delta IE)}$	15.03	8.20	7.92	11.68	11.41
RMSE	27.89	8.86	9.25	12.38	12.72
R^2	0.69	0.87	0.88	0.73	0.75
B3LYP					
$\langle \Delta IE \rangle$	-55.38	-28.46	-27.09	-27.55	-26.11
$\sigma_{(\Delta IE)}$	17.04	9.24	8.93	11.26	10.99
RMSE	57.85	29.87	28.47	29.68	28.26
R^2	0.61	0.84	0.85	0.77	0.79

Table S5: Interaction energy of [PFe$^{II/III}$-ImMe-EMS] and [PFe$^{II/III}$-ImMe2] complexes with water molecules in kcal/mol.

PFe^{II}-ImMe-EMS-(H_2O)$_6$]						
	IH		AMOEBA			PBE0
		DMA Water 14	FIT Water 14	DMA Water 03	FIT Water 03	
1	-77.07	-79.34	-76.19	-87.46	-84.33	-86.91
2	-79.13	-61.12	-58.90	-74.35	-71.69	-66.69
3	-76.70	-72.51	-71.59	-79.42	-78.80	-73.09
4	-124.92	-98.02	-95.99	-96.95	-95.27	-102.31
5	-146.95	-89.59	-87.53	-82.13	-80.40	-106.39
6	-150.24	-128.25	-124.45	-125.41	-121.97	-123.20
7	-140.61	-107.17	-107.98	-99.04	-100.23	-107.09
8	-146.38	-118.36	-115.27	-110.05	-106.72	-125.44
9	-119.50	-96.12	-95.61	-95.40	-94.66	-104.85
10	-102.98	-65.79	-64.32	-61.36	-59.55	-66.00
11	-98.35	-76.94	-74.49	-80.26	-77.74	-70.12
12	-124.12	-96.23	-92.53	-98.23	-94.16	-95.39
13	-116.94	-71.70	-69.64	-69.76	-67.41	-75.77
14	-103.89	-79.86	-80.23	-90.44	-90.90	-79.24
15	-110.90	-91.71	-91.30	-84.54	-83.91	-88.10
16	-103.29	-99.77	-95.77	-97.33	-92.83	-77.96
17	-105.41	-69.34	-67.05	-65.02	-62.35	-66.33
18	-133.86	-86.95	-86.83	-90.73	-90.30	-82.19
19	-117.55	-95.86	-96.02	-101.28	-101.38	-95.92
20	-143.39	-98.09	-99.28	-91.68	-92.74	-115.52
21	-110.94	-85.67	-85.82	-91.35	-91.29	-84.97
22	-75.46	-42.14	-41.10	-50.78	-49.63	-41.95
23	-143.59	-103.40	-102.84	-88.68	-88.05	-111.22
24	-115.27	-87.33	-86.69	-73.19	-73.03	-98.38
25	-82.88	-64.08	-61.25	-58.76	-55.99	-59.14
26	-85.89	-64.95	-63.31	-67.12	-65.74	-75.66
27	-59.03	-57.83	-56.28	-63.42	-61.77	-50.58
28	-161.13	-103.24	-101.67	-89.36	-87.56	-99.10
29	-108.48	-70.01	-69.29	-65.62	-64.78	-55.72
30	-78.03	-41.42	-43.87	-44.90	-47.24	-46.02
PFe^{III}-ImMe-EMS-(H_2O)$_6$]						
	IH		AMOEBA			PBE0
		DMA Water 14	FIT Water 14	DMA Water 03	FIT Water 03	
1	-60.11	-68.51	-66.40	-76.16	-74.20	-71.25
2	-63.87	-47.70	-47.93	-61.11	-61.01	-50.16
3	-53.62	-52.89	-52.89	-59.84	-60.12	--
4	-106.11	-81.40	-79.36	-80.66	-78.67	-83.00
5	-115.77	-65.99	-64.22	-59.32	-57.55	-75.05
6	-118.45	-108.38	-104.19	-105.15	-101.05	-95.55
7	-120.09	-90.83	-90.54	-82.46	-82.20	-84.74
8	-121.09	-101.31	-98.88	-92.88	-89.94	-101.77
9	-90.10	-73.74	-73.71	-72.87	-72.45	--
10	-86.63	-53.69	-53.46	-49.57	-48.76	-49.30
11	-79.11	-65.82	-65.10	-68.21	-67.49	-55.40
12	-103.32	-79.62	-77.15	-81.59	-78.78	-74.95
13	-92.77	-55.11	-52.69	-53.04	-50.32	-51.75
14	-81.32	-63.48	-64.25	-73.57	-74.40	-59.70
15	-100.04	-86.22	-85.39	-78.71	-77.65	-77.79
16	-90.11	-89.70	-86.70	-87.28	-83.66	-63.97
17	-87.09	-53.48	-51.26	-49.24	-46.66	-47.19
18	-118.29	-74.36	-74.35	-77.86	-77.87	-65.88
19	-100.60	-82.10	-83.01	-87.19	-88.10	-81.78
20	-119.12	-81.21	-81.89	-74.40	-74.82	-92.00
21	-99.86	-78.18	-76.72	-83.74	-81.98	-71.46
22	-61.31	-31.10	-30.92	-39.38	-39.03	-26.32
23	-125.45	-93.52	-92.30	-79.03	-77.37	--
24	-87.20	-68.32	-67.68	-54.35	-53.68	--
25	-69.89	-56.00	-53.07	-50.61	-47.59	--
26	-69.92	-53.43	-49.28	-55.67	-51.89	-58.76
27	-36.72	-39.98	-37.23	-45.57	-42.71	-28.39
28	-138.66	-87.73	-83.97	-74.11	-70.02	--
29	-75.31	-44.03	-43.99	-39.39	-39.03	--
30	-51.33	-23.18	-24.36	-26.38	-27.39	-22.49
PFe^{II}-ImMe$_2$-(H_2O)$_6$						
	IH		AMOEBA			PBE0
		DMA Water 14	FIT Water 14	DMA Water 03	FIT Water 03	
1	-54.44	-75.75	-71.95	-69.96	-65.81	-85.24
2	-98.38	-135.02	-132.64	-117.42	-114.79	-142.56
3	-43.58	-93.56	-91.81	-90.84	-88.97	-112.67
4	-45.66	-63.11	-62.77	-62.07	-62.30	-75.15
5	-38.71	-72.14	-72.17	-84.06	-83.77	-84.96
6	-63.92	-115.03	-112.68	-119.04	-116.32	-123.53
7	-71.76	-79.35	-79.84	-76.20	-76.67	-87.51
8	-70.14	-65.60	-63.60	-74.99	-72.99	--
9	-67.74	-78.46	-79.39	-81.43	-82.32	--
10	-32.93	-38.80	-35.10	-45.73	-41.81	-49.13
11	-76.79	-84.16	-83.70	-91.31	-90.73	-101.24
12	-43.87	-62.14	-60.47	-62.31	-60.97	-65.81
13	-70.15	-68.01	-66.99	-67.16	-65.99	-66.73
14	-63.08	-83.62	-80.38	-83.71	-80.62	-93.83
15	-85.21	-116.92	-115.17	-113.29	-111.62	-124.09

16	-70.40	-92.93	-90.05	-91.89	-88.96	-105.17
17	-63.63	-78.67	-77.69	-79.27	-78.59	-85.22
18	-82.56	-116.32	-114.61	-108.61	-107.11	-125.43
19	-95.34	-120.49	-119.99	-115.41	-115.18	-150.94
20	-88.27	-122.03	-120.61	-119.13	-117.59	-145.23
21	-43.38	-90.46	-87.57	-88.86	-85.92	-110.66
22	-68.26	-74.19	-72.94	-74.77	-73.45	-95.06
23	-113.48	-146.78	-144.53	-139.42	-136.95	-159.57
24	-67.39	-109.87	-107.42	-117.12	-114.66	-130.94
25	-65.98	-89.41	-87.14	-94.88	-92.42	-113.78
26	-92.50	-122.78	-120.94	-121.51	-119.18	-130.84
27	-63.72	-92.49	-89.84	-101.95	-99.28	--
28	-117.19	-135.92	-132.68	-127.76	-124.32	-147.19
29	-72.38	-110.06	-108.93	-106.74	-105.19	-118.26
30	-69.69	-109.74	-107.79	-112.33	-109.74	-128.54

PFe^{II}-ImMe-EMS-$(H_2O)_6$]

	PBE0	B3LYP
1	-86.91	-56.70
2	-66.69	-36.58
3	-73.09	-48.78
4	-102.31	-71.45
5	-106.39	-72.58
6	-123.20	-99.94
7	-107.09	-78.62
8	-125.44	-95.71
9	-104.85	-75.35
10	-66.00	-35.69
11	-70.12	--
12	-95.39	-66.98
13	-75.77	-47.35
14	-79.24	-49.59
15	-88.10	-59.04
16	-77.96	-48.39
17	-66.33	-35.56
18	-82.19	-50.63
19	-95.92	-71.27
20	-115.52	--
21	-84.97	-73.87
22	-41.95	-12.88
23	-111.22	-79.02
24	-98.38	-73.00
25	-59.14	-21.96
26	-75.66	-42.32
27	-50.58	-20.91
28	-99.10	-63.10
29	-55.72	-22.86
30	-46.02	-20.67

Tableau S6: Restrain parameters with force constants in kcal/Å² for distances and kcal/degree² for angles.

restrain-distance	Fe	N_{his}	250.0	1.9	2.0	
restrain-distance	Fe	S_{met}	250.0	1.9	2.0	
restrain-distance	Fe	N_{por1}	250.0	1.9	2.0	
restrain-distance	Fe	N_{por2}	250.0	1.9	2.0	
restrain-distance	Fe	N_{por3}	250.0	1.9	2.0	
restrain-distance	Fe	N_{por4}	250.0	1.9	2.0	
restrain-angle	N_{his}	Fe	N_{por1}	10.0	85.0	95.0
restrain-angle	N_{his}	Fe	N_{por2}	10.0	85.0	95.0
restrain-angle	N_{his}	Fe	N_{por3}	10.0	85.0	95.0
restrain-angle	N_{his}	Fe	N_{por4}	10.0	85.0	95.0
restrain-angle	S_{met}	Fe	N_{por1}	10.0	85.0	95.0
restrain-angle	S_{met}	Fe	N_{por2}	10.0	85.0	95.0
restrain-angle	S_{met}	Fe	N_{por3}	10.0	85.0	95.0
restrain-angle	S_{met}	Fe	N_{por4}	10.0	85.0	95.0

Tableau S7: Bond lengths (in Å) and valence angles (in degree) for the reduced heme structure (similar values are obtained for the ferric states for these parameters). MD simulations with the Standard deviations (in brackets).

Parameters	351C[a]	DFT[b]	PM7	AMOEBA
NE1-CE2	1.36	1.37	1.40 (0.02)	1.37 (0.03)
CE2-CD2	1.39	1.45	1.45 (0.02)	1.40 (0.03)
CD2-CD2	1.38	1.38	0.37 (0.02)	1.40 (0.03)
CE2-C2	1.38	1.40	1.38 (0.02)	1.40 (0.03)
NE2-CD2	1.36	1.38	1.40 (0.03)	1.34 (0.03)
CD2-CG	1.38	1.38	1.40 (0.03)	1.35 (0.02)
ND1-CG	1.36	1.38	1.40 (0.04)	1.37 (0.02)
ND1-CE1	1.36	1.36	1.38 (0.04)	1.38 (0.03)
CE1-NE2	1.36	1.33	1.37 (0.04)	1.38 (0.03)
SD-CE	1.80	1.80	1.81 (0.03)	1.81 (0.04)
CE2-C2-CE2	125	125	123 (2.6)	126 (0.3)
NE1-FE-NE1(1)	90	90	90 (2.0)	90 (0.3)
NE1-FE-NE1(2)	178	179	176 (2.0)	177 (1.8)
CD2-NE2-CE1	108	107	107 (2.3)	107 (2.5)
CG-CD2-NE2	107	105	108 (2.7)	111 (2.6)
ND1-CE1-NE2	109	109	109 (2.0)	109 (2.4)
CE-S-CG	105	104	102 (3.9)	100 (2.6)

a: geometrical data extracted from the X-ray structure contained in PDB file 351C. b: geometrical data obtained on a gas phase optimized geometry of the [PFe^{II}-ImMe-EMS] complex. c: Data obtained from classical MD simulation with the standard CHARMM force field.

Tableau S8: Bond lengths (in Å) and valence angles (in degree) for the reduced heme structure (similar values are obtained for the ferric states for these parameters). MD simulations with the Standard deviations (in brackets).

Parameters	1B5A[a]	DFT[b]	CHARMM[c]	AMOEBA-FIT
NE1-CE2	1.37	1.37	1.35 (0.02)	1.37 (0.02)
CE2-CD2	1.45	1.45	1.46 (0.03)	1.40 (0.03)
CD2-CD2	1.35	1.38	1.35 (0.03)	1.40 (0.03)
CE2-C2	1.36	1.4	1.35 (0.03)	1.40 (0.03)
NE2-CD2	1.33	1.38	1.38 (0.03)	1.38 (0.02)
CD2-CG	1.36	1.38	1.36 (0.03)	1.38 (0.03)
ND1-CG	1.38	1.38	1.37 (0.02)	1.37 (0.02)
ND1-CE1	1.33	1.35	1.37 (0.02)	1.35 (0.02)
CE1-NE2	1.36	1.33	1.32 (0.02)	1.34 (0.02)
CE2-C2-CE2	126	125	126 (2.3)	126 (2.8)
NE1-FE-NE1(1)	90	90	90 (1.8)	90 (2.8)
NE1-FE-NE1(2)	179	179	176 (2.2)	177 (1.7)
CD2-NE2-CE1	109	107	105 (1.8)	104 (2.3)
CG-CD2-NE2	108	109	112 (1.9)	112 (2.4)
ND1-CE1-NE2	108	109	106 (1.8)	111 (2.4)

a: geometrical data extracted from the X-ray structure contained in PDB file 1B5A. b: geometrical data obtained on a gas phase optimized geometry of the [PFe^{II}-ImMe2] complex. c: Data obtained from classical MD simulation with the standard CHARMM force field.

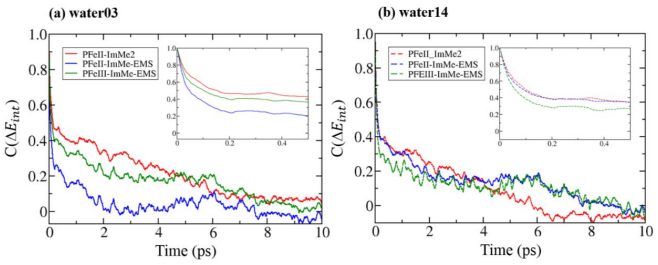


Figure S1: Normalized autocorrelation functions of the AMOEBA interaction energy of heme in the ferrous and ferric states with water aggregates (using the water03 (a) and water14 (b) water models). See section II.3 for details.

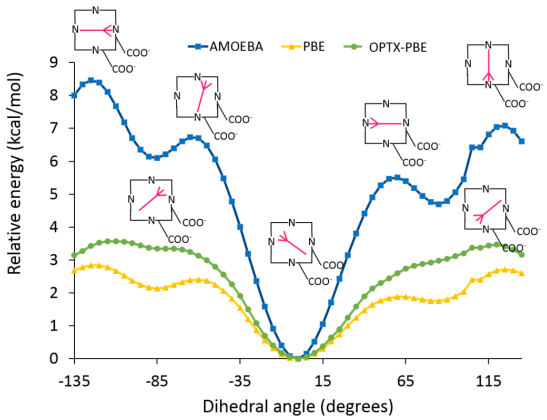


Figure S2: Energy profile associated to the rotation around the Nhis-Fe bond. The angle is one of the C_{his}-N_{his}-Fe-N_{heme} dihedral. The geometries corresponding to each extremum is schematized by a square holding four nitrogen atoms to symbolize the porphyrin ligand and by a pink three-branch fork representing the methyl-imidazole ligand. DFT calculations with PBE and OPTX-PBE functionals have been computed with deMon2k.

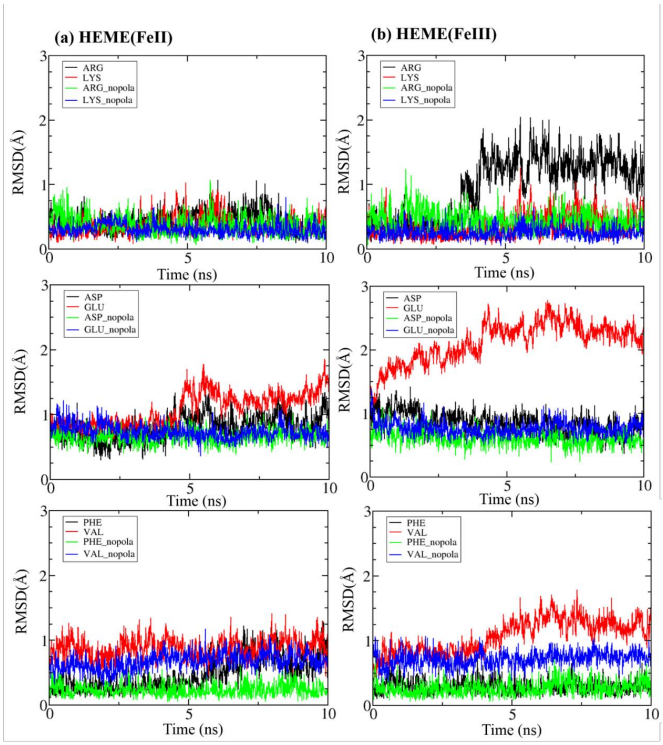


Figure S3: RMSD of aminoacid residues located within 10Å of the heme group using the AMOEBA or CHARMM (no_pol) force field.

MD simulation of heme cofactors embedded in water droplets with the PM7 methods

In a first step classical MD simulations were carried out with the CHARMM force field and the CHARMM c35b3 version. For each of the [PFe^{II}-ImMe-EMS] and [PFe^{II}-ImMe$_2$] complexes the heme was solvated in a box of TIP3P molecules the edge of which was set to 45 Å. After geometry optimization the systems was equilibrated by MD simulations at 300 K during 2 ns using periodic boundary condition and the Particle Mesh Ewald summation. The time step was set to 1 fs and the temperature was controlled with a Langevin algorithm. After equilibration we extracted aggregates comprised of the heme complexes and all the water molecules situated within a 12 Å radius sphere centered on the iron cation. The water droplets contained 231 and 205 water molecules for the [PFe^{II}-ImMe-EMS] and [PFe^{II}-ImMe$_2$] complexes respectively. The aggregates were then used in PM7 simulations.

We used the CUBY4 framework[1] which is interfaced to MOPAC2016[2] to carry out geometry optimization and MD simulations. We first carried 200 steps of geometry optimizations at the PM7 semi-empirical method[3]. Tests with PM6[4] were not conclusive since the iron coordination in the optimized geometry was not meaningful. Also simulations for the ferric state frequently failed to converge. We therefore restricted our simulation to the ferrous state. MD simulations were launched for 80 ps with a time step of 2 fs. Temperature was controlled with a Nose-Hoover thermostat (coupling frequency of 0.1 ps^{-1}). The shape of the water droplet was controlled with the confinement algorithm developed by Belgov and Roux et al. as implemented in CUBY4[5].

AMOEBA Force Field parameters

The parameters created for heme cofactor in its ferrous and ferric forms will be available on the TINKER-HP website soon.

REFERENCES

1. Řezáč, J., Cuby: An integrative framework for computational chemistry. *J. Comput. Chem.* **2016,** *37* (13), 1230–1237
2. Stewart, J. *MOPAC2016. Stewart Computational Chemistry,*, Colorado Springs, CO: 2016.
3. Stewart, J. J. P., Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013,** *19* (1), 1-32.
4. Stewart, J. J. P., Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009,** *15* (7), 765-805.
5. Beglov, D.; Roux, B., Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994,** *100* (12), 9050-9063.

CONCLUSION

In this chapter, we reported efforts toward the use of sophisticated force fields going beyond the point charge description of electrostatics. The objective underlying these efforts is to reach significantly more accurate evaluation of redox potentials of biological redox cofactors nested within proteins. AMOEBA FF was chosen as a workhorse. We have developed sets of electrostatic multipoles for heme in both ferrous and ferric states. Extensive validations have been carried out against DFT and MP2 calculations. In the case of interaction between heme and small molecules, AMOEBA parameters give better results especially for highly polarizable groups. The interaction energies calculated with AMOEBA have an excellent correlation with DFT results outperforming non-polarizable models. The water model was found to be important, the 2014-water model giving better results than the 2003-water model. Thanks to recent and decisive algorithmic developments in Tinker-HP software, we carried out MD simulations of heme proteins on the nanosecond time scale with the newly developed parameters.

The perspective of the development reported in this Chapter are wide. The next step would be to use the AMOEBA parameters to evaluate redox potentials in the series of heme proteins introduced in Chapter 1. The heme parameters for AMOEBA will also be useful for other research groups interested in hemoproteins. According the specific aim, these parameters may also be adapted for other chemical variants of the heme cofactors.

REFERENCES

1. Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D., Jr., Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013,** *9* (12), 5430.
2. Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C., Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A.* **2003,** *107* (48), 10353.
3. (a) Ren, P. Y.; Ponder, J. W., Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002,** *23* (16), 1497; (b) Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B.* **2003,** *107* (24), 5933.
4. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P., Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.* **2007,** *3* (6), 1960.
5. Hermida-Ramon, J. M.; Brdarski, S.; Karlstrom, G.; Berg, U., Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003,** *24* (2), 161.
6. Ponder, J. W.; Wu, C. J.; Ren, P. Y.; 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., Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B.* **2010,** *114* (8), 2549.

7. Ponder, J. W., TINKER: Software Tools for Molecular Design, 3.9. *Washington University School of Medicine: Saint Louis, MO* **2001**.

8. P. Eastman, J. S., J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comp. Biol.* **2017**, 13(7): e1005659.

9. Lagardere, L.; Jolly, L. H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y. Y.; Ponder, J. W.; Piquemal, J. P., Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018,** *9* (4), 956.

10. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B.* **2015,** *119* (29), 9423.

11. (a) Grossfield, A.; Ren, P.; Ponder, J. W., Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003,** *125* (50), 15671; (b) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P., Simulation of Ca2+ and Mg2+ Solvation Using Polarizable Atomic Multipole Potential. *J. Phys. Chem. B.* **2006,** *110* (37), 18553; (c) Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A., Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J. Chem. Phys.* **2006,** *125* (5), 054511; (d) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P., Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010,** *6* (7), 2059; (e) Semrouni, D.; Isley, W. C.; Clavaguéra, C.; Dognon, J.-P.; Cramer, C. J.; Gagliardi, L., Ab Initio Extension of the AMOEBA Polarizable Force Field to Fe2+. *J. Chem. Theory Comput.* **2013,** *9* (7), 3062.

12. (a) Semrouni, D.; Cramer, C. J.; Gagliardi, L., AMOEBA force field parameterization of the azabenzenes. *Theor. Chem. Acc.* **2014,** *134* (1); (b) Ren, P.; Wu, C.; Ponder, J. W., Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011,** *7* (10), 3143; (c) Yue, S.; Chuanjie, W.; W., P. J.; Pengyu, R., Multipole electrostatics in hydration free energy calculations. *J. Comput. Chem.* **2011,** *32* (5), 967.

13. (a) Jiang, J.; Wu, Y.; Wang, Z.-X.; Wu, C., Assessing the Performance of Popular Quantum Mechanics and Molecular Mechanics Methods and Revealing the Sequence-Dependent Energetic Features Using 100 Tetrapeptide Models. *J. Chem. Theory Comput.* **2010,** *6* (4), 1199; (b) Pengyu, R.; W., P. J., Consistent treatment of inter and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002,** *23* (16), 1497.

14. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013,** *9* (9), 4046.

15. Huiying, C.; Liaoran, C.; Xiangda, P.; Guohui, L., Polarizable force field development for lipids and their efficient applications in membrane proteins. *WIRES. Comput. Mol. Sci.* **2017,** *7* (5), e1312.

16. Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P., AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018,** *14* (4), 2084.

17. (a) Marjolin, A.; Gourlaouen, C.; Clavaguéra, C.; Ren, P. Y.; Wu, J. C.; Gresh, N.; Dognon, J.-P.; Piquemal, J.-P., Toward accurate solvation dynamics of lanthanides and actinides in water using polarizable force fields: from gas-phase energetics to hydration free energies. *Theor. Chem. Acc.* **2012,** *131* (4), 1198; (b) Clavaguéra, C.; Pollet, R.; Soudan, J. M.; Brenner, V.; Dognon, J. P., Molecular Dynamics Study of the Hydration of Lanthanum(III) and Europium(III) Including Many-Body Effects. *J. Phys. Chem. B.* **2005,** *109* (16), 7614; (c) Jing, Z.; Qi, R.; Liu, C.; Ren, P., Study of interactions between metal ions and protein model compounds by energy decomposition analyses and the AMOEBA force field. *J. Chem. Phys.* **2017,** *147* (16), 161733; (d) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P., Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012,** *8* (4), 1314.

18. (a) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L. Q.; Yang, W.; Ren, P. Y., The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force

Field. *J. Chem. Theory Comput.* **2012,** *8* (5), 1721; (b) Zhang, C.; Lu, C.; Wang, Q.; Ponder, J. W.; Ren, P., Polarizable Multipole-Based Force Field for Dimethyl and Trimethyl Phosphate. *J. Chem. Theory Comput.* **2015,** *11* (11), 5326; (c) Zhang, C.; Bell, D.; Harger, M.; Ren, P., Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J. Chem. Theory Comput.* **2017,** *13* (2), 666.

19. (a) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P., Calculation of protein-ligand binding free energy by using a polarizable potential. *P. Natl. Acad. Sci. USA* **2008,** *105* (17), 6290; (b) Shi, Y.; Zhu, C. Z.; Martin, S. F.; Ren, P., Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. *J. Phys. Chem. B.* **2012,** *116* (5), 1716.

20. Wu, X.; Clavaguera, C.; Lagardère, L.; Piquemal, J.-P.; de la Lande, A., AMOEBA Polarizable Force Field Parameters of the Heme Cofactor in Its Ferrous and Ferric Forms. *J. Chem. Theory Comput.* **2018,** *14* (5), 2705.

21. Allinger, N. L.; Yuh, Y. H.; Lii, J. H., Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989,** *111* (23), 8551.

22. Wilson, E. B., Decius, J. C., & Cross, P. C. (1955). *Molecular vibrations: The theory of infrared and Raman vibrational spectra*. New York: McGraw-Hill. **1955**.

23. Halgren, T. A., The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992,** *114* (20), 7827.

24. Stewart, R. F.; Davidson, E. R.; Simpson, W. T., Coherent X‐Ray Scattering for the Hydrogen Atom in the Hydrogen Molecule. *J. Chem. Phys.* **1965,** *42* (9), 3175.

25. Thole, B. T., Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981,** *59* (3), 341.

26. Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A., Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J. Chem. Phys.* **2001,** *115* (5), 2237.

27. Wang, B.; Truhlar, D. G., Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theory Comput.* **2010,** *6* (11), 3330.

28. (a) Rockers, J. A.; Wang, Q. T.; Liu, C. W.; Piquemal, J. P.; Ren, P. Y.; Ponder, J. W., An optimized charge penetration model for use with the AMOEBA force field. Phys. Chem. Chem. Phys. **2017,** *19* (1), 276; (b) Christophe, N.; Louis, L.; Étienne, P.; Nohad, G.; Qiantao, W.; R., B. D.; A., R. J.; W., P. J.; Y., R. P.; Jean-Philip, P., Scalable improvement of SPME multipolar electrostatics in anisotropic polarizable molecular mechanics using a general short-range penetration correction up to quadrupoles. *J. Comput. Chem.* **2016,** *37* (5), 494.

29. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theor. Comput.* **2013,** *9* (9), 4046.

30. Semrouni, D.; Ohanessian, G.; Clavaguera, C., Structural, energetic and dynamical properties of sodiated oligoglycines: relevance of a polarizable force field. *Phys. Chem. Chem. Phys.* **2010,** *12* (14), 3450.

31. Stone, A. J., Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981,** *83* (2), 233.

32. Stone, A. J., Mistributed Multipole Analysis for Gaussian Wavefonctions. *version 2.2.03* **2005-2007**.

33. Stone, A. J., Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128.

34. G., Á. J.; Christophe, C., A comprehensive approach to molecular charge density models: From distributed multipoles to fitted atomic charges. *Int. J. Quantum Chem.* **1994,** *52* (1), 17.

35. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B.* **2015,** *119* (29), 9423.

PART II

Real-Time Propagation of the Electronic Density in polarizable environment

Real-Time Propagation of the Electronic Density in polarizable environment

Electron transport (ET) is a fundamental process taking place in biological systems. For example light harvesting, the cellular respiratory chains, enzymatic reactions or defenses against oxidative stress all rely on transport of electrons over nanometers[1]. The temporal scales associated with these phenomena cover several orders of magnitude ranging from a few microseconds (e.g. long-range ET in cytochromes in the respiratory chain) down to tens of atto-seconds (e.g. charge migration within molecules). Understanding the molecular mechanism governing these fascinating processes is extremely motivating. The conceptual frameworks commonly used to rationalize electron transfers between molecules can be classified into two categories.

The first class covers ET limited by nuclear motion and can be understood using the Marcus theory. The theory considers two electronic diabatic states corresponding to the transferred electron either of the electron donor or on the electron acceptor. In order to preserve energy, ET can only take place for nuclear configurations corresponding to degeneracy of the two diabatic states. Therefore nuclear motion gate the electron tunneling for biological ET, because proteins exhibit multiscale dynamics (femto- to microseconds), so it can be the overall rate of electron transfer. In the regions of the conformational space associated with degeneracy of the diabatic states, tunneling of electron from the donor to the acceptor over the intervening medium takes place on the atto- to femtosecond time scale depending on the strength of the quantum mechanical coupling between the diabats. A great deal of effort has been deployed over the last decade to unravel the microscopic mechanisms governing protein-mediated tunneling. Semi-empirical models such as the pathway model were developed in the mid-1980s by Hopfied,

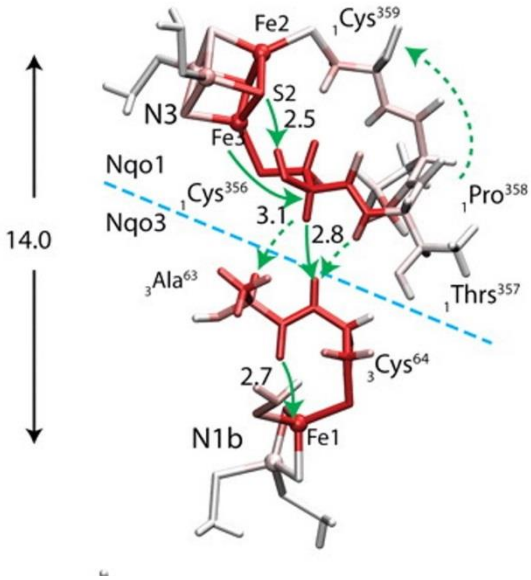


Figure 1. Tunneling currents involved in one of the inter iron-sulfur clusters electron transfer in the Respiratory complex 1. Interatomic distances are given in Å. The full and dashed arrows represent primary and secondary interatomic fluxes. The stronger the density flux through an atom, the darker its color. (Figure 2 from ref 3)

Beratan and Onuchic to map the ET pathway[2]. Some researchers sought the rationale of tunneling pathways based on the calculation by modern electronic structure theory programs, of tunneling density currents. In particular Stuchebruhkov thoroughly investigated how the electron flows from the donor to the acceptor[3] highlighting the protein residues that are the most important for efficient tunneling (Figure 1).

Another kind of electron transfer is the class of those governed by electron correlations as described by Cederbaum. They are called charge migration and proceed without any rearrangement of the nuclei. They also take place on attosecond time scales.

The theoretical model to simulate real-time time-dependent electronic structure response to external perturbations is the time-dependent Schrödinger equation. As for stationary calculations, approximations are required to simulate the time-dependent Schrödinger equation. Two families of methods can be distinguished. One on hand the wave function approaches such as the TD-HF (Time-dependent-Hartree–Fock), TD-CI (Time-dependent-configuration interaction), or the TD-MCSCF (Time-dependent multi configurational self-consistent field); on the other hand the Real-Time time-dependent density functional theory (RT-TDDFT). The latter has been gaining strong momentum in the last years due to the advantage ratio of accuracy toefficiency.

During the first years of the development, much effort was placed on devising efficient and robust algorithms. With these developments, we can now apply RT-TDDFT to molecular systems comprised of hundreds of atoms[4]. The current challenges are two-fold. On one hand, it is needed to continue developing highly efficient algorithms to simulate larger system and on the other hand, to account for environment effects, especially for heterogeneous environments. The objective of the present Part II is to contribute to the development of an original and highly effective RT-TDDFT methodology to deal with ED in large and heterogeneous environment.

In Chapter 4, we introduce the RT-TDDFT methodology. In Chapter 5, we report an original coupling between RT-TDDFT and a polarizable force field to take into account the heterogeneous environment.

1. (a) Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L., Nature of biological electron transfer. *Nature.* **1992,** *355*, 796; (b) Page, C. C.; Moser, C. C.; Chen, X.; Dutton, P. L., Natural engineering principles of electron tunnelling in biological oxidation–reduction. *Nature.* **1999,** *402*, 47; (c) Gray, H. B.; Winkler, J. R., Electron tunneling through proteins. *Q. Rev. Biophysics.* **2003,** *36* (3), 341.
2. (a) Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J., Electron tunneling through covalent and noncovalent pathways in proteins. *J. Chem. Phys.* **1987,** *86* (8), 4488; (b) Beratan, D. N.; Onuchic, J. N.,

Electron tunneling pathways in proteins: influences on the transfer rate. *Photosynth. Res.* **1989,** *22* (3), 173; (c) Beratan, D. N.; Onuchic, J. N.; Betts, J. N.; Bowler, B. E.; Gray, H. B., Electron tunneling pathways in ruthenated proteins. *J. Am. Chem. Soc.* **1990,** *112* (22), 7915.

3. Hayashi, T.; Stuchebrukhov, A. A., Electron tunneling in respiratory complex I. *Proc. Natl. Acad. Sci.* **2010,** *107* (45), 19157.

4. Provorse. R. M.; Isborn. M. C., Electron dynamics with real-time time-dependent density functional theory. *Int. J. Quantum Chem.* **2016,** *116* (10), 739-749.

CHAPTER 3

Efficient and Robust Implementation of RT-TDDFT in deMon2k

Efficient and Robust Implementation of RT-TDDFT in deMon2k

The theory of Time-Dependent Density Functional Theory (TDDFT) was developed in the 1980's by Runge and Gross[1]. They proved that the evolution of the electron density of a molecular system is uniquely defined by the time dependent external potential acting on the electrons (within an arbitrary constant). This theory opens the door toward the understanding of the response of a molecular system subjected to external perturbations like the electric field component of an electromagnetic wave. The TDDFT equations can be solved by application of the Linear Response theory giving rise to the so-called Linear-Response TDDFT (LR-TDDFT)[2] method. This formalism is adapted for weak perturbations. LR-TDDFT has found many applications in theoretical chemistry to calculate electronic excitation energies and oscillator strengths, making possible the simulation of absorption spectra of molecules[3]. LR-TDDFT has also been used in combination with molecular dynamics to simulate non-adiabatic chemical reactivity[4]. The acronym RT-(Real Time)-TDDFT on the other hand, refers to methodologies that attempt to propagate the equations-of-motion for the electron density (ρ) explicitly by numerical integration. RT-TDDFT simulations date back to the works of Theilhaber[2b], Calvayrac, Reinhard, and Suraud[5], and of Bertsch and Yabana[6]. The explicit propagation of the Runge and Gross equations has multiple advantages. First, the perturbations applied in RT-TDDFT are not necessarily weak as in LR-TDDFT which opens the possibility to simulate non-linear optical phenomena, for example the generation of High Harmonics[7] (HHG) or the ionization of molecules under strong electric fields[8]. Second, RT-TDDFT gives access to the realm of attosecond electron dynamics, the understanding of which is essential in the description of various physical-chemical processes such as the response of a molecule to irradiation by photons or massive particles (protons, electrons …). Finally the algorithms of simulating the evolution of the density look formally straightforward.

While RT-TDDFT has emerged in the physics community already in the 1990's, it has gained strong momentum in theoretical chemistry more recently. Implementations of RT-TDDFT for molecular systems have been reported in Octopus[9], GAUSSIAN[10], SIESTA[11], NWCHEM[12], QBOX[13], CPMD[14], and Q-chem[15] . The range of applications includes the simulation of spectroscopic observables (UV–visible spectra[16], core-level near-edge X-ray absorption spectra[17], photoelectron emission spectra[18]), the investigation of coherence and ultrafast charge-transfer dynamics in small molecules[19], the study of molecular conductance in polymers[20], or the simulation of non-adiabatic electron-nuclear dynamics[14].

A current challenge for RT-TDDFT approaches is to achieve efficient and robust implementations in order to address large molecular systems comprised of hundreds of atoms. The simulation of electron dynamics (ED) by RT-TDDFT requires a huge number of Kohn-Sham potential calculations, which can limit the accessible simulation lengths or the size of the simulated systems. In the context of *stationary*

DFT, algorithmic developments such as the Resolution of the Identity[21] (RI) or variational Density Fittings[22] (DF) have been decisive to reach high computational efficiency. RT-TDDFT methods could potentially benefit from such algorithms. This is the objective of the work described in this Chapter, namely to carry out a new implementation of RT-TDDFT within the deMon2k[23] software that provides this kind of advanced algorithms. In section I, we will detail the formalism employed in RT-TDDFT equations. We will cover the mathematical expressions for carrying out ED simulations, details on the auxiliary DFT framework used in deMon2k, the types of available perturbations to trigger electron dynamics and finally the tools implemented to analyze the time dependent electron density. In Section II, we will report various ED simulations results aimed at testing the computational performance of the implementation. We will occasionally refer to the article published in the Journal of Chemical Theory and Computation describing our implementation. This manuscript is reproduced at the end of Part II.

I. REAL-TIME AUXILIARY TDDFT

I.1 Electronic Equations-of-Motion

I.1.a Linear Combination of Atomic Orbitals formalism

Runge and Gross developed the many body wave function TD Schrödinger equation into the single-particle TD density functional theory[24]. They showed that there is a one-to-one correspondence between the time-dependent density $\rho(\boldsymbol{r},t)$ and the time-dependent potential $v_{ext}(\mathbf{r},\mathrm{t})$ for a given set of initial conditions. The most attractive way to calculate approximate densities is provided by the TD Kohn-Sham (TDKS)[25] framework that refers to a fictitious system of non-interacting electrons having the same ground-state density as the real system of interacting electrons. Each TDKS single particle wave function ψ_i obeys a time-dependent Schrödinger equation that reads, in atomic units:

$$i\frac{\partial\psi_i(\boldsymbol{r},t)}{\partial t} = \hat{H}[\rho(\mathbf{r},t)]\psi_i(\boldsymbol{r},t) \tag{1}$$

The non-relativistic TDKS operator $\hat{H}$ is a functional of the density which is comprised of four terms.

$$\hat{H}(\boldsymbol{r},t) = v_{ext}(\boldsymbol{r},t) - \frac{1}{2}\nabla^2 + \int\frac{\rho(\boldsymbol{r}',t)}{|\boldsymbol{r}-\boldsymbol{r}'|}dr' + v_{XC}[\rho(\boldsymbol{r},t)] \tag{2}$$

The first term on the r.h.s is the external potential v_{ext} felt by the electrons. It includes the potential created by the nuclei. It can be supplemented by other terms describing, for example the interaction with an external electric field or a charged projectile. In the context of hybrid QM/MM simulations the external potential is supplemented by the potential created by charged or polarized MM atoms (cf. Chapter 4). The last three terms are the kinetic energy operator, the classical Coulomb potential and

the exchange correlation (XC) potential. v_{XC} is a functional of the density that is in principle non local in space and time. In many implementations though the dependence in time of v_{XC} is neglected. This is called the adiabatic approximation[26]. Under this hypothesis, the approximate XC functionals developed for stationary DFT calculations can be re-used in RT-TDDFT without special reprogramming. Although the adiabatic approximation is valid for many applications[27], it introduces a supplementary approximation in the propagation besides the unknown nature of the exact XC functional. Attempts are currently underway to go beyond the adiabatic approximation[28], but we will not consider them in this Chapter.

The density ρ of the non-interacting reference system is obtained from the manifold of KS orbitals, the electronic occupations of which are denoted n_i.

$$\rho(\boldsymbol{r},t) = \sum_{i}^{all} n_i \psi_i^*(\boldsymbol{r},t)\psi_i(\boldsymbol{r},t) \tag{3}$$

The energy of the molecular systems $E[\rho(\boldsymbol{r},t)]$ reads,

$$E[\rho(\boldsymbol{r},t)] = \int v_{ext}(\mathbf{r},\mathrm{t})\rho(\boldsymbol{r},t)dr + \sum_{i}^{all}\int n_i\psi_i^*(\boldsymbol{r},t)\left(-\frac{1}{2}\nabla^2\right)\psi_i(\boldsymbol{r},t)\,dr + \frac{1}{2}\int\int\frac{\rho(\mathbf{r},t)\rho(\boldsymbol{r}',t)}{|\boldsymbol{r}-\boldsymbol{r}'|}d\boldsymbol{r}d\boldsymbol{r}' + E_{xc}[\rho(\boldsymbol{r},t)] \tag{4}$$

In practice, in order to code the above equation it is necessary to choose a methodology to represent the KS molecular orbitals (MOs). A common strategy is to define a basis set and to expand the KS MO as linear combinations of the basis set elements[29]. Localized atomic orbitals (e.g. Gaussian functions) or plane waves have been proposed in RT-TDDFT. The former choice is well suited for confined molecular systems ("cluster approach"). This is the choice made in NWChem[30], Gaussian[31], Qchem[32] or deMon2k[23]. The latter is more adapted for periodic systems (e.g. CPMD). An alternative which is more frequently encountered in codes developed by physicists is to represent KS MOs on discretized grids in real space (see for instance the Octopus[33] or TELEMAN[34] codes).

In deMon2k, the KS molecular orbitals are expanded as Linear Combinations of Gaussian-Type Orbitals (LCGTO). We use Greek letters μ , ν to denote AO. For simplicity the same notations are used to index them. The KS MOs read:

$$\psi_i(\boldsymbol{r},t) = \sum_{\mu=1}^{N_{AO}} c_{\mu i}(t)\mu(\boldsymbol{r}) \tag{5}$$

where N_{AO} is the number of atomic orbitals (AO). The MO coefficients $c_{\mu i}$ hold the dependence in time. It has to be noticed that because of the occurrence of the imaginary unit in the time-dependent Schrödinger equations, the MO coefficients are complex numbers in RT-TDDFT simulations ($c_{\mu i} = c_{\mu i}^{R} + ic_{\mu i}^{I}$, $c_{\mu i}^{*}$ denoting the complex conjugate of $c_{\mu i}$). The electron density can be written from the density matrix P, the elements of which are given by Eq. 7.

$$\rho(t, \boldsymbol{r}) = \sum_{\mu,\nu}^{N_{AO}} P_{\mu\nu}(t)\mu(\boldsymbol{r})\nu(\mathbf{r}) \tag{6}$$

$$P_{\mu\nu}(t) = 2 \sum_{i}^{N_{MO}/2} c_{\mu i}^{*}(t)c_{\nu i}(t) \tag{7a}$$

$$P_{\mu\nu}(t) = 2 \sum_{i}^{N_{MO}/2} \left(c_{\mu i}^{\mathrm{R}}(t)c_{\nu i}^{\mathrm{R}}(t) + c_{\mu i}^{\mathrm{I}}(t)c_{\nu i}^{\mathrm{I}}(t) + i\left(c_{\mu i}^{\mathrm{I}}(t)c_{\nu i}^{\mathrm{R}}(t) - c_{\mu i}^{\mathrm{R}}(t)c_{\nu i}^{\mathrm{I}}(t)\right)\right) \tag{7b}$$

$$P_{\mu\nu}(t) = 2 \sum_{i}^{N_{MO}/2} \left(P_{\mu\nu}^{\mathrm{R}}(t) + iP_{\mu\nu}^{\mathrm{I}}(t)\right) \tag{7c}$$

where N_{MO} is the number of occupied MO. We have defined the matrices P^{R} and P^{I} that collect the real and imaginary elements of the full density matrix. One can show that $P^{R}$is symmetric while P^{I} is anti-symmetric. Having introduced the LC-GTO framework we can rewrite Eq. 4 with the help of matrix and bracket notations.

$$E[\rho(\boldsymbol{r}, t)] = \sum_{\mu,\nu} P_{\mu\nu}H_{\mu\nu}^{ext} + \sum_{\mu,\nu} P_{\mu\nu}\left\langle \mu\left|-\frac{1}{2}\nabla^2\right|\nu\right\rangle + \sum_{\mu,\nu}\sum_{\sigma,\tau} P_{\mu\nu}(t)P_{\sigma\tau}(t)\langle\mu\nu\|\sigma\tau\rangle + E_{xc}[\rho(\boldsymbol{r}, t)] \tag{8}$$

$H_{\mu\nu}^{ext}$ and $H_{\mu\nu}^{kin} = \left\langle \mu\left|-\frac{1}{2}\nabla^2\right|\nu\right\rangle$ are matrix elements of the external and kinetic operators. Together they form the core Hamiltonian. The symbol ‖ stands for the coulomb operator ($1/|\boldsymbol{r} - \boldsymbol{r}'|$). The Kohn-Sham potential is defined by differentiating the energy with respect to the electron density. In matrix notations the matrix elements of the KS potential are readily obtained by differentiation with respect to the elements of the density matrix.

$$H_{\mu\nu} \equiv \frac{\partial E}{\partial P_{\mu\nu}} = H_{\mu\nu}^{ext} + \left\langle \mu\left|-\frac{1}{2}\nabla^2\right|\nu\right\rangle + \sum_{\sigma,\tau} P_{\sigma\tau}(t)\langle\mu\nu\|\sigma\tau\rangle + \frac{\partial E_{xc}[\rho(\boldsymbol{r}, t)]}{\partial P_{\mu\nu}} \tag{9}$$

I.1.b Numerical approaches to the Liouville-von Neuman equation

i) Generalities on propagators

Eq. 2 can be recast using the density matrix P into a Liouville-von Neumann type equation[35].

$$i\frac{\partial P'(t)}{\partial t} = [H'(t), P'(t)] \tag{10}$$

This expression assumes that the electronic wave functions are expressed on an orthogonal basis (as emphasized by the use of primes). Yet, in general the sets of atomic orbitals for polyatomic molecules are not orthogonal. To use Eq. 10 a transformation to an orthogonal basis set is needed. This procedure is similar to that employed in stationary DFT calculations where a transformation to an orthogonal basis is operated to facilitate the resolution of the Roothaan-Hall equations[29]. We use the Löwdin orthogonalization scheme relying on $S^{-1/2}$, S being the overlap matrix ($S_{\mu\nu} = \langle\mu|\nu\rangle$). We provide in Annex II of the Chapter the working equations to transform the density matrix or the KS matrix into the AO or MO basis.

A formal solution of Eq. 10 is given by Eq. 11, where U is the evolution operator.

$$P'(t) = U(t, t_0)P'(t_0)U^{\dagger}(t, t_0) \tag{11}$$

Eq. 11 provides the electron density at time t from the density at time t_0. It is however not useful by itself. A powerful alternative is to seek for numerical simulations of this equation. The starting point is to discretize the time t into a series of successive steps of length Δt (Eq. 12). The evolution operator for each discretized step is given by Eq. 13. $\mathcal{T}$ is the time-ordering operator, ensuring that operators associated with later times always appear to the left of those associated with earlier times.

$$U(t, t_0) = \prod_{i}^{n-1} U(t_i + \Delta t, t_i) \tag{12}$$

$$U(t_i + \Delta t, t_i) = \mathcal{T}exp\left\{-i\int_{t_i}^{t_i+\Delta t} H'(\tau)d\tau\right\} \tag{13}$$

Several different algorithms (propagators) have been proposed to construct U in discretized time[36]. Propagators differ by their respective performances and stabilities. On one hand the term ***performance*** is related to the wall-clock time required to propagate the density over a certain amount of time. This can be a time consuming task due to the calculation of integrals involved in electron-electron Coulomb interactions and in exchange correlation terms (see Eq. 9). In addition certain algorithms involve iterative procedures that require more than one calculation of the KS potential at each propagation step (details will be given later). Therefore a common strategy to many algorithms

is to minimize the number of propagation steps by taking the largest time step (Δt). It has to be kept in mind that there is a physical limit to the maximum size of Δt_{max} depending on the physical process of interest. If ω_{max} is the maximum frequency to be time resolved in the simulation, Δt_{max} should not be larger than $ca.\ 1/\omega_{max}$. In most of our applications Δt is set to 1 or 2 as. On the other hand the term ***stability*** refers to the quality of the propagated solution after a certain simulation length. The propagator should be unitary to ensure that the density matrix satisfies the following important properties: $P'^2 = P'$ (idempotent), $P'^{\dagger} = P'$ (self-adjoint) and $Tr(P') = N$, N being the number of electrons (conservation of the total number of electrons). Clearly, in many cases stability can be enhanced by decreasing the time-step but the numerical performance will be decreased. One should find a propagator that allows the best trade-off between performance and stability.

The TDKS equations fall into the category of systems of initial-value first-order ordinary differential equations (ODEs). However not all numerical methods to solve the ODEs are suitable to solve the TDKS equations. The literature contains several studies investigating the optimum propagation for RT-TDDFT such as the implicit midpoint rule (Euler, Runge-Kutta, Crank-Nicolson), the exponential midpoint rule, the time-reversal symmetry based propagator, splitting techniques or Magnus propagators, etc.[36a, 37 38] Recently, Castro and coworker tested several multi-step based propagators including exponential Runge-Kutta, and the commutator free Magnus schemes[39]. It turns out to be that the commutator free Magnus integrator is the most robust, simple and efficient propagator for the application reported by the authors. In fact, most studies conclude that the choice depends on the internal characteristics of the system investigated, on the frequency and intensity of the perturbation and on the subsequent response of the electron cloud. Consequently it is desirable to have various alternatives in a code for production runs.

We now focus more specifically on the propagators that we implemented in deMon2k. The first one is the Euler propagator. It is based on Lagrange's mean value theorem:

$$P'(t_i + \Delta t) = P'(t_i) - i[H'(t_i), P'(t_i)] * \Delta t \qquad (14)$$

The propagation of the density matrix requires only the value of the density matrix and of the Kohn-Sham matrix at the current time. This propagation scheme doesn't guarantee preservation of the norm of the KS wave-function. This can lead to divergence of the electronic propagation even with very small time steps. It is therefore not recommended to use it in production runs.

The Magnus propagator expresses the time-unordered exponential by a series of nested commutator integrals Ω_i [40]:

$$\mathcal{T}exp\left\{-i\int_{t}^{t+\Delta t}H'(\tau)d\tau\right\}=e^{W}=e^{\Omega_1+\Omega_2+\Omega_3\ldots} \tag{15}$$

$$\Omega_1(t+\Delta t,t)=-i\int_{t}^{t+\Delta t}H'(\tau)d\tau \tag{16}$$

$$\Omega_2(t+\Delta t,t)=-i\int_{t}^{t+\Delta t}d\tau_1\int_{t}^{\tau_1}d\tau_2[H'(\tau_1),H'(\tau_2)] \tag{17}$$

In practice, the series is truncated at the n^{th} order. The integrals are approximated by a quadrature formula[36a].

$$\Omega_1(t+\Delta t,t)\simeq -iH'\left(t+\frac{\Delta t}{2}\right).\Delta t \tag{18}$$

$$\Omega_2(t+\Delta t,t)\simeq -i\frac{\Delta t}{2}[H'(t_1)+H'(t_2)]-\frac{\sqrt{3}}{12}\Delta t^2[H'(t_2),H'(t_1)] \tag{19}$$

where $t_{1,2}=t+\left[(1/2)\mp\sqrt{3}/6\right].\Delta t$ are Gauss quadrature sampling points. Stopping at second order, $W\approx\Omega_1$, we will refer to it as the second-order Magnus propagator (SOMP). It is similar to the exponential midpoint rule. The SOMP requires the knowledge of only one KS potential (at time $t+\Delta t/2$). Stopping at fourth order, $W\approx\Omega_1+\Omega_2$, the fourth-order Magnus propagator (FOMP) requires the knowledge of the KS potential at $t+\Delta t/2$ (for evaluating Ω_1), at $t+\left[(1/2)-\sqrt{3}/6\right]\Delta t$ and at $t+\left[(1/2)+\sqrt{3}/6\right]\Delta t$ (for evaluating Ω_2). One immediately sees that carrying out one FOMP step requires three times the computational effort needed to perform a SOMP step. That said, FOMP generally permits longer time steps Δt for the same stability. Rubio and coll. reported comparisons of propagators for Na_8 clusters submitted to a laser[36a]. They arrived at the conclusion that the FOMP becomes more advantageous over the SOMP if Δt could be 1.5 times larger. However, one should still be cautious in using too large time steps since the error accumulates due to finite truncation of the Magnus expansion. For processes involving high frequencies, FOMP or even higher order truncated Magnus expansions become mandatory. Similar conclusions were reached by Cheng *et al*. on methane subjected to an electric field[20]. They also concluded that SOMP offered a well-balanced propagator between stability and performance. In view of the conclusions of earlier studies we decided to base our implementation on the SOMP.

ii) Iterative and Propagator-Corrector versions of SOMP

To apply the Magnus propagator to the electron density at time t, the values of the KS matrix at later time $H(t+\Delta t/2)$ should be used, which is unknown. To overcome this difficulty two methods have been implemented, namely an iterative method and a predictor-corrector (PC) method. The algorithm

of the iterative method is depicted in Figure 1. In the first step one builds a guess for $H^0(t_n + \Delta t/2)$ by extrapolation from $H(t_n)$ and $H(t_{n-1})$. Then one operates an SOMP step on $P^0(t_n)$ to obtain $P^0(t_n + \Delta t)$ using $H^0(t_n + \Delta t/2)$. The knowledge of $P^0(t_n + \Delta t)$ allows the construction of the KS potential at $t_n + \Delta t$ (third step) and a new KS potential at $t_n + \Delta t/2$ is constructed by interpolation from the potentials at t_n and $t_n + \Delta t$ (fourth step). It is denoted by $H^i(t_n + \Delta t/2)$ in Figure 1. With this improved potential at $t_n + \Delta t/2$, one loops back at step 2 and proceeds to another iteration. The process is repeated until convergence of the density or of the potential. In deMon2k we define a convergence threshold based on the variation of the auxiliary fitted density as it will be explained in I.2. This is similar to the criteria employed in the self-consistent-field approach to obtain stationary densities[41]. At convergence the potentials at times t_n and $t_n + \Delta t$ are saved in preparation for the subsequent propagation step. The electron density at $t_n + \Delta t$ is analyzed to provide insights in to the electron dynamics.

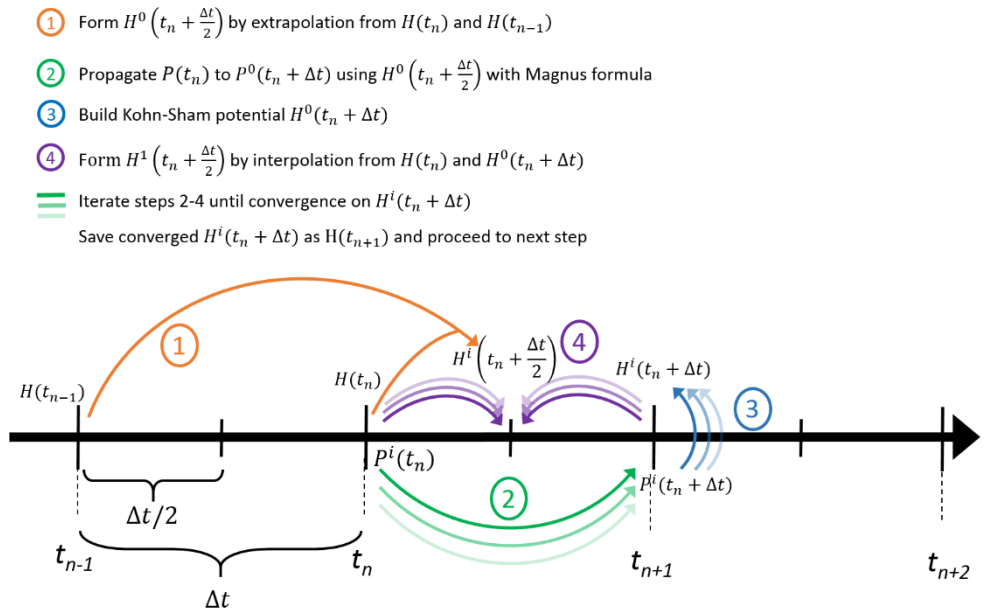


Figure 1. Iterative SOMP scheme for the propagation of the density from time t_n to time t_{n+1}.

The iterative method implies several calculations i) of the KS potential and ii) of the SOMP exponential $e^{-iH'(t+\Delta t/2)\Delta t}$ both being potentially time consuming. To alleviate the computational effort of the propagation Van Voorhis and co-workers proposed a PC method[20]. The algorithm is illustrated in Figure 2. The overall propagation $P(t_n)\ to\ P(t_n + \Delta t)$ is split into two phases: a prediction phase with a time step $\Delta t/4$, followed by a correction step with $\Delta t/2$. In the first step one builds a guess for $H(t_n + \Delta t/4)$ by extrapolation from $H(t_{n-1} + \Delta t/2)$ and $H(t_{n-2} + \Delta t/2)$. Then one operates a SOMP step on $P(t_n)$ to obtain $P(t_n + \Delta t/2)$ using $H(t_n + \Delta t/4)$. The knowledge of $P(t_n + \Delta t/2)$

allows the construction of the KS potential at $t_n + \Delta t/2$ (third step). In the corrector phase, the KS potential at $t_n + \Delta t/2$ is used to propagate $P(t_n + \Delta t/2)$ to obtain $P(t_n)$. The potentials at times $t_{n-1} + \Delta t/2$ and $t_n + \Delta t/2$ are saved in preparation for the subsequent propagation step.

In many cases the PC method combined with the SOMP is efficient and stable, but as usual cautious tests need to be carried out, for instance in case of strong perturbations. Tests with deMon2k will be reported in Section II.

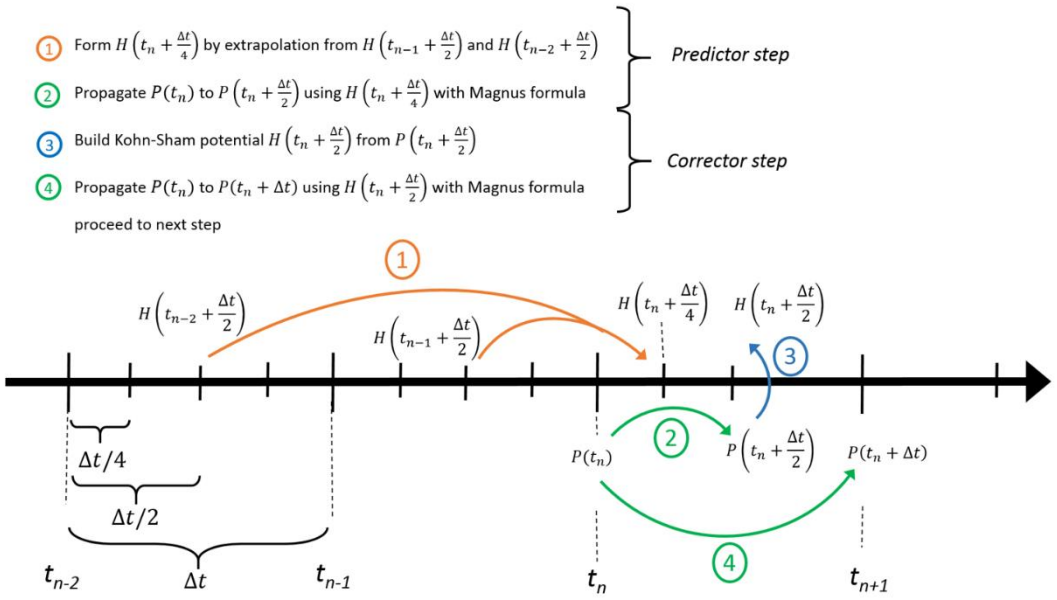


Figure 2. Predictor-Corrector SOMP scheme for the propagation of the density from time t_n to time t_{n+1}.

I.1.c The exponential of an operator

The application of the SOMP requires the evaluation of $e^{-iH'\Delta t}$. This is not a trivial task because of the complex nature of the argument of the exponential. The evaluation of the matrix exponential is another time consuming step besides the calculation of the KS potential. Several elegant solutions have been devised over the years to evaluate matrix exponentials[36b, 37, 42]. The most common methods are the diagonalization or the polynomial expansions (e.g. Taylor, Chebychev). More sophisticated methods are the Krylov subspace projection (Lanczos algorithm), the splitting techniques (Suzuki-Trotter) or the Leja and Pade interpolations. Each method has its pros and cons. The method employed to compute the exponential of the matrix has impact on the performance and on the stability of the propagation. This is still a very active research area. We focus here on the four methods we implemented in deMon2k.

i) Diagonalization

The first one proceeds *via* the diagonalization of the complex matrix W (Eq. 17). The diagonalization is performed using optimized external libraries (Lapack). This is usually a robust approach. It turns to be cumbersome for large matrices (*i.e.* large basis sets) since the cost grows as N_{AO}^3. Moreover parallelized versions of these libraries are not efficient. Therefore the diagonalization is well suited for serial calculations only.

$$e^W = Ue^w U^\dagger, \text{ with } WU = wU \tag{20}$$

ii) Taylor expansion

To overcome the scaling problem, the methods based on expansions are more advantageous. One way is to use a Taylor expansion for the exponential:

$$e^W = \sum_{n=0}^{\infty} \frac{1}{n!} W^n \tag{21}$$

in practice the number of terms introduced in the expansions is of course not infinite but limited to a finite number of terms (k). k has to be large enough to achieve convergence of the Taylor series. Depending on the systems and the physical processes investigated, the number of terms to be included in the series can be rather large. Numerical difficulties can also arise due to the fact that the method involves the multiplication of potentially very large floating point numbers by very small numbers ($1/n!$)

iii) Chebychev expansion

An alternative is to use Chebychev expansions[36a, 43]. The Chebychev polynomials are only defined in the range [-1,1]. The Hamiltonian matrix has to be scaled by a factor γ beforehand so that its spectrum lies within this range. We therefore define a scaled matrix $\widetilde{W}$ in the form of $-iH\tilde{a}$, with $\tilde{a} = \gamma\Delta t$ The Chebyshev expansion reads:

$$e^W = \mathcal{N} \sum_{n=0}^{\infty} (-i)^n c_n(\tilde{a}) T_n(\widetilde{W}) \tag{22}$$

where $\mathcal{N}$ is a normalization constant. $c_n = (2 - \delta_{n0}) J_n(\tilde{a})$ with J_n being the Bessel functions of the first kind ($J_n = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+n)!} \left(\frac{x}{2}\right)^{2m+n}$). T_n are the Chebychev polynomials of order *n*. They are defined by recurrence relationships.

$$T_0(\widetilde{W}) = I \tag{23a}$$

$$T_1(\widetilde{W}) = \widetilde{W} \tag{23b}$$

$$T_n(\widetilde{W}) = 2\widetilde{W}T_{n-1}(\widetilde{W}) - T_{n-2}(\widetilde{W}) \tag{23c}$$

The scaling factor (γ) is system dependent and has to be chosen carefully. One possibility is to scale with the upper ε_{max} and lower ε_{min} bounds of the *eigenspectrum* of H. $\gamma = (\varepsilon_{max} - \varepsilon_{min})/2$ or as suggested by Williams-Young et al.[43a] $\gamma = 3.(\varepsilon_{max} - \varepsilon_{min})/2$. This defines $\mathcal{N} = e^{-2i(\gamma+\varepsilon_{min})\Delta t}$.

iv) Baker-Campbell-Haussdorff expansion

The last method is based on the Baker-Campbell-Haussdorff (BCH) scheme[16b, 44]. It gives the exponential e^W and the propagation $P'(t + \Delta t) = e^W P'(t)\, e^{-W}$ at the same time. The KS matrix should be Hermitian to apply the BCH, which limits its applications.

$$P'(t+\Delta t) = P'(t) + \frac{1}{1}[W, P'(t)] + \frac{1}{2}\big[W, [W, P'(\mathrm{t})]\big] + \frac{1}{3!}\Big[W, \big[W, [W, P'(t)]\big]\Big] + \cdots \tag{24}$$

$$P'(t+\Delta t) = C_0 + \sum_{n=1}^{\infty} \frac{1}{n!}[W, C_{n-1}] \tag{25}$$

with $C_0 = P'(t)$. The three methods relying on expansions are in principle exact but in practice the accuracy is highly dependent on the choice of k. Two possibilities are available in deMon2k. Either one specifies the number of terms k to use throughout the ED simulation, or a tolerance threshold is given and, each time a matrix exponential has to be computed, a specific number of terms k is determined to reach a certain accuracy. In practice for each new matrix added to the series, we calculate the modules of the diagonal elements and take the sum of them. If the sum is below the threshold then the convergence of the expansion is considered to be obtained. We typically use thresholds of 1.0E-30.

I.1.d External perturbations

In general a simulation starts by a stationary SCF calculation to determine the ground state electron density of the systems of interest. Then ED simulations are carried out with the introduction of the perturbation. This is done through the addition of terms to the external potential v_{ext} term which enters Eq. 3. One possibility is to simulate the interaction of the electron cloud with the electric field of an electromagnetic wave. We make the assumption of the dipole interaction and express the interaction energy as:

$$E^{efield} = -\boldsymbol{\mu}(t) \cdot \boldsymbol{F}(t) \tag{26}$$

where $\boldsymbol{\mu}$ is the molecular dipole: $\boldsymbol{\mu}(t) = \sum_A Z_A \boldsymbol{R}_A - \int \rho(\boldsymbol{r}, t)\boldsymbol{r}d\boldsymbol{r} = \sum_A Z_A \boldsymbol{R}_A - \sum_{\mu,\nu} P_{\mu\nu} \boldsymbol{D}_{\mu\nu}$, with the elements of the dipole operator defined as $\boldsymbol{D}_{\mu\nu} = \langle \mu | \boldsymbol{r} | \nu \rangle$. The external potential to be included in the KS potential during the propagation is obtained by differentiation of the interaction energy with respect to the electron density. In matrix notation it reads:

$$\frac{\partial E^{efield}}{\partial P_{\mu\nu}} = \boldsymbol{F}(t)\boldsymbol{D}_{\mu\nu} \tag{27}$$

The general expression for the field is $\boldsymbol{F}(t) = F_{max} X(\mathrm{t}) \cos \omega t \, \boldsymbol{d}$, where F_{max} is the maximum field strength, X is the envelop of the pulse, $\boldsymbol{d}$ is the direction vector of the pulse and ω is the frequency of the light. All these parameters can be set by the user. We have implemented four different shapes of electric fields $\boldsymbol{F}$ as represented in Figure 3. These are the Gaussian pulses, squared sinusoidal pulse, Dirac (infinitely narrow pulse or "kick") and linear ramp.

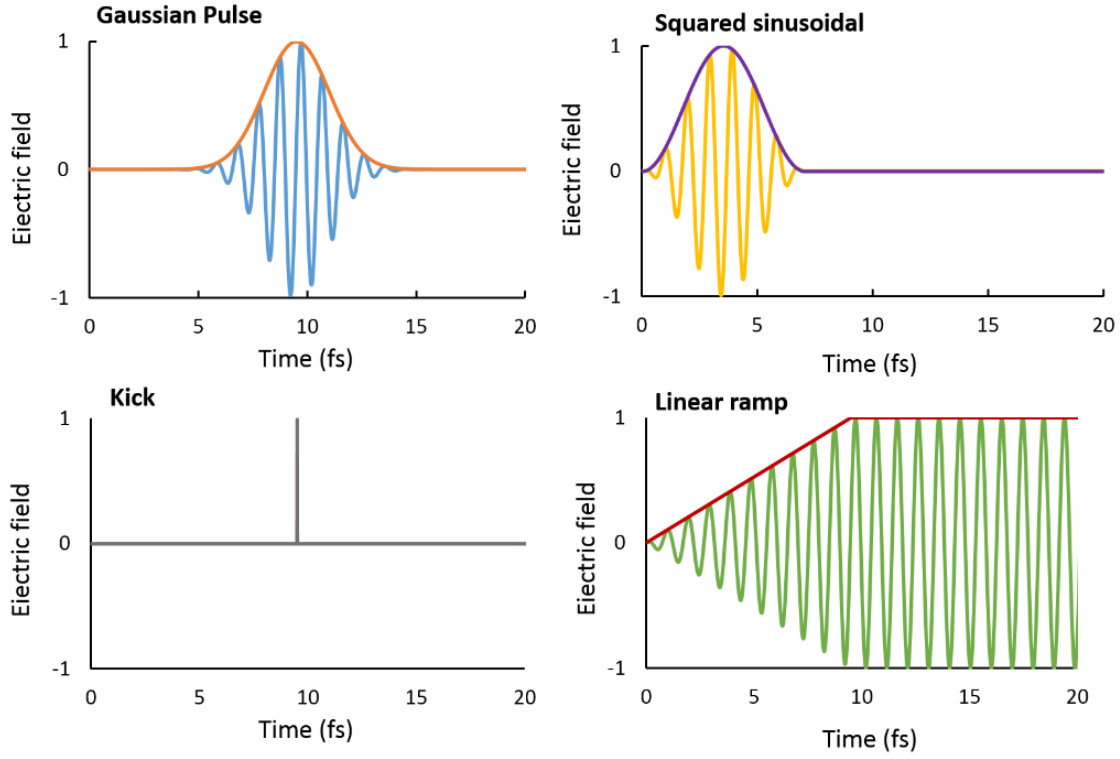


Figure 3. Examples of electric fields applicable in ED simulations with deMon2k with an arbitrary maximum field strength $\mathrm{F_{max}}$ equal to 1. Top-left: Gaussian pulse, the center is $\mathrm{t_0} = 9.5\ \mathrm{fs}$, its width $\mu_F = 1.5\ \mathrm{fs}$.The orange line corresponds to a white light (ω=0 a.u.) and the blue line a monochromatic pulse (ω=0.156 a.u). Top-right: same as Top-left but with a squared sinusoidal envelop and centered at 3.5 fs. Bottom-left infinitely narrow pulse centered at $9.5\ \mathrm{fs}$. Bottom-right: linear ramp with the maximum field achieved for $9.5\ \mathrm{fs}$. The orange line corresponds to white light (ω=0 a.u.) and the green line a monochromatic pulse (0.156 a.u).

Another type of perturbation available in deMon2k is the collision with a charged classical particle[45]. This allows the to simulation of inelastic collisions (i.e. collisions not affecting the projectile's trajectory)[46]. The interaction energy between the molecule and the projectile is given by Eq. 28.

$$E^{proj} = -\int \rho(\boldsymbol{r},t).\phi_{proj}(\boldsymbol{r},t).\boldsymbol{dr} + \sum_{A} Z_A \boldsymbol{R}_A.\phi_{proj}(\boldsymbol{r},t) \tag{28}$$

ϕ_{proj} is the potential created by the projectile with charge q_{proj}. For slow projectiles compared to the speed of light (c) the potential is calculated by a Coulomb law. For projectiles approaching the speed of light (for example alpha particles with kinetic energies above 1 MeV), a Liénard-Wiechert potential is appropriate[47]:

$$\phi_{proj}(\boldsymbol{r}) = \frac{q_{proj}}{\boldsymbol{R}\left[1 - \frac{v_{proj}^2 sin^2\theta}{c^2}\right]^{1/2}} = \frac{\gamma.q_{proj}}{\boldsymbol{R}} \tag{29}$$

where q_{proj} and v_{proj}, are the charge and velocity of the projectile, θ is the angle between the propagation direction and the distance vector between an electron and the particle ($\boldsymbol{R} = \boldsymbol{r} - \boldsymbol{r}_{proj}$). $\gamma = \left(1 - v_{proj}^2 sin^2\theta/c^2\right)^{-1/2}$ is the angle-dependent Lorentz factor. For particles travelling at speeds much lower than the speed of light ($v_{proj}^2 \ll c^2$) $\gamma \approx 1$ and Eq. 29 reduces to a standard Coulomb potential. On the other hand, when $v_{proj}^2 \to c^2$, $\gamma \to 1$ if $\theta \to 0$ (i.e. for electrons positioned on the particle trajectory) and $\gamma \to +\infty$ if $\theta \to \pi/2$ (*i.e.* for electrons positioned perpendicular to the particle trajectory). The corresponding potential term H^{proj} is added to the KS matrix.

I.1.e Complex Absorbing Potentials

An appealing application of RT-TDDFT is the simulation of ionization of molecules by application of strong electric fields or upon collisions with charged particles[18b]. Upon such perturbation a large amount of energy is deposited into the electron cloud which promotes electronic excitations. A fraction of electron density may escape the attraction potential created by the nuclei. Simulating these phenomena at the DFT level is however a challenging objective especially for computer codes relying on localized atomic basis sets like deMon2k. Indeed for describing the emission of electrons having non zero kinetic energy the basis set should span a large spatial area around the ionized molecules. The basis sets customarily used in quantum chemistry calculations haven't been designed for such applications. On the contrary they have been optimized to describe bound electrons in the vicinity of atomic nuclei.

To simulate ionization with RT-TDDFT we first use very diffuse basis sets, as developed by Schlegel and co-workers[48]. They are built on the aug-cc-pVTZ basis set with addition of several diffuse functions going up to f second quantum number. The lowest Gaussian exponents go down to 0.0064 for s functions. This is however not sufficient because the localized basis sets have physical boundaries whatever their diffuseness. We therefore add a Complex Absorbing Potential (CAP) to the Hamiltonian to absorb the emitted electrons when they are far from the molecule of interest. To better understand how CAP work in RT-TDDFT simulations let us calculate the norm of the electronic wave function:

$$\frac{d}{dt}\langle\psi|\psi\rangle = \left[\frac{d}{dt}\langle\psi|\right]|\psi\rangle + \langle\psi|\left[\frac{d}{dt}|\psi\rangle\right] \tag{30}$$

The time dependent Schrodinger equation gives $\frac{d}{dt}|\psi\rangle = \frac{1}{i}H|\psi\rangle$ and $\frac{d}{dt}\langle\psi| = \frac{-1}{i}\langle\psi|H^{\dagger}$, where $H^{\dagger}$ is the adjoint of H. The matrix elements of the Hamiltonian are $H_{ij} = \langle i|H|j\rangle$. If the Hamiltonian matrix elements are real, $H_{ij}^{\dagger} = \langle i|H^{\dagger}|j\rangle = \langle j|H|i\rangle^{*} = \langle j|H|i\rangle = H_{ij}$. Therefore Eq. 30 becomes

$$\frac{d}{dt}\langle\psi|\psi\rangle = \frac{-1}{i}\langle\psi|H^{\dagger}(=H)|\psi\rangle + \frac{1}{i}\langle\psi|H|\psi\rangle = 0 \tag{31}$$

For a Hermitian operator the norm of the electronic wave function is conserved. Now in presence of a CAP (V^{abs}), $H = H^{R} + iV^{abs}$, $H_{ij}^{\dagger} = \left\langle i\middle|\left(H^{R} + iV^{abs}\right)^{\dagger}\middle|j\right\rangle = \left\langle j\middle|H^{R} + iV^{abs}\middle|i\right\rangle^{*} = \left\langle j\middle|H^{R} - iV^{abs}\middle|i\right\rangle \neq H_{ij}$. We have now $\frac{d}{dt}|\psi\rangle = \frac{1}{i}H^{R}|\psi\rangle + V^{abs}|\psi\rangle$ and $\frac{d}{dt}\langle\psi| = \frac{-1}{i}\langle\psi|H^{\dagger} + \langle\psi|V^{abs}$. Inserting these equalities in Eq. 3 we have:

$$\frac{d}{dt}\langle\psi|\psi\rangle = 2\langle\psi|V^{abs}|\psi\rangle \tag{32}$$

When the Hamiltonian is not Hermitian the norm of the wave function is not conserved. The idea of introducing a CAP in RT-TDDFT simulations is therefore to allow for the injection or removal of electron density during ED simulations. This is an *ad hoc* methodology that must be used with care to avoid artefacts in the propagation. Ideally the CAP should affect only unbound electrons. Some authors suggested to define CAPs based on the energies of the KS MOs[49]. We have decided to implement in deMon2k a real space based definition of CAP[50]. The CAP function is written as a superposition of atom-centered spherical ${V^{abs}}_{a}$.

$$V^{abs}(R) = \min_{a} {V^{abs}}_{a}(R) \tag{33}$$

$$V^{abs}{}_a(R) = \begin{cases} 0 & for\ R < R° \\ V^{max} \sin^2\left(\frac{\pi}{2W}(R - R°)\right) & for\ R° + W < R < R° \\ V^{max} & R > R° + W \end{cases} \quad (34)$$

where V^{max} is the maximum value of the absorbing potential, W is the width of increase of the atomic CAP and $R°$ is the distance threshold at which the atomic CAP starts.

I.2 Auxiliary Density Functional Theory

Having described the algorithms to propagate the electron density at the TDDFT level, we will now be more specific about the implementation of these algorithms in deMon2k. Indeed a major advantage of deMon2k for stationary DFT based methods is its reliant on density fitting (DF) and other algorithmic tricks to accelerate the calculations[51]. Since our objective is to achieve an efficient implementation of RT-TDDFT building on the already optimized Auxiliary DFT (ADFT) framework we will describe it first in this section. The variational density fitting was originally introduced by Dunlap[52] to avoid the calculation of four-centers electron repulsion integrals (ERIS) entering the KS energy and potential (Eq. 8-9). Besides the Kohn-Sham density ρ which is built on the KS MOs (Eq. 6), an auxiliary density $\tilde{\rho}$ is expressed as a linear combination of auxiliary functions $k(\boldsymbol{r})$: $\tilde{\rho}(\boldsymbol{r}, t) = \sum_k x_k(t) k(\boldsymbol{r})$. The x_k are the density fitting coefficients. The DF formalism was proposed originally for stationary densities. Here we introduce the dependence on time of the auxiliary coefficients for future use in RT-TDDFT. For computational efficiency we use in deMon2k primitive Hermite Gaussian functions at atom K with the exponent ζ_k having the form given by Eq 35[53].

$$k(\boldsymbol{r}) = \left(\frac{\partial}{\partial K_x}\right)^{k_x} \left(\frac{\partial}{\partial K_y}\right)^{k_y} \left(\frac{\partial}{\partial K_z}\right)^{k_z} e^{-\zeta_k(\boldsymbol{r}-K)^2} \quad (35)$$

The fitting coefficients are obtained by minimization of a self-interacting error function[52]:

$$\varepsilon = \frac{1}{2}\langle \rho - \tilde{\rho} \| \rho - \tilde{\rho} \rangle = \frac{1}{2}\langle \rho \| \rho \rangle - \langle \rho \| \tilde{\rho} \rangle + \frac{1}{2}\langle \tilde{\rho} \| \tilde{\rho} \rangle \quad (36)$$

In matrix notation this gives:

$$\varepsilon = \frac{1}{2}\sum_{\mu,\nu}\sum_{\sigma,\tau} P_{\mu\nu}(t)\langle \mu\nu \| \sigma\tau \rangle - \sum_{\mu,\nu}\sum_{k} P_{\mu\nu}(t)\langle \mu\nu \| k \rangle x_k(\mathrm{t}) - \frac{1}{2}\sum_{k,l} x_k(t) x_l(t) \langle k \| l \rangle \quad (37)$$

The function ε is minimized with respect to all the coefficients x_k:

$$\left(\frac{\partial \varepsilon}{\partial x_k}\right)_P = -\sum_{\mu,v} P_{\mu v}\langle \mu v \| k\rangle + \sum_l x_l \langle l \| k\rangle = 0 \;\forall\; k \tag{38}$$

$$\left(\frac{\partial^2 \varepsilon}{\partial x_l \partial x_k}\right)_P = \langle l \| k\rangle \equiv G_{kl} \tag{39}$$

Because the auxiliary function Coulomb matrix G is always positive definite the above variation is indeed a minimization of the fitting error ε. As a result, the following solution for the auxiliary function fitting coefficients is obtained:

$$x(t) = G^{-1} J(t) \tag{40}$$

with the so-called Coulomb vector J being defined as:

$$J_k(t) = \sum_{\mu,v} P_{\mu,v}(t)\langle \mu v \| k\rangle \tag{41}$$

This set of coupled equations can be solved either by analytical or by iterative approaches[54]. In principle the minimization should be carried out under the constraint that the fitted density integrates to the total number of electrons[52]. This can be done with a Lagrange multiplier method. In practice however, experience from the deMon2k developers showed that the fitted density naturally integrates to the desired value without explicit introduction of such a constraint[†]. The latter is now dropped out from the set of fitting equations in deMon2k. Because of the positive nature of the function ε, we have

$$\frac{1}{2}\langle \rho \| \rho\rangle \geq \langle \rho \| \tilde{\rho}\rangle - \frac{1}{2}\langle \tilde{\rho} \| \tilde{\rho}\rangle \tag{42}$$

This inequality permits a variational substitution of the classical Coulomb repulsion in the total energy (Eq. 8).

$$\begin{aligned} E(t) = &\sum_{\mu,v} P_{\mu v}(t) H_{\mu v}^{ext} + \sum_{\mu,v} P_{\mu v}(t)\left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| v \right\rangle \\ &+ \sum_{\mu,v}\sum_k P_{\mu v}(t)\langle \mu v \| k\rangle x_k(\mathrm{t}) - \frac{1}{2}\sum_{k,l} x_k(t) x_l(t)\langle k \| l\rangle + E_{xc}[\rho] \end{aligned} \tag{43}$$

The elements of the Kohn-Sham potential are given by Eq. 44.

$$H_{\mu,v} \equiv \frac{\partial E}{\partial P_{\mu v}} = H_{\mu v}^{ext} + \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| v \right\rangle + \sum_k \langle \mu v \| k\rangle + \frac{\partial E_{xc}[\rho(\boldsymbol{r}, t)]}{\partial P_{\mu v}} \tag{44}$$

[†] A. M. Köster, private communication.

Compared to Eq. 8-9, there are no four-centers ERIS but only two- and three- centers ERIS entering the energy or the potential. This approximation reduces the formal scaling of the Coulomb integral calculation from N_{AO}^4 to $N_{AO}^2.M_{AF}$ with N_{AO} and M_{AF} denoting the number of atomic orbitals and auxiliary basis functions respectively. The number M is typically three to five times the number of basis functions, but this is much less than the number of products of AO.

Once the fitted coefficients obtained for a given Kohn-Sham density, they can be used for the calculation of the exchange-correlation (XC) energy and potential too[55]. $E_{xc}[\rho]$ is replaced by $E_{xc}[\tilde{\rho}]$ in Eq. 43. In that case the matrix elements of the Kohn-Sham potential are given by[55]:

$$H_{\mu\nu} \equiv \frac{\partial E}{\partial P_{\mu\nu}} = H_{\mu\nu}^{ext} + \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| v \right\rangle + \sum_{\bar{k}} \langle \mu\nu \| k \rangle (x_k(t) + z_k(\mathrm{t})) \tag{45}$$

with

$$z_k(\mathrm{t}) = \sum_{l} \langle k \| l \rangle^{-1} \langle l | V_{xc}[\tilde{\rho}(t)] \rangle \tag{46}$$

The combination of density fitting with RT-TDDFT is in principle straightforward. Returning back to Figure 1 and Figure 2 describing the iterative and PC methods of the SOMP, each time a new Kohn-Sham electron density ρ (or equivalently P) is formed by the SOMP, the fitting equations (Eqs. 40-41) are used to determine an auxiliary density. The fitted density is then used to calculate the next Kohn-Sham potential. Now it must be remembered that DF provides an approximate solution to the Kohn-Sham equations that depends for the most part on the choice of the auxiliary basis set. Although ADFT can be regarded as well tested now for stationary DFT calculations[51a] or Linear-Response TDDFT[56] it is not yet clear whether ADFT can be safely used in the context of RT-TDDFT. For example the electron dynamics per se might not be well reproduced in DF based ED simulations. Actually our tests, reported in the article reproduced at the end of Part II[57], showed that provided sufficiently flexible auxiliary basis sets are chosen, one can rely on density fitting techniques to reduce the computational cost of the RT-TDDFT propagations without altering the stability or the accuracy of the propagations. In section II we will report computational performance tests showing that the implementation of RT-TDDFT in deMon2k fully takes advantage of DF techniques of the code.

I.3 Analyzing tools for Electron Dynamics simulations

The formalism described above provides a prescription to simulate attosecond electron dynamics at the DFT level. It is also very important to devise dedicated methodologies to analyze the fluctuating electron density in the course of the simulations. Some physical observables can be obtained by

considering the associated quantum mechanical operator, for example the electronic energy (Eq. 43) or the multipole moments of the system. Other descriptive tools are also valuable and will be described in the following subsections.

I.3.a Molecular multipoles

The total number of electrons $N(t)$ and higher multipole moments such as the dipole moment $\boldsymbol{\mu}(t)$ and quadrupole tensor are given by Eqs. 47-49. The total spin charge (S) defined as the integrated difference between densities of alpha and beta electrons can also be calculated. It is of course only of interest for open-shell systems.

$$N(t) = \int \rho(\boldsymbol{r}, t)\, d\boldsymbol{r} \tag{47}$$

$$\boldsymbol{\mu}(t) = \sum_A Z_A \boldsymbol{R}_A - \int \rho(\boldsymbol{r}, t)\boldsymbol{r}.\, d\boldsymbol{r} \tag{48}$$

$$\boldsymbol{\Theta}^{\boldsymbol{A}}_{\boldsymbol{\alpha,\beta}}(t) = -\int \rho(\boldsymbol{r}, t)(r_{i,\alpha} - r_{A,\alpha})(r_{i,\beta} - r_{A,\beta}).\, d\boldsymbol{r} \tag{49}$$

$$S(t) = \int \left(\rho^{\alpha}(\boldsymbol{r}, t) - \rho^{\beta}(\boldsymbol{r}, t)\right) d\boldsymbol{r} \tag{50}$$

For isolated QM systems, the evolution of the total number of electrons is a good indicator of the stability of the ED simulation since this should remain constant. Variation of the number of electrons is a sign that the propagation is not unitary, *i.e*. that the simulation parameters are not well tuned for the system under investigation. When simulating collisions with charged particles one expected outcome is ionization of the molecule and emission of unbound electrons. A way to capture this phenomenon with a code working with localized atomic orbitals is to add a complex absorbing potential (CAP) in the imaginary part of the KS potential[50]. The CAP only acts on electron density typically beyond 10 Å of the molecule by removing these electrons from the pool. In this particular type of simulations N doesn't have to be conserved. On the contrary it describes ionization of the molecule since the CAP collects (in ideal situations) the density of unbound electrons.

The dipole moment is a key quantity to describe the deformation of the electron cloud upon application of a perturbation. It gives access to the polarizability tensor which is a basic quantity to simulate absorption spectra[16b, 58].

I.3.b Intrinsic multipoles on atoms or fragments

As for stationary DFT calculation the total density can be projected onto atoms. Therefore one can define atomic charges and so-called intrinsic atomic dipoles and quadrupoles provided a prescription for defining the atoms within the molecules are specified. In deMon2k Hirshfeld[59], Becke[60] and Voronoi deformation density[61] are for example available and have been made compatible with the use of RT-TDDFT[62]. Atomic charges are useful to describe charge transfers among atoms during ED simulations. Charge migrations[63] for example can be highlighted with this tool. On the other hand fluctuations of the intrinsic atomic dipoles provides information on the internal polarization of each atom.

Finally, the level of coarse graining of population analyses can be adapted to the level of fragments.

I.3.c Occupation of Kohn-Sham molecular orbitals

One way to analyze the ED is to define a set of KS MOs ($\boldsymbol{C}^k$) as reference and to follow their occupation numbers (n_k) during the simulation. The set of MOs obtained at convergence of the SCF procedure is used for example. The occupation numbers are obtained as $n_k(t) = \boldsymbol{C}^{k\dagger}\boldsymbol{P}'(t)\boldsymbol{C}^k$ where the $\boldsymbol{C}^k$ are the eigenvectors of the Kohn-Sham potential obtained from a SCF calculation and $\boldsymbol{P}'$ is the time dependent density matrix in the MO basis. Such analyses identify which MOs are associated with an electronic transition or ionization. It should be remembered that since any unitary transformation of the set of reference KS MOs is equally valid to determine the electron density and the electronic energy, this type of analysis is somehow arbitrary.

I.3.d Kinetic energy of electrons

To help understand electronic relaxation following perturbation one can evaluate the electron kinetic energy. Following Bader and Preston[64] the total kinetic density on each point in space ($K(\boldsymbol{r})$) is determined by the Laplacian of the total density ($L(\boldsymbol{r})$) and by the gradient of its components ($G(\boldsymbol{r})$).

$$K(\boldsymbol{r}) = L(\boldsymbol{r}) + G(\boldsymbol{r}) \tag{51}$$

$$L(\boldsymbol{r}) \equiv -\frac{1}{4}\nabla^2\rho(\boldsymbol{r}) \tag{52}$$

$$G(\boldsymbol{r}) \equiv \frac{1}{8}\sum_i \frac{\nabla\rho_i(\boldsymbol{r}).\nabla\rho_i(\boldsymbol{r})}{\rho_i(\boldsymbol{r})} \tag{53}$$

Integration of K over the entire space leads to the total electron kinetic energies while L, which can locally take positive or negative values, globally integrates to zero[64]. When the Laplacian contribution is integrated over atoms, or over molecular fragments we found that $\int Ldr$ is already close to zero (typically 10^{-2} Ha). The investigation of the kinetic energy per fragment provides a very insightful tool

to analyze sub-femtosecond relaxation processes. This tool has been used for example by the TheoSim group recently to investigate ionization following collision of the guanine DNA base by charged particles[45].

I.3.e Graphical representation of molecular fields

Electron Dynamics simulations are particularly amenable to pictorial representations. It is for example instructive to calculate the electron density ($\rho(\boldsymbol{r},t)$), the deformation density ($\rho(\boldsymbol{r},t)-\rho(\boldsymbol{r},t_0)$) or other fields such as the Time-Dependent Electron Localization Function[65] on grids of points and later to visualize them with appropriate software (e.g. VMD[66]). For illustrative purpose we depict in Figure 4 the negative and positive isosurfaces of the deformation density for a benzene molecule subjected to an external electric field. The energy of the light corresponds to the first excitation energy of benzene. A recurrence of 0.6 fs is observed.

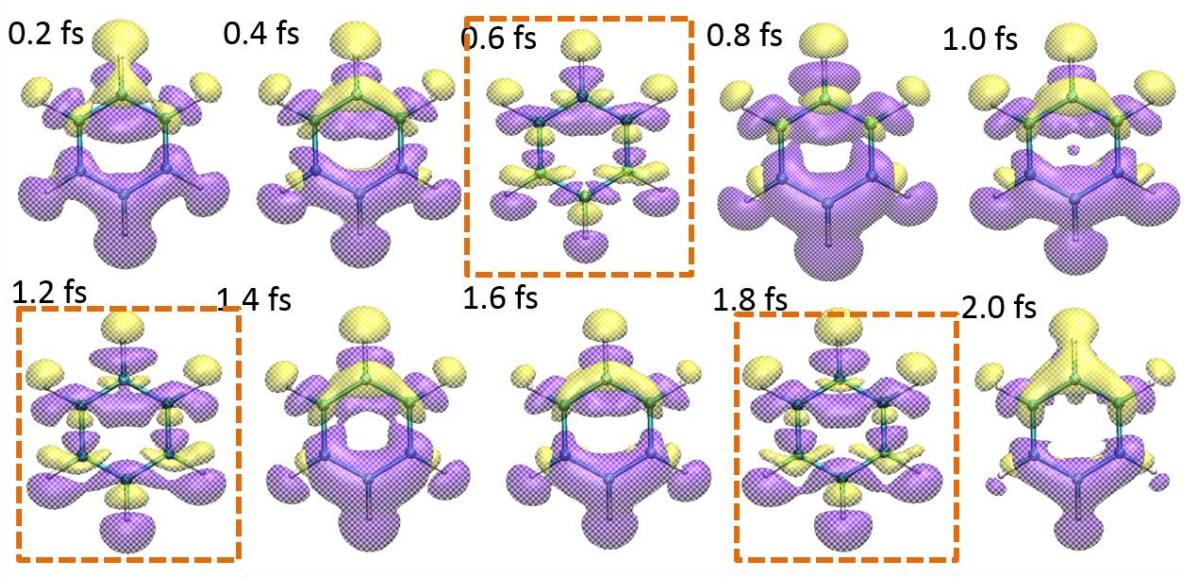


Figure 4. Isosurfaces of the deformation density (0.01 bohr^{-3}) of benzene subjected to a Gaussian monochromatic pulse corresponding to the transition from the ground to the first excited state. The field was oriented along the top-botton direction defined by the page. Yellow and violet surfaces correspond to region of accumulation and depletion of electron density. The orange rectangles illustrate the periodicity in the electron flows.

I.4 Implementation in deMon2k

deMon stands for "**de**nsity of **Mon**tréal". It refers to a computer code based on DFT that was created in the 1980's in the group of Prof. Salahub[67]. deMon2k is a new version of the program that emerged in the early 2000's from the fusion of the deMon and ALLCHEM programs. deMon2k has been

developed by various scientific groups since then, essentially in the CINVESTAV center (Prof. A. M. Köster group). deMon2k contains around 250,000 lines and 1,929 routines. It is programmed in Fortran 77 and parallelized using the Message Passing Interface (MPI) protocol. The implementation of a complete new functionality like RT-TDDFT in such a complex code is not a trivial task. It is mandatory to think thoroughly in advance about the implementation strategy which supposes a good knowledge of the structure of the code. This is an important step to avoid creating interferences ("bugs") with other parts of the code - which would make the implementation useless at the end - and to optimize efficiency.

We started to implement a basic serial version of RT-TDDFT for closed-shell molecules with the Euler propagator. This first work involved i) the introduction of new keywords into the code, the modification of existing options, ii) the coding of the density propagation engine, iii) the routines for performing back-and-forth transformations between the AO and MO representations, iv) the inclusion of time dependent electric fields. The structure of the RT-TDDFT has been decided at this stage. Our strategy was to borrow as much as possible the already existing subroutines. Effectively the strength of deMon2k is that it is equipped with particularly efficient algorithms to evaluate molecular integrals in the framework of DF. Toward this end the structure of the RT-TDDFT engine was devised following the same algorithmic structure as the SCF module. In this way, it has been possible to branch entry point subroutines for calculation of the KS potential (bldksmat.f) or for performing density fitting (getcdj.f or getcdc.f and subroutines therein). The basic implementation was carefully tested by comparison of results produced by deMon2k and by another code (NWchem[12]). Among the simplest diagnostics for testing the stability of the propagations are the conservation of electronic energy or total electron number. The polarizabilities and the absorption spectra of molecules obtained from RT-TDDFT were compared with results obtained by finite difference or perturbative approaches. We also carefully tested under which conditions auxiliary densities could be used in RT-TDDFT. We found that auxiliary basis sets containing polarization functions - the GEN-An* sets, to take the deMon2k syntax - were sufficient to achieve accurate simulations. We refer to the article reproduced at the end of Part II for details on all the validation tests.

Once validated, the basic implementation was continuously improved to add more functionalities. The list of functionalities now available are listed below. I coded myself most of the subroutines required to have these new options, some being coded or improved by other members of the group.

- Both open-shell and closed-shell systems are amenable to RT-TDDFT simulations

- Magnus propagator in both iterative and PC versions (described in I.1.b).

- Taylor, Chebyshev and BCH expansions for evaluating the exponential of complex matrices (described in I.1.c).

- Analysis tools during ED simulations (described in I.3).

- Combination with the CONVENTIONAL, DIRECT or MIXED schemes. These are options available in deMon2k to handle the calculation and storage of ERIS.

- Restart calculations.

- Parallelization of the code using the pre-existing MPI architecture in deMon2k.

- Complex absorbing potentials.

II. COMPUTATIONAL PERFORMANCE AND VALIDATION

Various validation tests of the RT-TDDFT module have been reported in the article reproduced at the end of Part II. These tests involved the calculation of static polarizabilities or of absorption spectra. In Section II we report complementary tests calculations and more in-depth analyses of the efficiency of the implementation. We will start by comparing the iterative and PC version of the SOMP (II.1). Then we report comparison of the different methods available in deMon2k for calculating the exponential of a matrix (II.2).

II.1 On the efficiency of the Propagator-Corrector scheme

In the methodology section we introduced two propagation methods for the SOMP, namely the iterative and Predictor-Corrector (PC) algorithms. Both have their own pro and cons. The objective of this section is to compare the relative efficiency of the two approaches, and to analyze the computational performance of our implementation in deMon2k.

II.1.a Computational details

We consider two molecules. One is coumarin which is a seventeen-atom organic molecule. The other one is an Fe(II) inorganic complex (heme) containing ninety-nine atoms. Both molecules have been geometrically optimized at the DFT level (Figure 5).

For the ED simulations we have used the PBE functional and the DZVP-GGA/GEN-A2* combination of atomic and auxiliary basis sets. This choice leads to a total of 84/1155 and 466/6130 atomic/auxiliary functions for coumarin and heme respectively. The initial electron density of the ground state was

obtained by a stationary DFT energy minimization through an SCF procedure. To collect computational performance data the simulations have been propagated during 200 as. We used the BCH expansion to perform the propagation steps. We have added as many terms as necessary in the BCH expansion to arrive at a convergence threshold of 10^{-30} bohr^{-3}. The intrinsic atomic multipoles (charge, dipole, and quadrupoles) have been calculated on-the-fly at every time step by numerical integration of fitted densities with a grid of medium accuracy[62]. To integrate the exchange-correlation potential and energy we have used fitted densities and a very fine grid of points associated to an accuracy of 10^{-8} Ha on diagonal elements of the XC matrix. The Electron Repulsion Integrals (ERIS) are handled with a mixed scheme to optimize computational efficiency of the Coulomb repulsion terms. Table 1 and Table 2 gather the timings for the two molecules. All the simulations have been run on twelve 2.67GHz Intel® Xeon® X5650 processors with the MPI protocol.

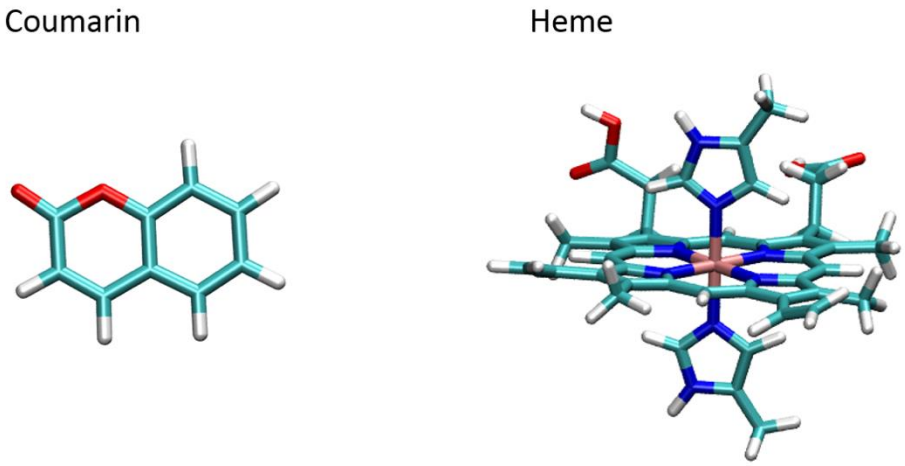


Figure 5. Optimized structure of coumarin and heme in gas phase. White: hydrogen, red: oxygen, blue: nitrogen, pink: iron, cyan: carbon.

II.1.b Stability and performance for weak perturbations

We start to investigate weak perturbations by an infinitely narrow electric pulse. In the first step of RT-TDDFT propagation an electric kick is applied to the electron cloud. This creates a perturbation in the electron density that, in turn, starts to fluctuate. The perturbing electric field vector is aligned on the x-axis and its strength is set to 0.0001 a.u. (0.05 V/nm, which corresponds to an intensity of 7.03×10^{8} W·cm^{-2}). Four mutually exclusive entries are reported in Table 1. The entry "AO<->MO " refers to the total time spent in switching from the AO to the MO basis (and *vice versa*). The entry "SOMP" refers to the calculation of $U(t,t_0)P'(t_0)U^{\dagger}(t,t_0)$, excluding the time taken to calculate the potential itself. The SOMP time is essentially the time taken to evaluate the exponential of the matrix with the BCH

expansion. The KS potential refers to the time spent in calculating the KS potential. This timing includes all the operations related to density fitting. Finally the entry "population analyses" refers to the atomic multipole calculations.

Table 1. Computational performances of the iterative (IT) and PC propagation algorithms. The reported wall-clock timings (in s) refer to a 200 as RT-TDDFT ED simulation on a coumarin molecule. The number in brackets are deviations from ideal scaling (see text for details).

Time step (as)	1		2		3		10	
	IT	PC	IT	PC	IT	PC	IT	PC
AO<->MO	18.4	12.2	9.4	6.2	6.2	4.1	1.9	1.2
			(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)
SOMP	121.1	72.5	77.0	45.6	59.4	33.2	31.5	17.1
			(0.8)	(0.8)	(0.7)	(0.7)	(0.4)	(0.4)
KS potential	624.9	219.7	310.6	106.9	196.3	68.1	60.1	19.8
			(1.0)	(1.0)	(1.1)	(1.1)	(1.0)	(1.1)
Population analysis	16.6	16.7	8.2	8.3	5.5	5.5	1.6	1.6
			(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)
Total time	774.5	322.8	402.5	168.4	265.8	112.1	95.4	40.8
			(1.0)	(1.0)	(1.0)	(1.0)	(0.8)	(0.8)
cost for 1 step	3.9	1.6	4.0	1.7	4.0	1.7	4.8	2.0
cost for 1 fs	3872.5	1613.9	2012.6	842.1	1322.3	557.9	476.9	204.2

Starting with coumarin we note that the PC algorithm is average 2.44 times faster than the iterative algorithm whatever the value of Δt. This is explained by the higher number of KS potential calculations and of Magnus propagation steps required in the iterative scheme. On average three iterations are needed for each propagation step, leading to three KS potential calculations and three SOMP steps. With the PC scheme, only one KS potential calculation and two SOMP steps are needed for each propagation step. The evaluation of the KS potential is the most time consuming, followed by the SOMP step. Transformation between the atomic orbitals and molecular orbitals as well as population analysis are much less computationally demanding.

To analyze the computational time more in depth, we report in Figure 6 a pie chart illustrating the proportion of the different tasks. For space reasons, we consider only simulations with 1 as time steps. Matrix multiplications are needed in atomic orbital to molecular-orbital basis transformations and in Magnus propagation. Together they represent 18% of the total time. In fact basis transformation is a minor task representing only 2.3%. The time for KS potential evaluation encompasses three main contributions, the calculation of the XC potential, the calculation of electron Coulomb repulsion terms in which we include both ERIS calculations and density fitting steps, and finally, the calculation of the core Hamiltonian. The XC contribution represents 76 % of the overall cost. We remark that a very accurate grid has been chosen, inducing a significant computational cost. Calculation of the Coulomb contribution is almost negligible (2%). This remarkable performance is achieved thanks to the mixed scheme implemented in deMon2k to store short-range ERIS in RAM and to compute long-range ERIS

by double asymptotic expansions[68]. The calculation of core integrals is less than 1% only. In fact because the core Hamiltonian is independent on the density it only needs to be calculated once at the beginning of the propagation. The repeated analysis of the time-dependent electron density induces only a small supplementary cost to the calculation (<1%).

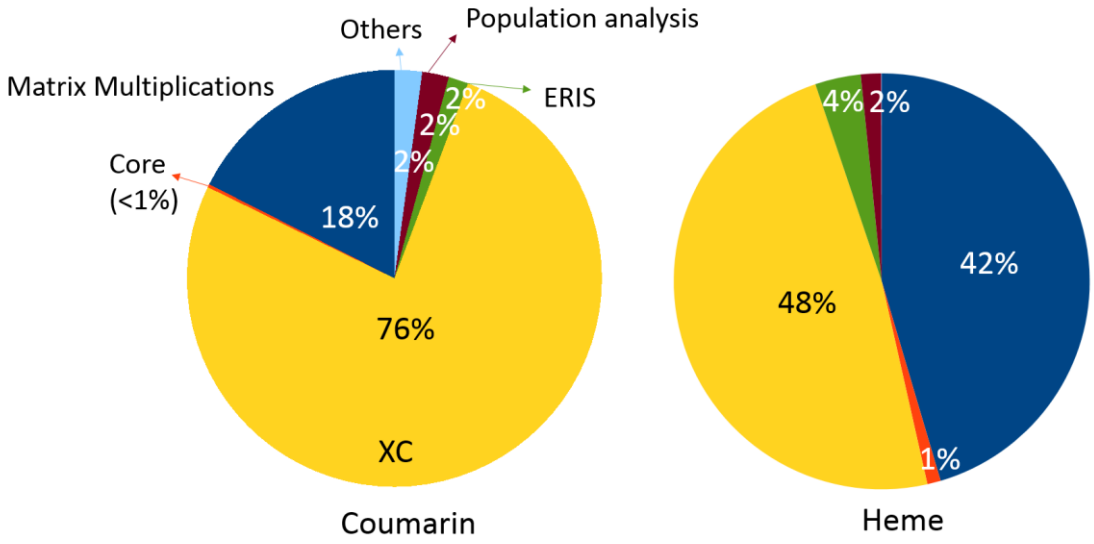


Figure 6. Computational timing for 200 as of RT-TDDFT ED simulations of coumarin and heme. The time step was set to 1 as and the iterative SOMP was chosen. A grid of high accuracy was employed to evaluate the XC contribution to the potential.

We now look at the influence of the time step size on the computational performance. The evolution of overall simulation time follows the size of Δt. For example it takes 332.85 s to perform 200 as of simulation with a time step of 1 as (PC), but only 112.1 s with a Δt of 3 as, that is a reduction by factor 3. To quantify the scaling we provide in Table 1 numbers quantifying the deviation from an "ideal scaling" (numbers in brackets). For each entry this indicator is calculated from the ratio of timings using the 1as time step ED and multiplied by the ratio of integration time step. Therefore a value of 1 indicates the computational effort for this entry is strictly proportional to Δt. This is the case for instance of population analyses that are obtained by numerical integration on a grid of points. This is an operation which is totally disconnected from the choice of propagation algorithm. Therefore the total time for population analysis depends only on the number of propagation steps and not on the length of the time step. Most indicators are close to one. This indicates that increasing the time step doesn't induce an increase of computational cost of each step. The only deviation from unity we observe is for the "Magnus propagation" entry which amounts to 0.7 and even 0.4 for a 3 as and 10 as time steps respectively. This is due to the higher number of nested commutators that are included in the BCH expansion to ensure stability of the propagation.

Alternatively said, the computational time saving obtained by choosing a larger time step is partially lost by the fact that more terms are needed in the evaluation of the BCH expansion to ensure stability. Nevertheless one can increase the integration time length without affecting propagation stability and still gain a proportional advantage in terms of computer time. This is the case both for the iterative and the propagator-corrector methods.

Most of these conclusions are relevant for ED simulations on the larger heme complex (Table 2). The PC algorithm is average 2.11 times faster than the iterative one. A main difference is, as expected, the higher overall computational cost which is due to the much larger number of basis functions in heme compared to coumarin. Matrix multiplication (42%) and XC potential calculations (48%) are now almost equally time consuming (Figure 6). Therefore for a larger system like heme, the scaling for linear algebra is less favorable than for the numerical integration of the XC contributions on grids of points.

Table 2. Computational performances of the iterative and PC propagation algorithms. The reported wall-clock timings (in s) refer to a 200 as RT-TDDFT ED simulation on a heme complex.

Time step (as)	1		2		3	
	IT	PC	IT	PC	IT	PC
AO<->MO	586.8	3,967	293	197	195	132
			(1.0)	(1.0)	(1.0)	(1.0)
SOMP	11,826.4	6,764	9,149	5,004	8,073	4,409
			(0.6)	(0.7)	(0.5)	(0.5)
KS potential	15,319.6	5,085	7,502	2,550	5,076	1,700
			(1.0)	(1.0)	(1.0)	(1.0)
Population analysis	431.0	435	215	214	143	143
			(1.0)	(1.0)	(1.0)	(1.0)
Total time	27,325.4	12,455	16,747	7,863	13,218	6,320
			(0.8)	(0.8)	(0.7)	(0.7)
cost for 1 step	136.6	62.3	168	79	197	94
cost for 1 fs	136,626.9	31,136.6	167,473	39,314	65,761	31,446

Stable simulations for heme turned out to be more difficult to achieve. For instance no stable simulation could be obtained with a 4 as integration time step[‡]. It also becomes less and less efficient to increase the time step as more and more terms are needed to converge the BCH expansion, which results in an increase of the computational cost of each time step. For example, a single PC propagation step takes 1.04 min *vs*. 1.57 min with 1 as or 3 as time step lengths respectively. This is probably due to the more complex electronic structure of heme compared to coumarin (e.g. presence of d orbitals on the iron cation). A recent study also suggested that core electrons are those animated by the fastest motion thereby setting a lighter limit to the integration time length. The deep iron core electrons might therefore also contribute to making RT-TDDFT ED simulations of heme difficult. Satisfactorily though,

[‡] When using the more robust diagonalization of the propagator the simulation remained stable with a 10 as integration time step suggesting that the instability was due to numerical error accumulation in the BCH routines.

we find that the iterative as well as the PC algorithms implemented in deMon2k permit to carry out such simulations in reasonable times.

II.1.c Stability and performance for strong perturbations

The PC scheme has a clear advantage in computational time compared to the iterative one. However in the previous tests, a weak perturbation was applied – an infinitely narrow electric kick of strength 0.0001 a.u (0.1 V/nm). We now analyze similar simulations with a stronger electric field over a longer period of time. The mathematical expression of the external electric field is given by:

$$\boldsymbol{F}(t) = F_{max} exp[(-(t-t_0)^2/2{u_F}^2]\cos(\omega t)\boldsymbol{d} \qquad (54)$$

where F_{max} is the maximum field strength, t_0 and u_F are respectively the center and width of the Gaussian envelope and ω is the frequency of the electric field. These parameters were set to 0.005 a.u. (5 V/nm), 0.1 fs, 0.015 fs and 0.114 a.u, respectively. The value for ω is the excitation energy of 3.10 eV (400 nm), which corresponds to typical Soret band of heme systems. Therefore application of this electric pulse should trigger transition of the ground state density to the Soret band. Results are shown in Table 3.

We find again a computational advantage of the PC over the iterative scheme. In fact most of the conclusions drawn for the weak kick are also transferrable to the case of a Gaussian pulse.

Table 3. Computational performances of the iterative (IT) and PC propagation algorithms. The reported wall-clock timings (in s) refer to a 200 as RT-TDDFT ED simulation with a Gaussian pulse on a heme.

Time step (as)	1		2		3	
	IT	PC	IT	PC	IT	PC
AO<->MO	714	461	350	238	263	159
			(1.0)	(1.0)	(0.9)	(1.0)
SOMP	14,857	8,103	11,172	6,294	11,061	5,507
			(0.7)	(0.6)	(0.4)	(0.5)
KS potential	21,318	7,174	10,437	3,490	7,911	2,395
			(1.0)	(1.0)	(0.9)	(1.0)
Population analysis	465	467	233	234	156	156
			(1.0)	(1.0)	(1.0)	(1.0)
Total time	36,437	15,965	21,742	10,153	19,063	8,161
			(0.8)	(0.8)	(0.6)	(0.7)
cost for 1 step	182	80	217.4	102	285	121
cost for 1 fs[a]	182,183	79,826	108,709	50,764	94,839	40,604

a: value extrapolated from an ED simulation run for 200 as

II.2 On the efficiency of the exponential of an operator

As introduced in section I, the methods available in deMon2k to evaluate the exponential of the matrix entering the SOMP are i) the diagonalization of the e^W (*diago*), ii) the Baker–Campbell–Hausdorff formula (*BCH*), iii) the Taylor expansion (*Taylor*) of the e^W and for iv) the series expansion of e^W using

Chebyshev polynomials (*Chebyshev*). The numerical performances of each type of method depend on the nature of linear algebra operations, and for this reason are system dependent. Methods ii, iii and iv depend on a user defined convergence criterion. According to this criterion, different expansion terms (k) will be used (Eqs. 21-25). We have considered five molecular systems consisting of a coumarin molecule solvated by droplets of water molecules of different sizes (Figure 7). Systems S1, S2, S3, S4 and S6 respectively encompass 76, 196, 436, 916 and 1256 electrons and 195, 495, 1095, 2295 and 3145 atomic orbitals (DZVP-GGA/GEN-A2* combination of atomic orbitals and auxiliary basis sets). The geometries have been extracted from a classical MD simulation of coumarin solvated in a box of solvent using periodic boundary conditions.

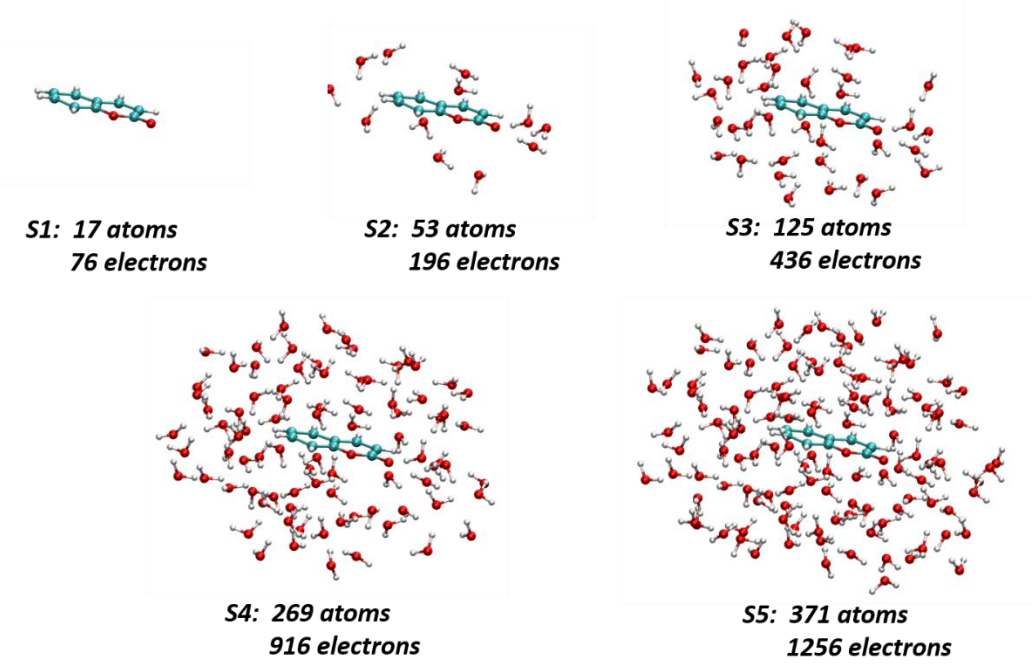


Figure 7. Five molecular systems used for testing different methods to perform the SOMP.

We report in Table 4 the wall-clock timings corresponding to a propagation of the electron density for 400 steps. The SOMP/PC scheme has been used to simulate the response of the ground state electron cloud to an electric field kick of strength 0.001 a.u. applied along the x-direction. We switched for this series of tests from the highly accurate (10^{-8} Ha) numerical grid[§] as used in the previous section to a grid of accuracy 10^{-7} Ha. The latter grid permits a noticeable decrease of computational cost of the XC contribution to the Kohn-Sham potential while still retaining high accuracy. It is also more

[§] deMon2k uses Lebedev grids centered on atoms to integrate XC contributions. The number of angular points for each radial shell is optimized for the molecular structure with an accuracy criterion applying on the diagonal elements of the XC potential matrix. This is the criteria we are referring to in the main text.

representative of the computational set-ups now commonly used in the group to carry out RT-TDDFT simulations with deMon2k. For the *BCH*, *Taylor* and *Chebyshev,* iterations over the number of terms introduced in the series were pursued with a tolerance criterion of 10^{-30}. All the simulations have been carried out on the Cedar supercomputer hosted by ComputeCanada (Intel Broadwell® CPUs at 2.1Ghz, model E5-2683 v4). Three 96 tasks distributed over 3 nodes have been used. The available RAM memory enabled the use of the MIXED ERIS option of deMon2k even for the largest system, thereby avoiding recalculation of Coulomb integrals in the course of the ED simulations.

Table 4. Computational timings (in s) using different algorithms for evaluating the matrix exponential entering the SOMP.

		KS potential	SOMP	AO<->MO	Total
Diagonalization					
	S1	57	316	277	687
	S2	207	1,878	419	2,589
	S3	681	16,408	852	18,270
	S4	2,632	195,781	3,644	206,860
	S5	3,797	398,520	6,656	730,796
BCH					
	S1	59	1,472	148	1717
	S2	214	2,612	268	3174
	S3	795	8,787	955	10,929
	S4	2,601	31,335	3,686	39,308
	S5	6,364	70,295	7,642	96,962
Taylor					
	S1	60	1,725	145	1,966
	S2	199	3,511	325	4,110
	S3	706	8,681	861	10,571
	S4	2,270	31,083	3,116	39,044
	S5	5,646	69,446	6,888	84,566
Chebyshev					
	S1	59	2,164	141	2,398
	S2	194	4,043	289	4,604
	S3	701	11,598	868	13,482
	S4	2,432	45,028	3,482	52,594
	S5	5,612	92,781	6,948	107,853

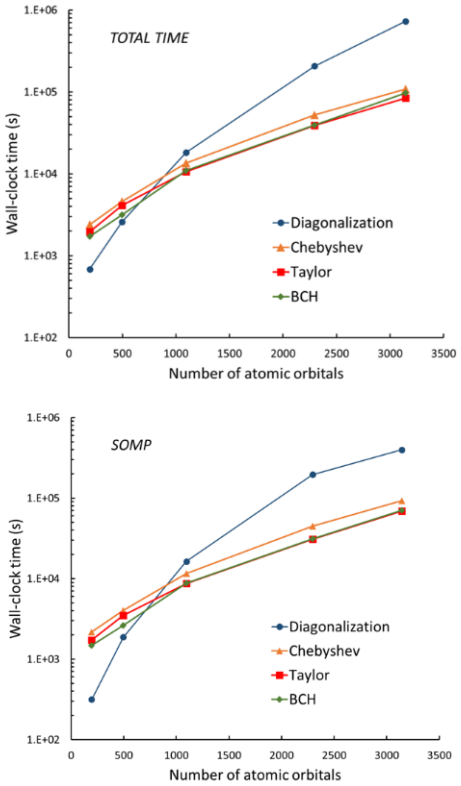


Figure 8. Computational timings for propagating RT-TDDFT simulations on coumarin with different matrix exponentiation methods. Top: Total time of the computation. Bottom: timing of the SOMP step

Table 4 collects the timings for the five systems with different exponentiation matrix methods. The total time is decomposed into three main tasks, namely "Kohn-Sham potential", SOMP and "AO<->MO" transformation (secondary tasks that contribute slightly to the total time are not collected in the Table). The "AO<->MO" transformations are not computationally time consuming. As expected the time taken for basis transformations is independent on the exponentiation method for a given cluster size. The small differences we see for different simulations (for example 6,656 4,642, 6,888 and

6,948 for S5) can be attributed to differences in machines and speed on the internode connection at the moment of the simulations. The SOMP is by far the most time consuming task in all the simulations. This is a consequence of the decrease of the grid accuracy for the XC contribution. This result shows that our implementation of RT-TDDFT takes full advantage of the algorithmic machinery available in deMon2k to evaluate molecular integrals (Density fitting, mixed ERIS scheme, XC numerical integration…). It therefore places the bottleneck of RT-TDDFT simulation in the SOMP steps. To better analyze this point we depict in Figure 8 these timings as a function of the number of atomic orbitals (N_{AO}). The diagonalization method which formally scales as N_{AO}^3 becomes uncompetitive after around 700 atomic orbitals. It is on the other hand far more efficient for small systems (<50 atoms) than the expansion methods. The Chebyshev, Taylor and BCH expansions show similar scaling with a slight advantage for the Taylor scheme. It should be stressed though that the comparison of the three methods is dependent on the choice of the tolerance threshold. In Figure 9 we represent the dipole moment evolution for the largest S5 system during 100 as. It is clearly similar for the three types of propagation.

We depict in Figure 8 a graphical representation of the total and SOMP timings. The expansion methods are more advantageous to simulate large molecular systems. Nevertheless the scaling of the implementation is not fully satisfactory. Indeed for the S5 cluster that contains 1256 electrons and 3145 N_{AO}, it takes 23 h to run 100 as on 96 processors with the Taylor expansion. This seems to set-up a limit to the maximum size to the molecular systems that can be investigated with deMon2k. Reduction of the computational cost of the SOMP step is therefore highly desirable to perform ED simulations on larger systems comprising for instance 1,000 atoms. Actually the analysis of our first implementation of the expansion method indicates that it is not optimum. For simplicity we borrowed the already existing subroutines available in the code to perform the matrix multiplications needed in Eqs. 21-25 (subroutine mpmumat.f[69]). This subroutine is adequate to perform one single matrix multiplication using the MPI protocol. It is used for instance in the SCF driver. However in the context of many successive matrix multiplications as required for the Taylor, Chebyshev or BCH expansion, it becomes unproductive; for example because the mpmulmat starts by distributing the rows of the matrix to the slave nodes. When applying this subroutine successively, the rows of the initial matrix are repeatedly sent to the slaves, which is not necessary. This is probably a main source of loss of computational efficiency as communications among processors are slow processes. An alternative would be to resort to dedicated subroutines (e.g. SCALAPACK) that optimize the workload among processors or to use different types of architectures like GPUs (Graphical Processor Units).

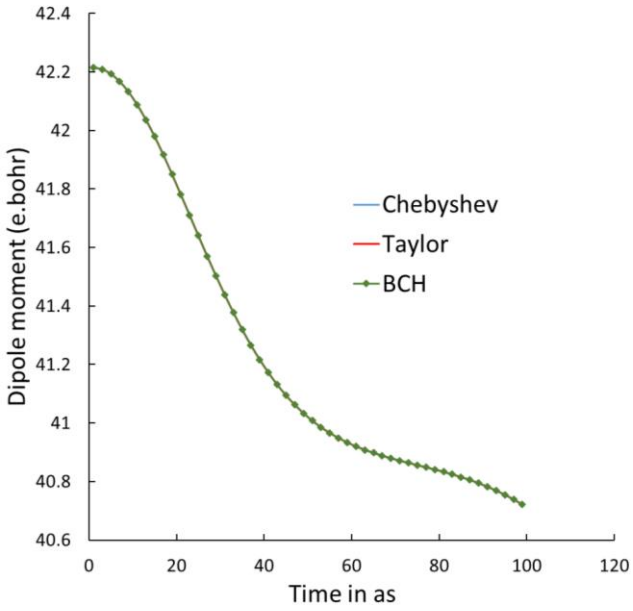


Figure 9. Dipole moment for the S5 systems during 100 as using three different exponentiation methods.

CONCLUSION

In this chapter, we have introduced the theory of RT-TDDFT to simulate ultrafast electron dynamics. An efficient and robust implementation of this method has been realized in deMon2k. With this implementation, real-time electron density response to different external perturbations in large molecular systems are possible. For propagating the electron density, we have chosen the second-order Magnus propagator (SOMP) which is a well-balanced propagator between stability and performance. Various alternative algorithms to achieve this propagation have been implemented to different systems and perturbations. For the external perturbation, simulate different types of electric fields and inelastic collisions with a charged particle may be simulated. Several tools are available to analyze the real time fluctuating electron density including molecular multipoles, intrinsic multipoles, occupation of molecular orbitals, kinetic energy of electrons and graphical representation of molecular fields.

Thanks to the use of the density fitting formalism, our implementation is very efficient. Tests of auxiliary densities with RT-TDDFT show that using auxiliary basis sets containing polarization functions (GEN-An*) is sufficient to achieve accurate simulations. Moreover, during the implementation, we took full advantage of the highly optimized algorithmic machinery in deMon2k such as the mixed ERIS scheme and XC numerical integration. We validated our implementation by comparing polarizabilities and the absorption spectra of molecules with results obtained by finite difference or perturbative approaches.

Performance tests of different algorithms and depth analyses of efficiency of the implementation have been carried out. In the test of two different versions of SOMP, we found that a computational advantage of the Predictor-Corrector over the iterative scheme in both weak and strong perturbations. The speed up depends both on the systems and the time step. Roughly speaking, PC is around 2.2 times faster than iterative schemes. With the tests of different methods for calculating the exponential of a matrix, we conclude that diagonalization is efficient for small systems containing less than 500 atomic orbitals, while expansion methods (Chebyshev, Taylor and BCH) should be used for systems of larger sizes. A molecular system that contains around 3000 atomic orbitals can be carried out with our implementation. This is achievable thanks to the use of fitting density and efficient algorithms in deMon2k which reduce considerably the computational time of the KS potential.

The bottleneck of RT-TDDFT simulation is in the SOMP steps. Future efforts will be pursued to increase the efficiency and stability of this propagation. The efficiency can be improved by optimizing the workload among processors of MPI or to use different types of architectures like GPUs (Graphical Processor Units). Other propagators and algorithms could be tested to increase the stability such as higher order truncated Magnus expansion, commutator free Magnus schemes[39]. We would like to update the code to be compatible with hybrid functionals which will introduce a complex KS matrix. special functional should be implemented to go beyond the adiabatic approximation of the exchange-correlation part.

REFERENCES

1. Runge, E.; Gross, E. K. U., Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984,** *52* (12), 997.
2. (a) Casida, M. E., Time-dependent density-functional response theory for molecules: Focus on functionals. *Abstr. Pap. Am. Chem. S.* **1997,** *213*, 271-COMP; (b) Theilhaber, J., Ab initiosimulations of sodium using time-dependent density-functional theory. *Phys. Rev. B.* **1992,** *46* (20), 12990.
3. (a) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R., Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold. *J. Chem. Phys.* **1998,** *108* (11), 4439; (b) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J., Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **2008,** *128* (4), 044118.
4. (a) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U., Trajectory Surface Hopping within Linear Response Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2007,** *98* (2), 023001; (b) Craig, C. F.; Duncan, W. R.; Prezhdo, O. V., Trajectory Surface Hopping in the Time-Dependent Kohn-Sham Approach for Electron-Nuclear Dynamics. *Phys. Rev. Lett.* **2005,** *95* (16), 163001.
5. Calvayrac, F.; Reinhard, P. G.; Suraud, E., Nonlinear plasmon response in highly excited metallic clusters. *Phys. Rev. B.* **1995,** *52* (24), R17056.
6. Yabana, K.; Bertsch, G. F., Time-dependent local-density approximation in real time. *Phys. Rev. B.* **1996,** *54* (7), 4484.

7. Tong, X.-M.; Chu, S.-I., Time-dependent density-functional theory for strong-field multiphoton processes: Application to the study of the role of dynamical electron correlation in multiple high-order harmonic generation. *Phys. Rev. A* **1998,** *57* (1), 452.
8. Penka, E. F.; Couture-Bienvenue, E.; Bandrauk, A. D., Ionization and harmonic generation in CO and H_2CO and their cations with ultrashort intense laser pulses with time-dependent density-functional theory. *Phys. Rev. A* **2014,** *89* (2), 023414.
9. (a) Castro, A.; Appel, H.; Oliveira, M.; Rozzi, C. A.; Andrade, X.; Lorenzen, F.; Marques, M. A. L.; Gross, E. K. U.; Rubio, A., octopus: a tool for the application of time-dependent density functional theory. *Phys. Status. Solidi. B.* **2006,** *243* (11), 2465; (b) Andrade, X.; Alberdi-Rodriguez, J.; Strubbe, D. A.; Oliveira, M. J. T.; Nogueira, F.; Castro, A.; Muguerza, J.; Arruabarrena, A.; Louie, S. G.; Aspuru-Guzik, A.; Rubio, A.; Marques, M. A. L., Time-dependent density-functional theory in massively parallel computer architectures: the OCTOPUS project. *J. Phys-Condens. Mat.* **2012,** *24* (23); (c) Andrade, X.; Aspuru-Guzik, A., Real-Space Density Functional Theory on Graphical Processing Units: Computational Approach and Comparison to Gaussian Basis Set Methods. *J. Chem. Theor. Comput.* **2013,** *9* (10), 4360.
10. (a) Li, X.; Smith, S. M.; Markevitch, A. N.; Romanov, D. A.; Levis, R. J.; Schlegel, H. B., A time-dependent Hartree–Fock approach for studying the electronic optical response of molecules in intense fields. *Phys. Chem. Chem. Phys.* **2005,** *7* (2), 233; (b) Liang, W.; Chapman, C. T.; Li, X., Efficient first-principles electronic dynamics. *J. Chem. Phys.* **2011,** *134* (18), 184102.
11. Takimoto, Y.; Vila, F. D.; Rehr, J. J., Real-time time-dependent density functional theory approach for frequency-dependent nonlinear optical response in photonic molecules. *J. Chem. Phys.* **2007,** *127* (15), 154114.
12. Lopata, K.; Govind, N., Modeling Fast Electron Dynamics with Real-Time Time-Dependent Density Functional Theory: Application to Small Molecules and Chromophores. *J. Chem. Theor. Comput.* **2011,** *7* (5), 1344.
13. Schleife, A.; Draeger, E. W.; Kanai, Y.; Correa, A. A., Plane-wave pseudopotential implementation of explicit integrators for time-dependent Kohn-Sham equations in large-scale simulations. *J. Chem. Phys.* **2012,** *137* (22), 22A546.
14. Tavernelli, I.; Röhrig, U. F.; Rothlisberger, U., Molecular dynamics in electronically excited states using time-dependent density functional theory. *Mol. Phys.* **2005,** *103* (6-8), 963.
15. Nguyen, T. S.; Parkhill, J., Nonadiabatic Dynamics for Electrons at Second-Order: Real-Time TDDFT and OSCF2. *J. Chem. Theor. Comput.* **2015,** *11* (7), 2918.
16. (a) Nguyen, T. S.; Parkhill, J., Nonadiabatic Dynamics for Electrons at Second-Order: Real-Time TDDFT and OSCF2. *J. Chem. Theor. Comput.* **2015,** *11* (7), 2918-2924; (b) Lopata, K.; Govind, N., Modeling Fast Electron Dynamics with Real-Time Time-Dependent Density Functional Theory: Application to Small Molecules and Chromophores. *J. Chem. Theor. Comput.* **2011,** *7* (5), 1344.
17. Lopata, K.; Van Kuiken, B. E.; Khalil, M.; Govind, N., Linear-Response and Real-Time Time-Dependent Density Functional Theory Studies of Core-Level Near-Edge X-Ray Absorption. *J. Chem. Theor. Comput.* **2012,** *8* (9), 3284.
18. (a) Mitrić, R.; Werner, U.; Bonačić-Koutecký, V., Nonadiabatic dynamics and simulation of time resolved photoelectron spectra within time-dependent density functional theory: Ultrafast photoswitching in benzylideneaniline. *J. Chem. Phys.* **2008,** *129* (16), 164118; (b) Vincendon, M.; Dinh, P. M.; Romaniello, P.; Reinhard, P.-G.; Suraud, É., Photoelectron spectra from full time dependent self-interaction correction. *Eur. Phys. J. D* **2013,** *67* (5), 97.
19. (a) Wang, F.; Schmidhammer, U.; de la Lande, A.; Mostafavi, M., Ultra-fast Charge Migration Competes with Proton Transfer in the Early Chemistry of $H_2O^{\circ+}$. *Phys. Chem. Chem. Phys.* **2017,** *19* (4), 2894-2899; (b) Bruner, A.; Hernandez, S.; Mauger, F.; Abanador, P. M.; LaMaster, D. J.; Gaarde, M. B.; Schafer, K. J.; Lopata, K., Attosecond Charge Migration with TDDFT: Accurate Dynamics from a Well-Defined Initial State. *J. Phys. Chem. Lett.* **2017,** *8* (17), 3991.
20. Cheng, C.-L.; Evans, J. S.; Van Voorhis, T., Simulating molecular conductance using real-time density functional theory. *Phys. Rev. B* **2006,** *74* (15), 155112.

21. (a) Vahtras, O.; Almlöf, J.; Feyereisen, M. W., Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993,** *213* (5), 514-518; (b) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R., Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995,** *240* (4), 283.
22. Dunlap, B. I.; Rösch, N.; Trickey, S. B., Variational fitting methods for electronic structure calculations. *Mol. Phys.* **2010,** *108* (21-23), 3167.
23. Köster, A. M.; Geudtner, G.; Alvarez-Ibarra, A.; Calaminici, P.; Casida, M. E.; Carmona-Espindola, J.; Dominguez, V.; Flores-Moreno, R.; Gamboa, G. U.; Goursot, A.; Heine, T.; Ipatov, A.; de la Lande, A.; Janetzko, F.; del Campo, J.-M.; Mejia-Rodriguez, D.; Reveles, J.; Vasquez-Perez, J.; Vela, A.; Zuniga-Gutierrez , B.; Salahub, D. R. *deMon2k Version 5*, Mexico City, **2016**.
24. Runge, E.; Gross, E. K. U., Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984,** *52* (12), 997.
25. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965,** *140* (4A), 1133.
26. Bauernschmitt, R.; Ahlrichs, R., Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem. Phys. Lett.* **1996,** *256* (4), 454.
27. Thiele, M.; Gross, E. K. U.; Kümmel, S., Adiabatic Approximation in Nonperturbative Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2008,** *100* (15), 153004.
28. (a) Ullrich, C. A.; Burke, K., Excitation energies from time-dependent density-functional theory beyond the adiabatic approximation. *J. Chem. Phys.* **2004,** *121* (1), 28; (b) Suzuki, Y.; Lacombe, L.; Watanabe, K.; Maitra, N. T., Exact Time-Dependent Exchange-Correlation Potential in Electron Scattering Processes. *Phys. Rev. Lett.* **2017,** *119* (26), 263401; (c) Elliott, P.; Fuks, J. I.; Rubio, A.; Maitra, N. T., Universal Dynamical Steps in the Exact Time-Dependent Exchange-Correlation Potential. *Phys. Rev. Lett.* **2012,** *109* (26), 266404.
29. Szabó, A., & Ostlund, N. S. (1996). *Modern quantum chemistry : introduction to advanced electronic structure theory.* Mineola (N.Y.): Dover publications. **1996**
30. Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A., NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010,** *181* (9), 1477.
31. 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.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F. o.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Ayengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, **2009**.
32. Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill,

D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015,** *113* (2), 184.

33. Andrade, X.; Strubbe, D.; De Giovannini, U.; Larsen, A. H.; Oliveira, M. J. T.; Alberdi-Rodriguez, J.; Varas, A.; Theophilou, I.; Helbig, N.; Verstraete, M. J.; Stella, L.; Nogueira, F.; Aspuru-Guzik, A.; Castro, A.; Marques, M. A. L.; Rubio, A., Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. *Phys. Chem. Chem. Phys.* **2015,** *17* (47), 31371.

34. *TELEMAN - Beijing(Pékin)-Wuxi-Toulouse-Erlangen-Le Mans collaboration on Time-Dependent Electronic Dynamics in Molecules And Nanosystems*, **2011**.

35. Berendsen, H. J. C., *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. Cambridge University Press: Cambridge, **2007**.

36. (a) Castro, A.; Marques, M. A. L.; Rubio, A., Propagators for the time-dependent Kohn–Sham equations. *J. Chem. Phys.* **2004,** *121* (8), 3425-3433; (b) Bandrauk, A. D.; Lu, H., Exponential propagators (integrator) for the time-dependent Schrödinger equation. *J. Theor. Comput. Chem.* **2013,** *12* (06), 1340001.

37. Kidd, D.; Covington, C.; Varga, K., Exponential integrators in time-dependent density-functional calculations. *Phys. Rev. E.* **2017,** *96* (6), 063307.

38. Russakoff, A.; Li, Y.; He, S.; Varga, K., Accuracy and computational efficiency of real-time subspace propagation schemes for the time-dependent density functional theory. *J. Chem. Phys.* **2016,** *144* (20), 204125.

39. Gómez Pueyo, A.; Marques, M. A. L.; Rubio, A.; Castro, A., Propagators for the time-dependent Kohn-Sham equations: multistep, Runge-Kutta, exponential Runge-Kutta, and commutator free Magnus methods. *J. Chem. Theor. Comput.* **2018**, 14 (6), 3040.

40. Magnus, W., On the exponential solution of differential equations for a linear operator. *Commun. Pure. App. Math.* **1954,** *7* (4), 649.

41. Köster, A. M.; Campo, J. M. d.; Janetzko, F.; Zuniga-Gutierrez, B., A MinMax self-consistent-field approach for auxiliary density functional theory. *J. Chem. Phys.* **2009,** *130* (11), 114106.

42. Hochbruck, M.; Ostermann, A., Exponential integrators. *Acta. Numerica.* **2010,** *19*, 209.

43. (a) Williams-Young, D.; Goings, J. J.; Li, X., Accelerating Real-Time Time-Dependent Density Functional Theory with a Nonrecursive Chebyshev Expansion of the Quantum Propagator. *J. Chem. Theor. Comput.* **2016,** *12* (11), 5333; (b) Wang, F.; Yam, C. Y.; Chen, G.; Fan, K., Density matrix based time-dependent density functional theory and the solution of its linear response in real time domain. *J. Chem. Phys.* **2007,** *126* (13), 134104.

44. Gilmore, R., Baker‐Campbell‐Hausdorff formulas. *J. Math. Phys.* **1974,** *15* (12), 2090-2092.

45. Parise, A.; Alvarez-Ibarra, A.; Wu, X.; Zhao, X.; Pilmé, J.; Lande, A. d. l., Quantum Chemical Topology of the Electron Localization Function in the Field of Attosecond Electron Dynamics. *J. Phys. Chem. Lett.* **2018**, 844.

46. (a) Covington, C.; Hartig, K.; Russakoff, A.; Kulpins, R.; Varga, K., Time-dependent density-functional-theory investigation of the collisions of protons and α particles with uracil and adenine. *Phys. Rev. A.* **2017,** *95* (5), 052701; (b) Nagano, R.; Yabana, K.; Tazawa, T.; Abe, Y., Application of the time-dependent local density approximation to collision between a highly charged ion and an atom. *J. Phys. B: At., Mol. Opt. Phys.* **1999,** *32* (4), L65; (c) Hong, X.; Wang, F.; Wu, Y.; Gou, B.; Wang, J., H^+-H_2O collisions studied by time-dependent density-functional theory combined with the molecular dynamics method. *Phys. Rev. A.* **2016,** *93* (6), 062706.

47. Feynman, R. P., Leighton, R. B., & Sands, M. L. *The Feynman lectures on physics*. Reading, Mass: Addison-Wesley Pub. Co. **1963**
48. Krause, P.; Sonk, J. A.; Schlegel, H. B., Strong field ionization rates simulated with time-dependent configuration interaction and an absorbing potential. *J. Chem. Phys.* **2014,** *140* (17), 174113.
49. Lopata, K.; Govind, N., Near and Above Ionization Electronic Excitations with Non-Hermitian Real-Time Time-Dependent Density Functional Theory. *J. Chem. Theor. Comput.* **2013,** *9* (11), 4939.
50. Krause, P.; Schlegel, H. B., Strong-field ionization rates of linear polyenes simulated with time-dependent configuration interaction with an absorbing potential. *J. Chem. Phys.* **2014,** *141* (17), 174104.
51. (a) Calaminici, P.; Alvarez-Ibarra, A.; Cruz-Olvera, D.; Domínguez-Soria, V.-D.; Flores-Moreno, R.; Gamboa, G. U.; Geudtner, G.; Goursot, A.; Mejía-Rodríguez, D.; Salahub, D. R.; Zuniga-Gutierrez, B.; Köster, A., Auxiliary Density Functional Theory: From Molecules to Nanostructures. In *Handbook of Computational Chemistry*, Leszczynski, J., Ed. Springer Netherlands: Dordrecht, **2016**; pp 1-67; (b) Alvarez-Ibarra, A.; Calaminici, P.; Goursot, A.; Gómez-Castro, C. Z.; Grande-Aztatzi, R.; Mineva, T.; Salahub, D. R.; Vásquez-Pérez, J. M.; Vela, A.; Zuniga-Gutierrez, B.; Köster, A. M., Chapter 7 - First Principles Computational Biochemistry with deMon2k A2 - Ul-Haq, Zaheer. In *Frontiers in Computational Chemistry*, Madura, J. D., Ed. Bentham Science Publishers: **2015**; pp 281.
52. Mintmire, J. W.; Dunlap, B. I., Fitting the Coulomb potential variationally in linear-combination-of-atomic-orbitals density-functional calculations. *Phys. Rev. A* **1982,** *25* (1), 88.
53. Köster, A. M., Hermite Gaussian auxiliary functions for the variational fitting of the Coulomb potential in density functional methods. *J. Chem. Phys.* **2003,** *118* (22), 9943.
54. Domínguez-Soria, V. D.; Geudtner, G.; Morales, J. L.; Calaminici, P.; Köster, A. M., Robust and efficient density fitting. *J. Chem. Phys.* **2009,** *131* (12), 124102.
55. Köster, A. M.; Reveles, J. U.; del Campo, J. M., Calculation of exchange-correlation potentials with auxiliary function densities. *J. Chem. Phys.* **2004,** *121* (8), 3417.
56. Carmona-Espíndola, J.; Köster, A. M., Photoabsorption spectra from time-dependent auxiliary density functional theory. *Can. J. Chem.* **2013,** *91* (9), 795.
57. Wu, X.; Teuler, J.-M.; Cailliez, F.; Clavaguéra, C.; Salahub, D. R.; de la Lande, A., Simulating Electron Dynamics in Polarizable Environments. *J. Chem. Theor. Comput.* **2017,** *13* (9), 3985.
58. Sun, J.; Song, J.; Zhao, Y.; Liang, W.-Z., Real-time propagation of the reduced one-electron density matrix in atom-centered Gaussian orbitals: Application to absorption spectra of silicon clusters. *J. Chem. Phys.* **2007,** *127* (23), 234107.
59. Hirshfeld, F. L., Bonded-atom fragments for describing molecular charge densities. *Theoretica. chimica. acta.* **1977,** *44* (2), 129.
60. Becke, A. D., A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phy.* **1988,** *88* (4), 2547.
61. Célia, F. G.; Jan-Willem, H.; Jan, B. E.; Matthias, B. F., Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *J. Comput Chem.* **2004,** *25* (2), 189.
62. de la Lande, A.; Clavaguéra, C.; Köster, A., On the accuracy of population analyses based on fitted densities. *J. Mol. Model.* **2017,** *23* (4), 99.
63. Cederbaum, L. S.; Zobeley, J., Ultrafast Charge Migration by Electron Correlation. *Chem. Phys. Lett.* **1999,** *307* (3–4), 205.
64. Bader, R. F. W.; Preston, H. J. T., The kinetic energy of molecular charge distributions and molecular stability. *Int. J. Quantum Chem.* **1969,** *3* (3), 327.
65. Burnus, T.; Marques, M. A. L.; Gross, E. K. U., Time-dependent electron localization function. *Phys. Rev. A* **2005,** *71* (1), 010501.
66. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33.
67. St-Amant, A.; Salahub, D. R., New algorithm for the optimization of geometries in local density functional theory. *Chem. Phys. Lett.* **1990,** *169* (5), 387.

68. Alvarez-Ibarra, A.; Köster, A. M., A new mixed self-consistent field procedure. *Mol. Phys.* **2015,** *113* (19-20), 3128.
69. Gerald, G.; Florian, Köster, A. M.; Alberto, V.; Patrizia, C., Parallelization of the deMon2k code. *J. Comput. Chem.* **2006,** *27* (4), 483.

CHAPTER 4

Electron Dynamics in Contact with Polarizable Environments

Electron Dynamics in Contact with Polarizable Environments

In Chapter 3 we introduced a Real-Time Time-Dependent Density Functional Theory (RT-TDDFT) methodology to simulate electron dynamics in isolated molecular systems. Thanks to the involvement of auxiliary fitted density, the implementation allows the treatment of with large molecular systems comprised of tens of atoms, and possibly several hundreds of atoms once linear algebra operations are optimized. Although encouraging this computational set-up will face hurdles when attempts will be made to simulate molecular systems comprised of thousands or hundreds of thousands of atoms such as those encountered in biology (*e.g*. proteins, DNA, lipid layers…). One easily foresees the incredible complexity of simulating the electronic response of very large molecular systems subjected to a widely spread perturbation. At the present time plain RT-TDDFT simulations on nanometer scales seems intractable. On the other hand one may be interested in simulating electron dynamics explicitly only within a restricted area of space. The space beyond that specific region would be considered as its environment. This point-of-view recalls the strategy of continuum polarizable models[1] or of hybrid QM/MM (Quantum Mechanics/Molecular Mechanics) approaches[2]. In fact certain classes of perturbations can be localized, as for example the initial collision of a molecular fragment by a fast moving particle. It seems natural to seek for coupled schemes between RT-TDDFT and approximate representations of environments. The simplest option is to embed the region-of-special-interest (RSI) by a static environment[3]. In hybrid QM/MM, for example this can be achieved by adding the electrostatic potential created by the atomic charges of the MM atoms into the quantum Hamiltonian. This solution doesn't seem sufficient though because of the intuitive need to account for the subtle interplay between the electron dynamics which is treated explicitly within the region of special interest, and the electron density in the environment.

If the environment can be regarded as homogeneous the combination of RT-TDDFT with implicit polarizable continuum models (PCM) is a possible option. This kind of implementation has been recently reported by a few groups[4]. RT-TDDFT/PCM ED simulations lose atomic resolution of the environment. For highly inhomogeneous environment or/and if the coupling between the RSI and it environment is subtle, polarizable hybrid QM/MM approaches are good alternative candidates. To the best of our knowledge the only hybrid RT-TDDFT/MMpol implementation reported at the beginning of this work was due to Dinh *et*

al..[5†] These authors combined a grid based implementation of RT-TDDFT to a molecular mechanics force field to describe the MgO (001) solid surface. They were interested in the optical response properties of sodium clusters adsorbed on magnesium oxides. Electronic induction was introduced by distinguishing core from valence electrons on the oxygen atoms (O^{2-}). The average position of the core electrons + nucleus and the average position of valence electrons had the possibility to be different depending on the electrical environment, thereby creating induced dipoles. Although interesting, a generalization of this approach to polarizable FF for biomolecular simulation doesn't seem straightforward without complete parametrization of a dedicated FF. Methodologies adapted to simulate ED in biomolecules need to be devised. This is the objective of the work reported in this Chapter.

In the first section we introduce the methodology we chose to couple RT-TDDFT to polarizable force field. A particular focus will be put on the choice of induction model for the types of physical processes of interest with RT-TDDFT, as well on the introduction of retardation effects. With this new methodology in hand we analyze in section II the dynamics of the response of the environment of a peptide subjected to a strong perturbation. It will be shown how the introduction of a polarizable force field permits to introduced a new kind of dissipation mechanism into the algorithm.

I. METHODOLOGY

I.1 The charge point dipole mode for attosecond electron dynamics?

I.1.a. Stationary vs. dynamical description of induction

As seen in Part I there are various ways to include electronic induction in polarizable FF, hence in QM/MMpol methodologies, namely the fluctuating charge model[7], the Drude particle model[8] and the charge point dipole model[2]. These approaches have been validated for stationary calculations although coupled to molecular dynamics simulations. The stationary solutions are achieved by iterative procedures or approximated by propagating a fictitious dynamics of the "induction degrees of freedoms"[9]. In this Chapter we are interested in different dynamical regimes, namely the sub-femtosecond time domain. We wish to simulate the as-fs dynamical responses of the QM region's electron cloud after application of a

[†] In the course of the PhD, another implementation of a RT-TDDFT/MMpol scheme sharing important similarities with our approach appeared in the literature[6]. Donati, G.; Wildman, A.; Caprasecca, S.; Lingerfelt, D. B.; Lipparini, F.; Mennucci, B.; Li, X., Coupling Real-Time Time-Dependent Density Functional Theory with Polarizable Force Field. *The journal of physical chemistry letters* **2017,** *8* (21), 5283-5289. (a few weeks after publication of our own work).

perturbation. The choice of an induction model in the context of RT-TDDFT/MMpol simulations has to be done carefully. Can the well-tested QM/MMpol approaches developed for electronic stationary states be safely adapted in the context of as ED? A first obvious difference is that we don't look for stationary states of the electron density and the polarization state but we wish to describe their coupled dynamics. Second, for the typical applications we envision, the interaction of matter with strong laser fields or collisions with charged particles, it is not guaranteed not to push the standard induction models beyond their limits of validity. For example the strength of the electric fields applying on MM atoms may become large or may fluctuate rapidly on the as time scale. Finally it should be recalled that light travels at finite speed (approximately 3Å/as in vacuum) and that the question of retardation in the propagation of electric fields, which underlies all induction models, has to be posed in the context of as RT-TDDFT/MMpol simulations. In this subsection we justify our model for induction that we plan to couple to RT-TDDFT.

I.1.b. Expressions of atomic induced dipoles

We decided to rely on the description of induction by means of point dipole moments on atoms. The general expression for the induced point dipole ($\boldsymbol{\mu}$) on an atom subjected to electric field $\boldsymbol{F}$ can be expressed as a series expansion:

$$\boldsymbol{\mu}_i(F) = \sum_j \alpha_{ij}(\omega)\, F_j + \frac{1}{2}\sum_{j,k} \beta_{ijk}(\omega) F_j F_k + \frac{1}{6}\sum_{j,k,l} \gamma_{ijkl}(\omega) F_j F_k F_l + \cdots \qquad (1)$$

where $i, j, k, l = \{x, y, z\}$ are the Cartesian components of vectors or tensors. α_{ij}, β_{ijk} and γ_{ijkl} are the components of the polarizability, first hyperpolarizability and second hyperpolarizability tensors respectively. These tensors depend in principle on the frequency of the electric field applied on the MM atom K. In the context of RT-TDDFT simulations, if the electron density on the RSI generates a fast fluctuating electric field on the as time scale, this could be accounted for by the dynamical character of the polarizability. On the other hand the first and second hyperpolarizabilities go beyond the linear regime in case the fields generated by the RSI would be very strong. Building a polarizable FF based on Eq. 1 can be cumbersome and it might not be necessary to reach such a high level of sophistication. To decide which simplifications can be made we therefore need to know within which ranges of strength and of frequency the electric fields generated by the RSI are susceptible to fall in actual ED (Electron Dynamics) simulations.

I.1.c. Field strength and frequency in typical ED simulations

To investigate this point we simulate the dynamics of the electron cloud of a coumarin molecule in the gas phase after application of a perturbation. During the ED simulation we probe the value of the electric field

created by the electrons and by the nuclei at various distances. Four types of perturbation are tested. The first type of perturbation is the application of an electric field kick of strength 0.01 a.u (5 V/nm), which is a high, although transient value. The field has been applied perpendicular to the aromatic plane. For the second perturbation a Gaussian electric pulse of strength 0.001 a.u. (500mV/nm) is applied along the z-axis direction (in the molecular plane). The full-width at half-maximum of the envelop pulse is 1.65 fs and it is centered at 5 fs. The envelop is multiplied by a cosine function the frequency of which corresponds to an energy of 0.1528 Ha (close to the first excitation energy of coumarine). The two aforementioned perturbations are expected to bring the electron density in low lying excited states. The next two perturbations are designed to cause ionization of coumarin. They are consequently much stronger.

The third type of perturbation named "linear ramp" corresponds to application of a strong ionizing laser field aligned along the x-axis (perpendicular to the aromatic plane). The applied field increases linearly from 0 to 5 fs where it reaches a constant value of 0.1 u.a. To deal with this kind of process a complex absorbing potential has been placed 10 Å away from the coumarin to absorb emitted electrons (see Chapter 3 for details)[10]. In addition very diffuse atomic basis sets are used on H atoms to expand the MOs of electrons localized a few Angstroms away from the nuclei[10a]. The last type of perturbation named "radiation" is a collision with a 0.1 MeV proton travelling along the x-axis. Such a collision puts the electron cloud in high lying excited states with a fraction of electron becoming unbound[11]. This results in the emission of so-called secondary electrons that are absorbed by the CAP. We probe the electric field generated by the electron cloud along each ED simulation at several points. The points have been chosen taking the geometry of the microsolvated coumarin investigated in Chapter 3 (Figure 7, aggregate S5). The electric field is calculated at the positions of water molecule oxygen atoms[‡]. We have defined four ranges of distances from the coumarine: 0-3 Å; 3-4 Å, 4-6 Å and 6-7 Å. We report in Figure 1 the evolution of the maximum and minimum field strength for each layer and each type of perturbation.

[‡] Note that no water molecules are actually included in the RT-TDDFT simulations. We simply probe the electric field at the positions where these molecules could be positioned in a polarizable RT-TDDFT/MMpol simulation.

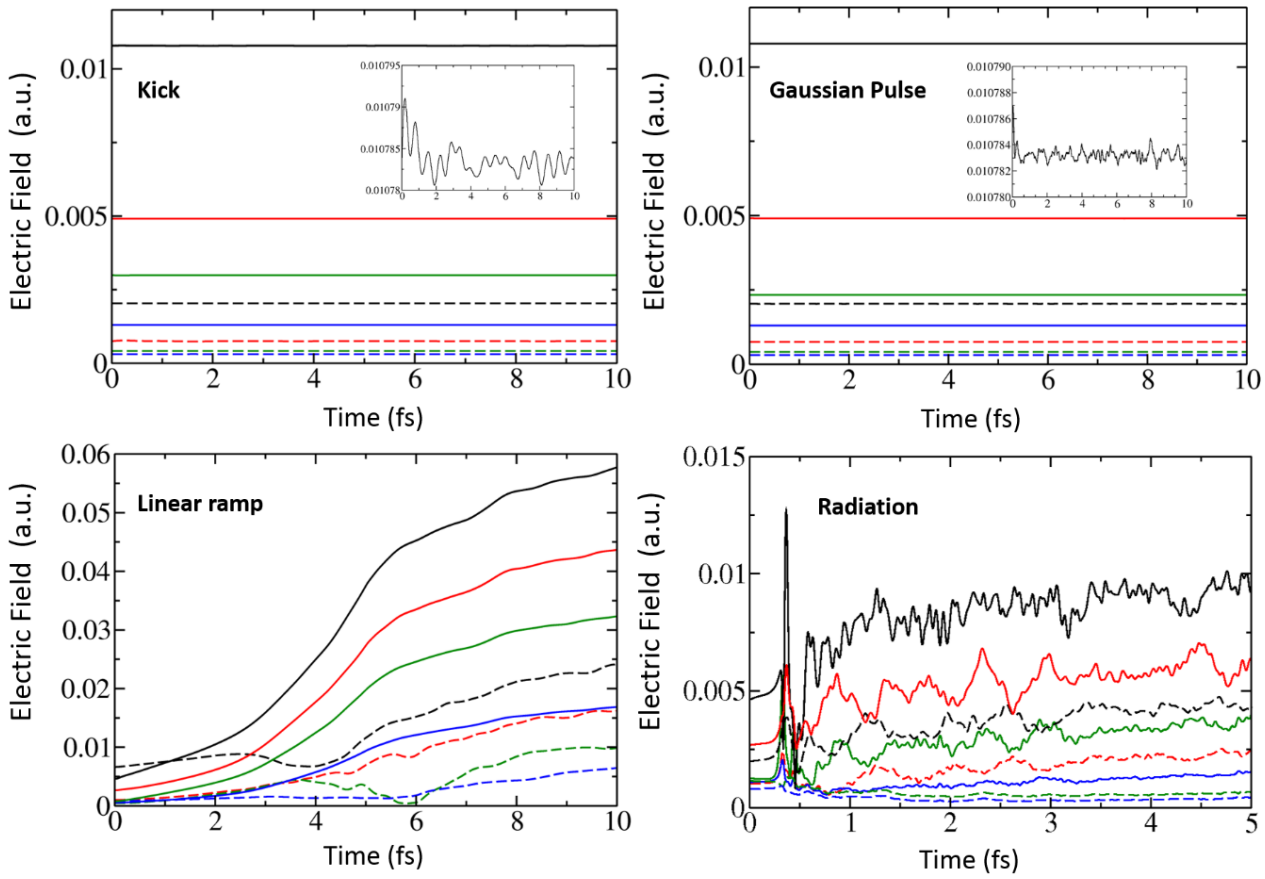


Figure 1. Variation of electric field on oxygen of water molecule in different shells after four different kind of perturbation on coumarin. The electric fields encompass both the electron and nucleus contributions. Color code: 0-3 Å in black,3-4 Å in red, 4-6 Å in green, 6-7 Å in blue. The plain and dashed curves correspond to the highest and the lowest electric fields for each layer. PBE/DZVP-GGA/GEN-A2* ED simulations relying on second order Magnus propagator with a time step of 2 as. For the ionizing radiations (bottom) a complex absorbing potential was placed 10 Å away from the coumarin to absorb emitted electrons. Very diffuse atomic basis sets were used on H atoms to expand the MOs of unbound electrons. Almost 3 e- were emitted in the strong laser field ionization and 0.3 e- upon proton collision (at 10 fs).

Roughly, the longer the distance from coumarin the weaker the electric field. This is an expected trend. For the first two kinds of perturbations (upper panel) the electric field generated by the coumarin is weak (<0.005 u.a.) except for the closest probing points (0.01 u.a.). For these two simulations the initial perturbation is not strong essentially putting the electron cloud in low lying excited states. The amplitude of the fluctuations of the fields generated by the electron density are very small (already <0.00001 u.a. for the *closest* probing point). We are clearly in a weakly perturbation regime. If we were conducting RT-TDDFT/MMpol simulations, the induced dipoles on the oxygen atoms would be proportional to the electric field ($\mu_i(F) \cong \sum_j \alpha_{ij}(\omega)\, F_j$). The introduction of first and second hyperpolarizability would not be necessary.

We also see that the fluctuations of the electric field at the closest probing point are slow compared to the ED taking place on the as time scale. For an eventual MM atom positioned near the coumarin the variation of electric field would be uncoupled from its own polarization. Therefore we could safely drop out the frequency dependence of the polarizability tensor and consider static polarizabilities: $\mu_i(F) \cong \sum_j \alpha_{ij} F_j$.

In the case of strong field ionization (bottom-left) the electric field generated by the electron cloud is much more important (> 0.01 a.u. even at 7 Å). This is because the electron density is largely displaced by the applied field. In this simulation almost 3e- have been absorbed by the complex absorbing potential. The question arises: *whether such high fields would put the description of electronics beyond the linear regime?* Figure 2 brings a negative answer to this question. It shows the DFT based induced dipole (z-component) produced water, cysteine and phenol upon application of a constant external field. Even for highly polarizable molecules such as a cysteine (an amino acid encompassing a thiol function) or a phenol, we find a linear relationship between the applied electric field and the induced dipole moment on the [0, 0.05] range, in atomic units. We do not see electric field fluctuations on the sub-femtosecond time scale for this type of strong perturbation.

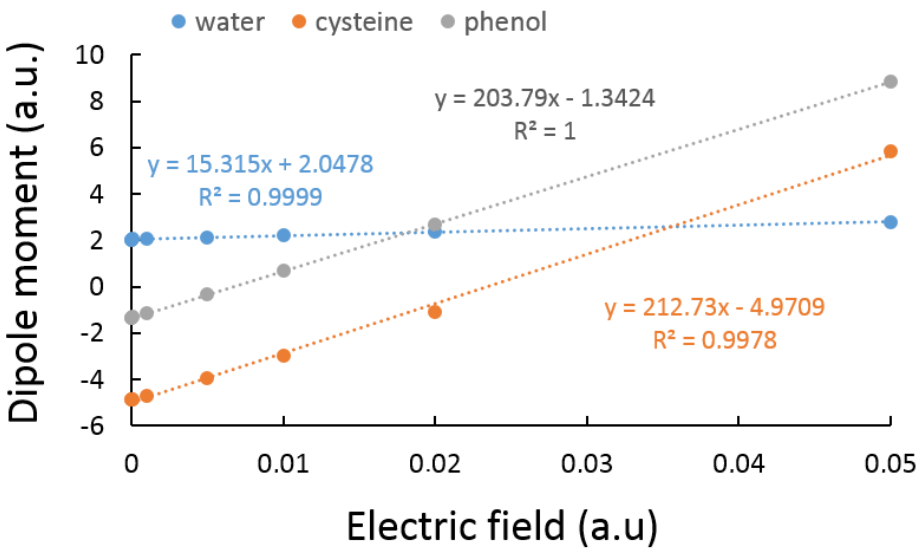


Figure 2. z-component of the induced dipole moment computed at the DFT level as a function of applied electric field (along the Oz axis). The dashed lines correspond to linear regression fit of the DFT points.

Finally, considering the case of collision of coumarin by fast protons, we find an intermediate situation. The electric field increases moderately around the coumarin, making the linear response probably valid.

On the other hand the amplitudes of the fluctuations are rather pronounced (e.g. 0.005 considering distances from the molecule of around 4Å). Yet these fluctuations are not fast enough compared to the attosecond dynamics of the electrons. The inclusion of dynamical polarizabilities is therefore not needed.

In summary the ED simulation reported in Figure 1 tends to indicate that for typical applications of RT-TDDFT, a model of induction in which the induced dipoles on MM atoms is proportional to the electric field, neglecting hyperpolarizabilities and dynamic polarizabilities, is perfectly adapted. Instead of Eq. 1 an adequate FF to be coupled to RT-TDDFT could be based on $\mu_i(F) \cong \sum_j \alpha_{ij} F_j$. We can make a further simplification by assuming that the atom polarizabilities are isotropic $\boldsymbol{\mu}_K = \alpha_K \boldsymbol{F}_K$. This set of hypotheses defines a model of induction for the MMpol force field. From this, supplementary approximations can be made. The polarizability could be made dependent on the type of atom[7], for example an oxygen atom within an water molecule may have different intrinsic polarizability than an oxygen atom within a peptide bond.

I.2 Electron dynamics in contact with a polarizable environment

I.2.a. Determination of point dipoles

Having defined the induction model we now describe the mathematical expressions that are necessary to proceed to implementation in deMon2k. From now the indices i, A, K respectively refer to electrons, atom nuclei from the QM region and from MM atoms. Each MM site K is characterized i) by a permanent charge (q_K) and a static polarizability α_K which is taken to be isotropic. The latter permits to determine an induced dipole $\boldsymbol{\mu}_K$ from the electric field felt by atom K ($\boldsymbol{F}_K$). $\boldsymbol{F}_K$ includes the electric field created by other MM atoms that arises both from other permanent multipoles $\boldsymbol{F}_K^{(0)}$ and from other induced dipoles $\boldsymbol{F}_K^{ind}$. For simplicity we will assume that only permanent charges are present in the FF (i.e. no permanent dipoles, quadrupoles, like in AMOEBA). $\boldsymbol{F}_K$ also includes the electric field created by the QM region, that is from the QM nuclei $\boldsymbol{F}_K^{ZQM}$ and from the electron density (ρ) $\boldsymbol{F}_K^{\rho}$. In principle we could also include the field created by an external perturbation but we will not consider this possibility here for simplicity.

$$\boldsymbol{\mu}_K = \alpha_K \boldsymbol{F}_K = \alpha_K \left(\boldsymbol{F}_K^{(0)} + \boldsymbol{F}_K^{ZQM} + \boldsymbol{F}_K^{\rho} + \boldsymbol{F}_K^{ind} \right) \quad (2)$$

The mathematical expressions to compute the electric fields are given below:

$$\boldsymbol{F}_K^{ZQM} = \sum_{A \in QM} \frac{Z_A}{{r_{KA}}^3} \boldsymbol{r}_{KA} \tag{3}$$

$$\boldsymbol{F}_K^{(0)} = \sum_{\substack{L \in MM \\ L \neq K}} \frac{q_L}{{r_{KL}}^3} \boldsymbol{r}_{KL} \tag{4}$$

$$\boldsymbol{F}_K^{ind} = -\sum_{\substack{L \in MM \\ L \neq K}} \boldsymbol{T}_{KL} \boldsymbol{\mu}_L(t) \tag{5}$$

$$\boldsymbol{T}_{KL} = \frac{1}{r_{KL}^3} \boldsymbol{I} - \frac{3}{r_{KL}^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \tag{6}$$

$$\boldsymbol{F}_K^{\rho} = -\int \frac{\rho(\boldsymbol{r}, t)}{{r_K}^3} \boldsymbol{r}_K \, d\boldsymbol{r} \tag{7}$$

$\boldsymbol{r}_{KL}$ is the vector between atoms K and L; $\boldsymbol{T}_{KL}$ is the dipole-dipole interaction tensor and $\boldsymbol{I}$ is the identity matrix. Z_A is the nuclear charge of QM nucleus A. In ED simulations carried out with RT-TDDFT the nuclei may either be fixed or displaced by Newton's laws (Ehrenfest molecular dynamics[12]). In this work we only consider the former possibility and carry out ED simulations at fixed nuclear positions. In that case $\boldsymbol{r}_{KA}$, $\boldsymbol{r}_{KL}$, hence $\boldsymbol{F}_K^{ZQM}$ and $\boldsymbol{F}_K^{(0)}$ are constant in our simulations. The other electric fields are time-dependent.

I.2.b. Energy expression of QM/MMpol with electrostatic embedding

The total QM/MMpol Hamiltonian reads:

$$E^{QM/MMpol}[\rho(\boldsymbol{r}, t), \boldsymbol{\mu}(t)] = E^{QM}[\rho(\boldsymbol{r}, t)] + E^{embed}[\rho(\boldsymbol{r}, t), \boldsymbol{\mu}(t)] + E^{MM}[\boldsymbol{\mu}(t)] \tag{8}$$

where E^{QM} collects the energy of the QM region. In the Kohn-Sham DFT framework we have $E^{QM} = T_s[\rho] + \int \rho(r) v_{ext} dr + J[\rho] + E_{xc}[\rho] + E_{class}^{QM}$ that is, respectively, the sum of kinetic energy of the reference electron gas, the interaction energy with the external potential, the classical Coulomb repulsion, the exchange-correlation energy and the classical repulsion among QM nuclei (E_{class}^{QM}). E^{embed} is the embedding energy.

$$E^{embed} = \sum_{K \in MM} \int \frac{\rho(\boldsymbol{r}, \boldsymbol{t})}{|\boldsymbol{r}_K - \boldsymbol{r}|} q_K d\boldsymbol{r} - \frac{1}{2} \sum_{K \in MM} \int \frac{\rho(\boldsymbol{r}, \boldsymbol{t})}{|\boldsymbol{r}_K - \boldsymbol{r}|^3} \boldsymbol{\mu}_K(t) \cdot (\boldsymbol{r}_K - \boldsymbol{r}) \, d\boldsymbol{r} + E_{class}^{embed} \tag{9}$$

$$E_{class}^{embed} = \sum_{K\in MM}\sum_{A\in QM}\frac{Z_A q_K}{|\boldsymbol{r}_K - \boldsymbol{r}_A|} - \sum_{A\in QM}\frac{1}{2}\boldsymbol{\mu}_K(t)\cdot\boldsymbol{F}_A^{ZQM} \quad (10)$$

For a polarizable FF E^{embed} encompasses the interaction with MM charges (q_K) and with the induced dipoles ($\boldsymbol{\mu}_K$) (respectively the first two terms on the r.h.s. of Eq. 9) E_{class}^{embed} is the classical energy of MM atoms with the QM nuclei. E^{embed} clearly appears as a coupling term between the QM and MM regions as it depends both on the electron density (QM region) and on the induced dipoles (MM region).

Finally E^{MM} is the MM energy for the MM region. The MM energy involves bonded terms and non-bonded terms. The classical interaction among MM charges and the induction energy within the MM region: $-\frac{1}{2}\sum_{K\in MM}\boldsymbol{\mu}_K(t)\cdot\boldsymbol{F}_K^{(0)}$.

For convenience we can also define an induction energy according to:

$$E^{ind} = \frac{1}{2}\sum_{K\in MM}\int\frac{\rho(\boldsymbol{r},\boldsymbol{t})}{|\boldsymbol{r}_K - \boldsymbol{r}|^3}\boldsymbol{\mu}_K(t)\cdot(\boldsymbol{r}_K - \boldsymbol{r})\,d\boldsymbol{r} - \frac{1}{2}\sum_{A\in MM}\boldsymbol{\mu}_K(t)\cdot\boldsymbol{F}_A^{ZQM} - \frac{1}{2}\sum_{K\in MM}\boldsymbol{\mu}_K(t)\cdot\boldsymbol{F}_K^{(0)} \quad (11)$$

The potential entering the KS potential is obtained by differentiation of the embedding energy with respect to the electron density: $\partial E^{embed}/\partial\rho$. Adopting the matrix notations introduced in Chapter 4 we have:

$$E^{embed} = \sum_{K\in MM}\sum_{\sigma,\tau}q_K P_{\sigma\tau}\left\langle\sigma\left|\frac{1}{|\boldsymbol{r}_K - \boldsymbol{r}|}\right|\tau\right\rangle - \frac{1}{2}\sum_{K\in MM}\sum_{\sigma,\tau}P_{\sigma\tau}\left\langle\sigma\left|\frac{\boldsymbol{\mu}_i(\boldsymbol{r}_K - \boldsymbol{r})}{|\boldsymbol{r}_K - \boldsymbol{r}|^3}\right|\tau\right\rangle + E_{class}^{embed} \quad (12)$$

$$\frac{\partial E^{embed}}{\partial P_{\sigma\tau}} = \sum_{K\in MM}q_K\left\langle\sigma\left|\frac{1}{|\boldsymbol{r}_K - \boldsymbol{r}|}\right|\tau\right\rangle - \frac{1}{2}\sum_{K\in MM}\left\langle\sigma\left|\frac{\boldsymbol{\mu}_i(\boldsymbol{r}_K - \boldsymbol{r})}{|\boldsymbol{r}_K - \boldsymbol{r}|^3}\right|\tau\right\rangle \quad (13)$$

$P_{\sigma\tau}$ are matrix elements of the Kohn-Sham density matrix. These two terms need to be added in the Kohn-Sham Hamiltonian to polarize the electron density by the MM region.

I.2.c. Coupling between RT-TDDFT MMpol?

i) Stationary/dynamical solution

To solve the time-independent KS equations to determine the stationary states of the system of interest, a common procedure is to relax the MM induced dipoles at every SCF cycle. The MM dipoles are then injected in the next SCF cycle to calculate a new embedding potential. The convergence threshold for

converging the MM dipole moments is tightened along with the SCF convergence to reach, at global convergence, a user-defined value, typically 10^{-8} to 10^{-10} D. On the other hand if one is interested in the time-dependent solutions of the KS equations more subtle algorithms are needed because of the time dependence of each terms of eq. 8. In principle one needs to set up the coupled equations of motion for the overall system. This is not a trivial task because of the composite quantum-classical nature of the system. One may think of coupling RT-TDDFT for the electron cloud to a fictitious dynamics of the MM induced dipoles, in the spirit of what is done for molecular dynamics simulations with MMpol[9]. In this PhD work we consider a simpler scheme in which we make the assumption that the MM dipoles completely relax at each RT-TDFT step. In other words, we look for the stationary polarization state of the environment along with the non-stationary propagation of the electron cloud. We call this a RT-TDDFT/MMpol-stationary scheme.

ii) Implementation in deMon2k

The coupling between RT-TDDFT and MMpol has been carried out based on the implementation of RT-TDDFT described in Chapter 3. The strategy we have followed has been to build on the pre-existing "in-deMon2k QM/MM" methodology[13]. Indeed deMon2k includes an internal QM/MM approach meaning that both MM and QM (DFT) calculations are carried out by the program without needing program interfaces. The in-deMon2k QM/MM is based on the TINKER[14] format. It is currently compatible with the OPLS[15], CHARMM[16] or Amber[17] force fields. Our strategy has been to upgrade this methodology to QM/MMpol, *i.e*. introduction electronic induction into the code, and the coupling of it to RT-TDDFT. The hope was that QM/MMpol calculation which needs several information passing between the DFT and MMpol calculations would benefit from fully local implementation.

An important feature of the QM/MMpol implementation was to use auxiliary fitted densities to evaluate the electric field created by the electron density ($\boldsymbol{F}_K^{\rho}$). This field is needed to evaluate the MM induced dipoles (Eq. 7). This can be a time consuming task. We showed the substitution of ρ by $\tilde{\rho}$ decreased dramatically the computational cost of evaluating $\boldsymbol{F}_K^{\rho}$ without affecting accuracy, as long as sufficiently flexible (i.e. including polarization functions) auxiliary basis sets are used, namely GEN-An*. These tests are described in details in the article reproduced at the end of Part II.

iii) Validation of the coupling scheme

To test the RT-TDDFT/ MMpol-stationary- scheme we consider a peptide (Tyr-Gly-Gly-Phe-Met) treated by DFT immersed in a box of 4,030 polarizable POL3[18] water molecules (Figure 3). The full system was previously equilibrated by classical MD simulations. We have used the PBE functional and the DZVP-GGA

atomic basis set in combination with the GEN-A2* auxiliary sets. The fitted density has been used to calculate both the Coulomb and XC potentials[19]. A grid of high accuracy is used to integrate the XC contributions (10^{-7} Ha). The induced dipoles have been updated at every SCF cycle by an iterative procedure until the Root-Mean-Square between two successive cycles is below 10^{-9} D. After SCF convergence the electronic density of the peptide is perturbed by a Gaussian shaped electric field centered at 10 as and with standard deviation 1 as. Three field strengths have been tested: 0.001, 0.01 and 0.1 a.u. The simulations have been run for 1 fs with the propagator-corrector Magnus and the Becker-Campbell-Haussdorff (see Chapter 3). 30 terms have been used in the expansion.

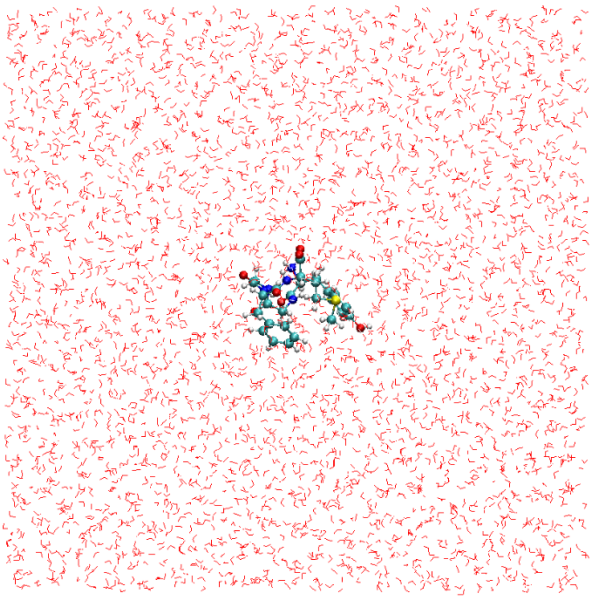


Figure 3. Metenkephalin solvated in a water box. The peptide is described by DFT while the aqueous environment is described by polarizable POL3 water model.

The objective is to evaluate the error due to the RT-TDDFT/ MMpol-stationary- scheme. For short enough time steps the decoupling approximation is certainly valid. Indeed we find that the MM induced dipoles evolve very smoothly with such short time scales. No more than one iteration is needed to converge the induced dipole. The 0.1 as ED simulation will thus serve as reference. We have repeated the simulations with longer time steps of 0.25, 0.5, 0.75 and 1 as. Figure 4 depicts the differences of RT-TDDFT/MMpol total energy (Eq. 8), of induction energy (Eq. 12), and of embedding energy (Eq. 9) as a function of time taking the 0.1 as time-step simulation as reference.

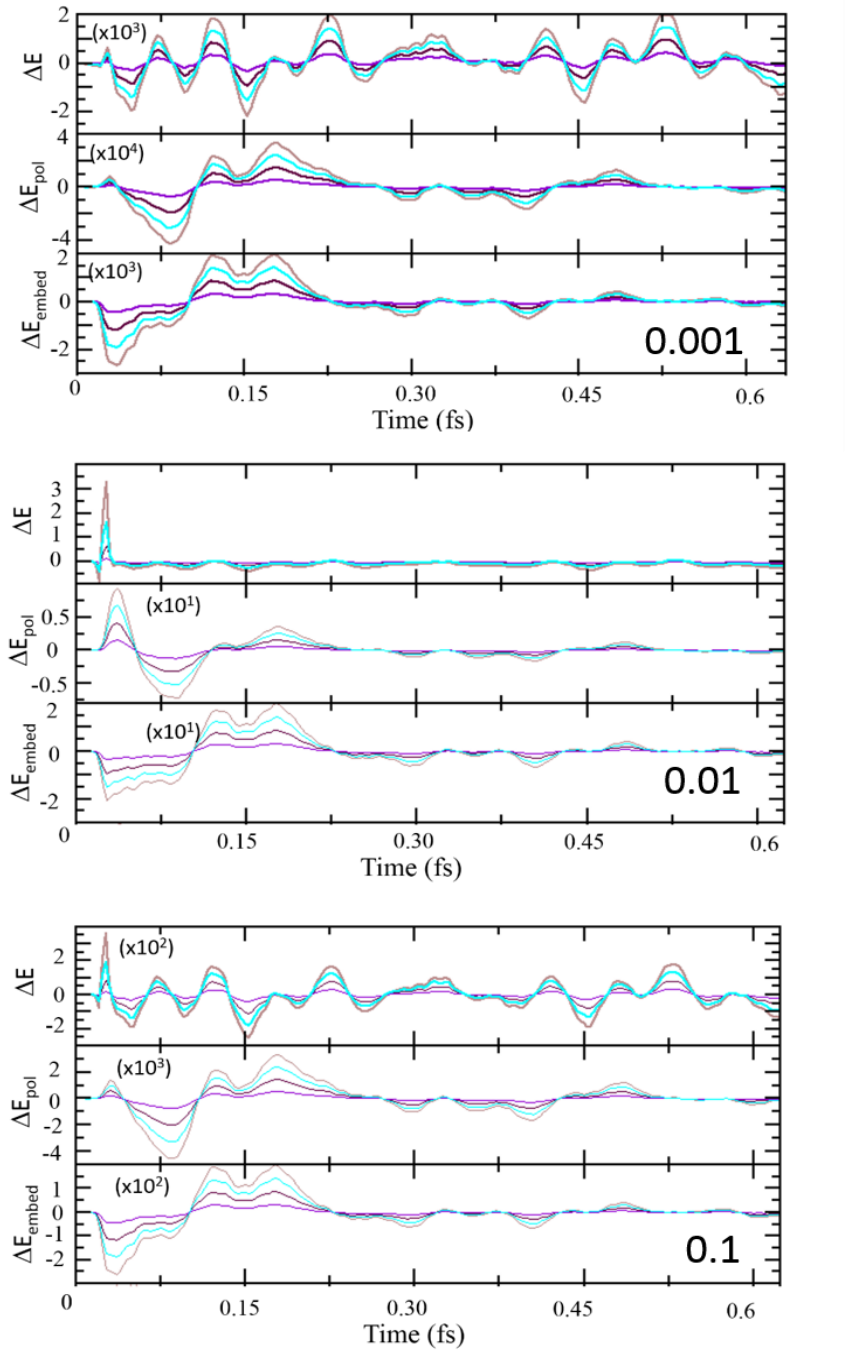


Figure 4. Error estimations of the MMpol-stationary-RT-TDDFT scheme. The 0.1as time-step simulation is taken as a reference and the data for larger time steps are given with respect to the reference. light brown: 1 as, cyan : 0.75 as, marron: 0.5 as, violet: 0.25 as. All the energies are given in kcal/mol scaled by coefficients given on each graph. ΔE is the total QM/MM energy, ΔE_{pol} is the total polarization energy and ΔE_{embed} is the embedding energy (permanent charges and induced dipole). The initial field strengths is set to 0.001 (Top), 0.01 (Middle) and 0.1 (Bottom) a.u.

We first consider simulations with a weak initial perturbing pulse (Figure 4, Top, 0.001 a.u.). Clearly, the larger the time step the larger the difference with the 0.1as ED simulation. For the total energy and for the embedding energy the maximum error is of the order of a few thousandths of a kcal/mol with a 1 as time step. It is an order of magnitude smaller for the polarization energy. These values are rather small compared to the variations of the total energy in these simulations caused by the initial perturbation with the external electric field (around 0.06 kcal/mol). Interestingly the energy errors fluctuate around zero. This suggests that the simulations with time steps larger than 1as eventually depart from the reference trajectory but do not diverge from it. We also find that the peptide dipole as well as the water dipoles of the first solvation layer (*i.e.* those mainly impacted by the electron dynamics taking place on the peptide) were within $1.0E^{-5}$D from those of the reference trajectory. This is a very small value. In simulations in which the initial perturbing electric field strength was increased to 0.01 a.u. (Figure 4, Middle), the same trends are obtained albeit with a factor of ten in the amplitudes of the errors. This again seems acceptable in view of the overall total energy change (6 kcal/mol). For an even stronger perturbing field (0.1 a.u., (Figure 4, Bottom) the errors in total, polarization energy and embedding energies are of the order of a kcal/mol, a tenth of a kcal/mol and a hundredth of a kcal/mol, respectively. These values are quite high, but again much smaller than the fluctuations of the total energy of the molecule (around 50 kcal/mol).

Altogether these tests justify the non-stationary/stationary coupling scheme between RT-TDDFT for the QM part and stationary MMpol for the environment although one should be careful to adapt the propagation time step to the amplitude of the electronic fluctuation that takes place in the QM region. The most suitable time step might depend on the particular system of interest.

In principle though there should be a time step beyond which the decoupling between electrons and MM dipoles ceases to be valid. When we increased the time step (2 or 5 as) the electronic propagation was not stable anymore and diverged in a few steps. RT-TDDFT propagations are usually very sensitive to discontinuities that may arise in the time-dependent KS potential. Hence, a plausible explanation for the numerical instabilities observed in RT-TDDFT/MMpol simulations for the largest time steps may stem to potential discontinuities caused by significant variations of MM induced dipoles between two propagation steps. Interestingly, sudden instability of electron dynamics propagation may thus well be a sign of the breakdown of the decoupling hypothesis between the electron cloud dynamics and induced MM dipoles.

I.3 On the introduction of retarded electric fields

I.3.a. Why consider retardation effects?

In hybrid QM/MMpol calculations relying on polarizable force fields the mutual interactions between the QM and MM regions are mediated by electric fields generated within each region. In a charge point dipole model for example, the electric field generated by the quantum nuclei and the electron cloud contributes to determine induced dipoles on MM atoms. Conversely the electric field generated by the fixed MM charges and by the MM induced dipoles are incorporated into the Kohn-Sham Hamiltonian (assuming one uses DFT). Electric fields propagate at the speed of light (c). In vacuum c is around $3.10^8\ m/s \approx 137\ a.u.$ (atomic units), that is around $3\ \text{Å}/as$. It is less in condensed matter as reflected by the refractive index (n). For molecular dynamics simulations relying on stationary DFT/MMpol potential energies, the typical time steps employed are of the order of 1 fs. One can safely assume infinite propagation of electric fields mediating QM/MMpol interactions.

The situation is different in the context of RT-TDDFT/MMpol simulations. The typical propagation time steps (Δt) are of the order of 1 as. As illustrated in Figure 5, an MM atom K situated for example 12 Å away from the RSI would feel fluctuations of electric field generated by the electron cloud with a delay of 4 as (4 propagation steps if Δt is set to 1 as). Similarly the polarization state of MM atoms would take some times to reach other MM atoms or the RSI. The consideration of delays in the propagation of the electric fields is a point that has to be addressed and which is specific to RT-TDDFT/MMpol type of methodology couplings.

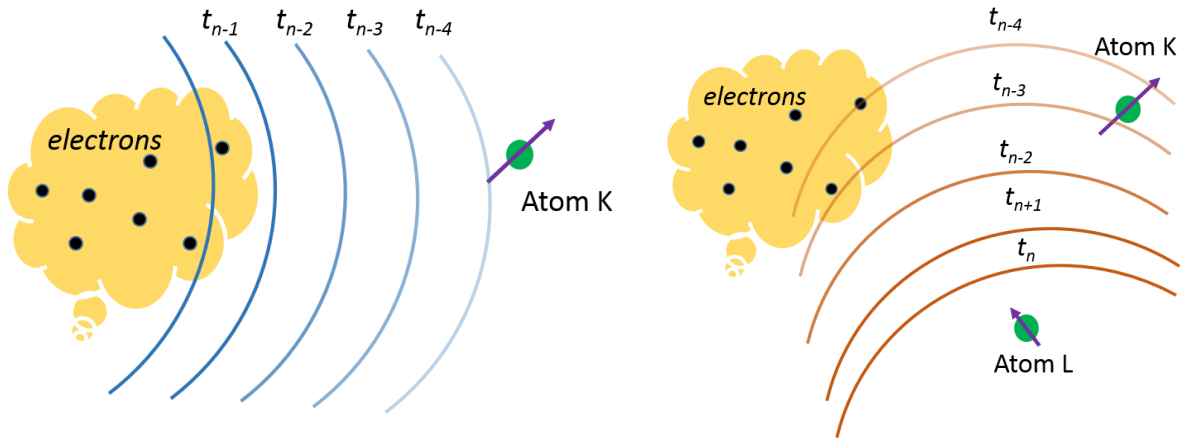


Figure 5. Qualitative illustration of the finite speed propagation of electric fields mediating QM/MM interactions on the attosecond domain. Left: the electric field created by the fluctuating electron cloud of the RSI (in yellow) takes four propagation time steps to reach MM atom K which holds a point dipole μ_K. The timings t_i refers to propagation times in the timeframe of ED simulation on the RSI. Right: the electric field created by the induced dipole on MM atom L takes a few propagation time steps to reach the RSI or MM atom K. The timings t_i refers to propagation times in the timeframe of atom L.

I.3.b. How to introduce retardation effects?

i) Retardation electric fields created by induced dipole

When building the KS potential at time $t_n + \Delta t$ during the iterative process one needs to account for the fact that the field created by other induced dipoles takes time to reach atom *K*. Let's first consider $\boldsymbol{F}_K^{ind}$ and imagine we can ride on atom *K*. The electric field created by other induced dipoles on MM atoms *L* should not be that arising from the dipoles μ_L at $t_n + \Delta t$ but, rather, at a time earlier in the past. The further MM atoms *L*, the longer the delay for the electric field from this atom to reach MM atom *K*. The delay is determined by the distance between atoms (d_{KL}) and by the speed of light in the medium of interest, *i.e.* $t_{delay} = d_{KL}/nc$ with n the refractive index of the medium and c the speed of light in vacuum. As a rule of thumb for $n = 1$, $c \approx 137\ a.u.$, approximately 3 Å/as. Introducing the delay is straightforward if one knows the history of induced dipoles. Being three-dimensional vectors they can be stored easily in RAM (Random Access Memory) or on machine hard disks. When $\boldsymbol{F}_K^{ind}$ on atom *K* is needed, a loop over all other MM atoms (*L*) is carried out. For each atom *L*, d_{KL} is calculated, which defines t_{delay}, hence $\mu_L(t_{n+1} - t_{delay})$ using the history of the μ_L induced dipole. Because, in general, t_{delay} doesn't correspond to an integer multiple of Δt, we use cubic splines to interpolate $\mu_L(t_{n+1} - t_{delay})$ [20]. The number of records, *M*, is determined from the speed of light and from a cutoff which is used to screen the field created by the induced dipoles. $M = C(R_{cut}nc/\Delta t)$ where $C(X)$ is the least integer greater than or equal to *X*. R_{cut} is usually set to around 50 Å.

ii) Retardation electric field created by the electric density

Similar considerations apply for the electric field produced by the electron cloud ($\boldsymbol{F}_K^{\rho}$). In such a case it is the history of the electron density that must be stored. Some difficulties arise though. First, the memory required for the storage of large and numerous density matrices is demanding and would alter computational performance. To circumvent this difficulty, we don't use the Kohn-Sham density to evaluate $\boldsymbol{F}_K^{\rho}$ but instead an auxiliary density function $\tilde{\rho}$. We recall that $\tilde{\rho}$ is expressed as a linear combination of auxiliary functions f: $\tilde{\rho}(\boldsymbol{r}, t - t_{delay}) = \sum_f x_f(\boldsymbol{r}, t - t_{delay}) f(\boldsymbol{r})$ (see Chapter 3). The time-dependent auxiliary coefficients x_f are determined from the KS density by a variational fitting procedure[21]. The number of auxiliary functions is usually three to four times that of atomic orbitals, but storing the history of the x_f coefficients which define $\tilde{\rho}$ is much less demanding than storing the history of density matrices. Furthermore, we showed that $\tilde{\rho}$ could be used safely in place of ρ for evaluating the electric field created by the QM region in stationary or time-dependent DFT/MMpol calculations (that replacing $\boldsymbol{F}_K^{\rho}$ by $\boldsymbol{F}_K^{\tilde{\rho}}$), or to evaluate atomic multipoles. A second difficulty is the practical definition of the distance to be used

between electrons and atom K. Indeed, electrons are delocalized and one faces the well-known problem of defining an atom in a molecule. To solve this problem, we take as distance d_{KA} that between MM atom K and the QM atom *A* holding the auxiliary function $f(\boldsymbol{r})$.

The above consideration introduce retardation effects in the determination of MM induced dipoles. We now turn to the Kohn-Sham potential. The same strategy as before is used to determine the effective $\boldsymbol{\mu}_K(t_{n+1} - t_{delay})$ using the history of induced dipoles and cubic spline interpolations[20].

I.3.c. Some tests with a simple model

As a first test we consider a dimer of water molecules at the QM/MMpol level. One is described by DFT and another by the polarizable POL3 model[18] (Figure 6). In the latter the oxygen and hydrogen atoms hold charges -0.730 and 0.365 and atomic polarizabilities of 0.528 and 0.170 Å^3 respectively. The ground state density is obtained in an SCF calculation. A Gaussian shaped electric field is applied during the ED simulation. It is centered at t_0+ 2as where t_0 is the initial time. The Gaussian pulse has a standard deviation of 0.1 as and a maximum strength of 0.9 e/bohr^2. It is oriented along the hydrogen bond direction between the two molecules.

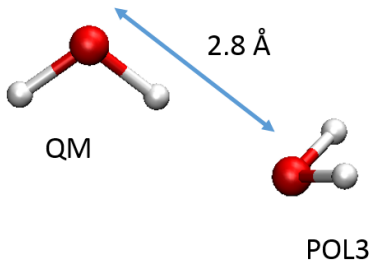


Figure 6. Water dimer used to validate the implementation of retardation in hybrid RT-TDDFT/MMpol simulations.

Figure 7 depicts the molecular dipole moment of the QM (Top) and MMpol (Bottom) water molecules for three ED simulations differing by the refractive indices defined to determine the retarded times, namely 10.0 (black, considered as equivalent to infinite speed of light), 1.0 (red) and 0.5 (green). When ignoring retardation (black curves), the dipole moment on the MM molecule instantaneously follows the perturbation on the QM region caused by the external electric pulse. Note however that while the dipole of the QM water varies by 0.07 D that of the MM water varies ten times less (0.007 D). This is a consequence of the rapidly decaying electric potential created by the electron density (eq. 22). The MMpol

molecule is slightly polarized. With n=1 and 0.5 the MM dipoles respond with delays of around 1.5 and 3 as respectively, as expected from the distance separating the molecules (ca. 1.8 Å for the H…O distance). Then the induction dynamics is clearly different for the three cases. The tight correlation between the QM and MM dipoles observed for n=10.0 is weakened. We also note that the back reaction of the MM dipole on the electron cloud is negligible as the dipole moment of the QM water follows the same trend for the simulation with different refractive indexes.

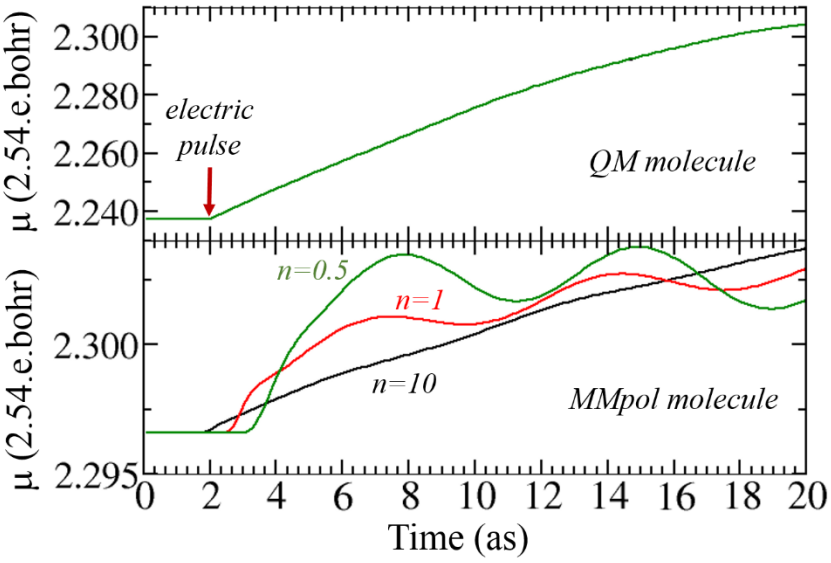


Figure 7. Dipole moment on the QM (Top) and MMpol (Bottom) water molecules after application of a short electric pulse on the QM molecule. Three values of the refractive index are considered, namely 10 (black), 1 (red) and 0.5 (green). Note the different scales between the upper and lower graphs. On the upper graph the three curves are superimposed.

II. RESPONSE MECHANISMS WITH HYBRID RT-TDDFT/MMpol

The methodology used for RT-TDDFT/MMpol with deMon2k has been described in section I. The method allows in principle to account for mutual polarization between the QM region where attosecond dynamics takes place and the environment. Dissipation of energy between the QM region and the environement are also in principle caught up by this computational set-up. In section II we investigate the response mechanisms of a peptide immersed in a water box after application of an ultrashort electric pulse. We first consider simulations without inclusion of retardation in the propagation of electric fields. These effects are analyzed in a second step.

II.1. Dynamics of the response of the environment

To analyze the response of the environment we consider the same system as in I.2.c, namely a methionine enkephalin peptide solvated in a box of POL3 water molecules. After tight SCF convergence, the peptide is perturbed by a Gaussian shaped electric pulse centered at 20 as with 3 as width. The field strengths are set to either 0.001, 0.01 or 0.1 a. u. The simulations have been conducted for 3 fs with a time step of 3 as using the predictor-corrector-Magnus/BCH propagator. We report the variation of the induced dipoles on MM atoms with respect to the initial time ($\Delta\mu(t) = \mu(t) - \mu(0)$) and their normalized auto-correlation functions ($C(t)$, ACF). Both quantities are averaged by hydration layers as indicated by the angular brackets $\langle \dots \rangle$.

We start by considering the upper graphs that correspond to perturbing field strength of 0.001 a.u.. As expected the longer the distance between the water molecules and the peptide, the smaller the impact on the induced dipoles. The first hydration layer is the one that experiences the highest variations of induced dipoles. As evident from the black curve in Figure 8, Top-Left, the average induced dipoles undergo damped oscillations. These are caused by energy dissipation in the MM environment, which is possible thanks to the use of a polarizable FF. Dissipation is very pronounced for the first hydration layer but it is also seen for the outer hydration layers. The induced dipoles for molecules pertaining to the inner hydration layer completely lose correlation within a few tens of as, while beyond 15 Å, the average ACF remains close to 0.8 at 200 as. The characteristic response time is distance dependent. This characteristic time increases for each successive hydration shells. Some of the averaged ACF become negative which is to be related to the oscillatory nature of the variations of $\langle \Delta\mu \rangle$. We finally remark that the response of MM induced dipoles is not fully instantaneous but also exhibits relaxation components over hundreds of attoseconds.

When the strength of the initial perturbing field is increased to 0.01 a.u. (Figure 8 middle) the amplitude of oscillation of the average induced dipoles is larger by a factor of ten. This is true for each hydration shell. When the field strength is further increased to 0.1 a.u., a further increase of response amplitudes is observed for $\langle \Delta\mu \rangle$. The ACFs exhibit more complex evolutions with the increase of perturbing field strength. For the weakest perturbing field strength (0.001a.u.) we already mentioned that the response was distance dependent (Top-Right). For a perturbing field strength of 0.01 a.u. the response of induced dipoles is not distance dependent within the first 50as, apart for water molecules situated beyond 15 Å (Middle-Right). Only after this time a scattering of the average ACFs becomes apparent. Finally, for a perturbing field strength of 0.1 a.u. all the average ACFs but one (again for water molecules situated

beyond 15 Å) are almost superimposed (Bottom-Right). The response mechanism of MM induced dipoles is therefore not distance dependent at all within 15 Å. All these results reflect subtle response mechanisms that deserve a detailed analysis.

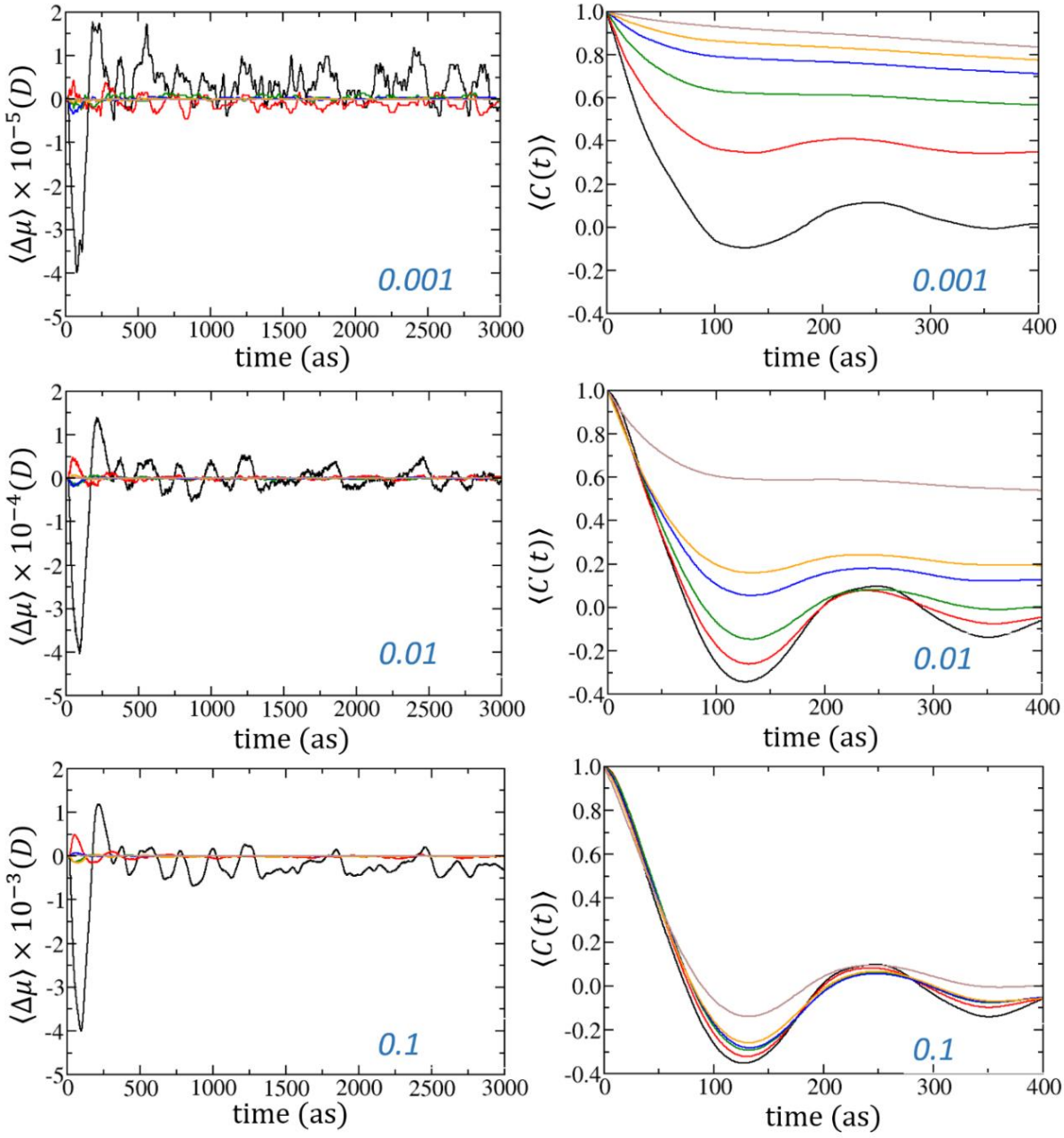


Figure 8. Left: variations of the average induced dipole moment by hydration layers ⟨Δμ(t)⟩=⟨μ(t)-μ(0)⟩. Note the change of scales for each graph. Right: normalized auto-correlation function of the water molecule induced dipole moments averaged by hydration layers for three values of the electric field affecting the peptide at the beginning of the RT-TDDFT/MMpol simulation. Color code: water molecules between 0-3 Å in black, 3-6 Å in red, 6-9 Å in green, 9-12 Å in blue, 12-15 Å in orange and beyond 15 Å in brown. The uncertainties on the curves are around 0.06.

We recall that the induced dipoles are determined by the electric field created by the other MM atoms ($\boldsymbol{F}_i^{(0)} + \boldsymbol{F}_i^{ind}$) and by the QM region ($\boldsymbol{F}_i^{\rho} + \boldsymbol{F}_K^{ZQM}$) (Eqs. 2-7). In the present RT-TDDFT simulations only $\boldsymbol{F}_i^{ind}$ and $\boldsymbol{F}_i^{\rho}$ can account for the variations of the MM induced dipoles since the nuclei are fixed. We also recall that we employ here a mixed non-stationary/stationary RT-TDDFT/MMpol scheme. Accordingly the response of the MM induced dipoles caused by variations of $\boldsymbol{F}_i^{QM\rho}$ are expected to be enhanced in our scheme compared to what they would be in fully dynamical simulation. Nonetheless the average $\langle \Delta\mu(t) \rangle$ and the associated ACF extracted from a 0.1 as time-step simulation were found to be very similar to the graphs shown on Figure 9, thereby indicating that the artificial enhancement of the dipole relaxation due to the RT-TDDFT/MMpol coupling scheme is moderate. The response of MM dipoles should be less pronounced as the distance r increases because of the decay of $\boldsymbol{F}_i^{\rho}$ with distance. On the other hand, the response caused by variation of $\boldsymbol{F}_i^{ind}$ should be associated with a certain delay, because it requires the other induced dipoles to be affected. For example, the induced dipoles of water molecules situated between 6 and 9 Å will be affected by induced dipoles of innermost hydration waters only when their induced dipoles have varied. The prevalence of one mechanism over the other should depend on the relative strength of $\boldsymbol{F}_i^{ind}$ and $\boldsymbol{F}_i^{\rho}$.

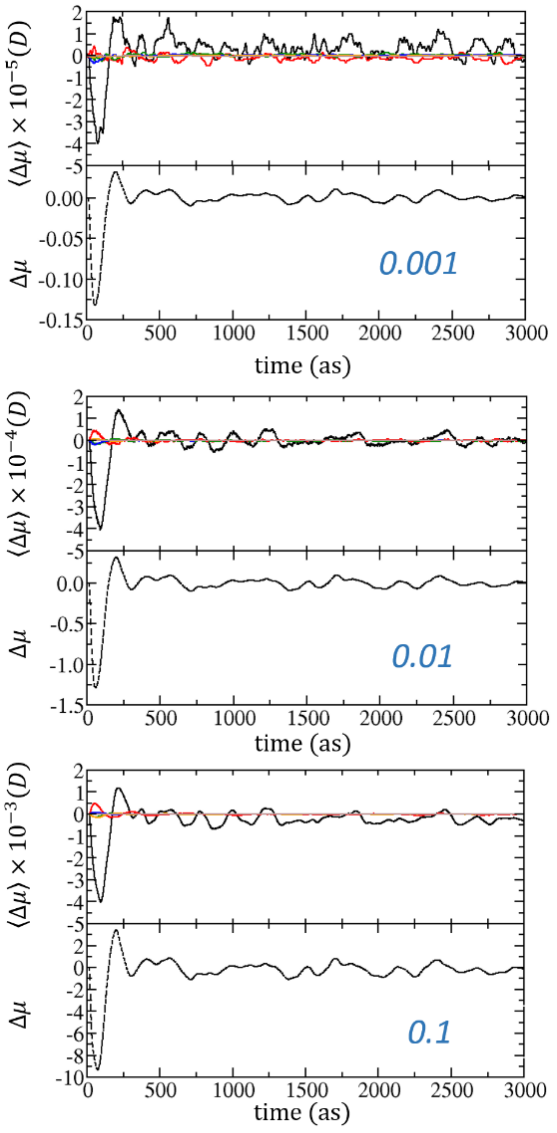


Figure 9. Correlation between the variations of the induced dipole moments on water molecules averaged by hydration shells ⟨Δμ(t)⟩=⟨μ(t)-μ(0)⟩ (upper-part of each graph) and the variation of dipole moment of the peptide (Lower-part of each graph). The three main graphs correspond to three perturbing field strengths of 0.001a.u (Top-Lefts), 0.01 (Bottom-Left) and 0.1 a.u. (Top-Right). The colors are defined in the caption of Figure 8.

For the inner hydration layer (< 3 Å) $\langle C(t) \rangle$ is almost identical whatever the initial perturbation field. It decays to 0.3 in around 50 as (although the variations of induced dipole amplitudes are different for each perturbing electric field). For this hydration layer the source of variation of MM induced dipoles is primarily $\boldsymbol{F}_i^{QM}$. It is the time-dependent field created by

the electron cloud of the peptide that determines the response of MM induced dipoles. The oscillations of the MM induced dipoles essentially follow that of the peptide dipole moment (Figure 9). For the outer hydration layers, the response mechanism depends on the relative importance of $\boldsymbol{F}_i^{ind}$ and $\boldsymbol{F}_i^{\rho}$, the latter being itself dependent on the perturbing field strength. For the strongest perturbing field (0.1 a.u.) the response mechanism of the MM induced dipoles is completely imposed by $\boldsymbol{F}_i^{\rho}$ for all hydration shells (except for water molecules beyond 15 Å). This explains why the average ACFs are almost superimposed. The amplitude of the response decays with distance but the speed at which the induced dipoles vary is the same. In this regime $\boldsymbol{F}_i^{\rho} \gg \boldsymbol{F}_i^{ind}$ so that $\boldsymbol{F}_i^{\rho}$ imposes the response mechanism: the MM induced dipoles within 15 Å follow the variations of the peptide dipole moment. For the intermediate perturbing field (0.01 a. u.) $\boldsymbol{F}_i^{\rho}$ dominates the response mechanism for the shorter distances (<9Å) and for short times (<50 as). At longer distances or after a certain time, 50 as here, $\boldsymbol{F}_i^{ind}$ becomes more important and starts to introduce a distance dependence in the response delay of the MM induced dipoles.

We finally remark that including nuclear motion in the simulation protocol would further make the response mechanisms more complex by allowing $\boldsymbol{F}_i^{(0)}$ to be time-dependent. This would enable to simulate the reorientation polarization of the environment molecule. We also neglected retardation in the propagation of the electric fields mediating the QM/MMpol interations. This will be the topic of the next section.

II.2. Response Dynamics with delayed potentials

In this last section we investigate if the delays taken by the electric fields mediating the interaction between the QM and MMpol regions impact the relaxation mechanisms that we outlined in the previous section. Toward this end we repeated the simulations using the algorithms described in Section I.3 to introduce retarded fields. To avoid artefacts from the RT-TDDFT/stationaryMMpol scheme we use a time step of 0.1 as for these simulations. The results are reported on Figure 10.

The simplest case to start with is the one initiated by the strongest external perturbation (Figure 10, bottom). In this case the induced dipole created on the peptide is so large that $\boldsymbol{F}_K^{\rho}$ dominates over $\boldsymbol{F}_K^{ind}$ on all MM atoms *K* within 15 Å. This is seen on the ACF (Figure 10, bottom, right) for all solvation layers (< 15 Å) that all follow the same trends. The first solvation layer is clearly the most affected. A maximum polarization response is achieved around 80 as after the pulse (which corresponds to the ACF dropping to zero). The simulations have been carried out for $c = 137\ a.u.$ and $c = +\infty$ (which is equivalent to

neglecting retardation). The curves reported in Figure 10 are almost exactly superimposable indicating no influence of retardation. To account for this result, let us first consider the innermost solvation layer (0 to 3 Å) . The water molecules in this layer feel the variations of electric field originating from the peptide with delays of just a few as since $c \approx 3$Å/as. Effectively, as seen in the insets of Figure 10 which depict zooms on the 0-50 as time window, the plain curves (ED simulations with $c = 137\ a.u.$) are delayed by a few attoseconds over the dashed curves (ED simulations with $c = +\infty$). This delay is much less than the overall response mechanism that takes several tens of as. This means that the speed at which $\boldsymbol{F}_K^{\rho}$ is fluctuating is not fast enough compared to the time at which it propagates to induce clear retardation effects on the MM dipoles. The response dynamics remains imposed by $\boldsymbol{F}_K^{\rho}$. One might expect that retardation would be more pronounced for outer solvation layers because the distance to cover for $\boldsymbol{F}_K^{\rho}$ to reach remote K MM atoms is larger. However, the results shown on Figure 10 don't confirm this expectation. In addition, because the electric field created by a dipole decays as r^{-3} retardation effects are damped rapidly with distance.

When considering simulations with weaker initial perturbing electric fields, conclusions are essentially the same. No retardation effects are highlighted. Now $\boldsymbol{F}_K^{\rho}$ is weaker because the initial perturbation of the electron density is smaller. The reasons outlined in the previous case to account for the absence of visible retardation effects in the solvation layers still hold. Regarding $\boldsymbol{F}_K^{ind}$, that now competes with $\boldsymbol{F}_K^{\rho}$ to determine the induced dipole on atom *K*, a similar reasoning applies. Indeed, the dipoles that contribute the most to $\boldsymbol{F}_K^{ind}$ are those of the closest MM atoms *L*, say at a few Å. However, to observe retardation effects the fluctuations of electric fields created by dipoles $\boldsymbol{\mu}_L$ would have to be very pronounced and would have to take place on the attosecond timescales. In fact, as seen from the ACF depicted on Figure 10 the relaxation dynamics of induced dipoles requires tens of attoseconds.

In summary, we have shown that retardation in the mutual electrostatic interactions between the molecule and its environment can be safely neglected. This is due to the fact that electric fields generated by the molecule or its environment do not fluctuate sufficiently rapidly to create noticeable retardation effects. Our simulations in Section II consisted in optical excitations of the central molecule by a laser field. The induced dipole on the molecule was the main source of perturbation of its environment. Because the field created by a dipole decays rapidly with the distance (~$1/r^3$), any retardation in the potential is rapidly damped. In the eventuality of ionization of the central molecule one might eventually expect different conclusions since the overall charge variation might produce a rapidly fluctuating electric field. Our

conclusions which are grounded on the realization of hybrid RT-TDDFT/MMpol simulations also hold for other hybrid schemes like those combining RT-TDDFT with implicit polarizable continuum models[4a, 22].

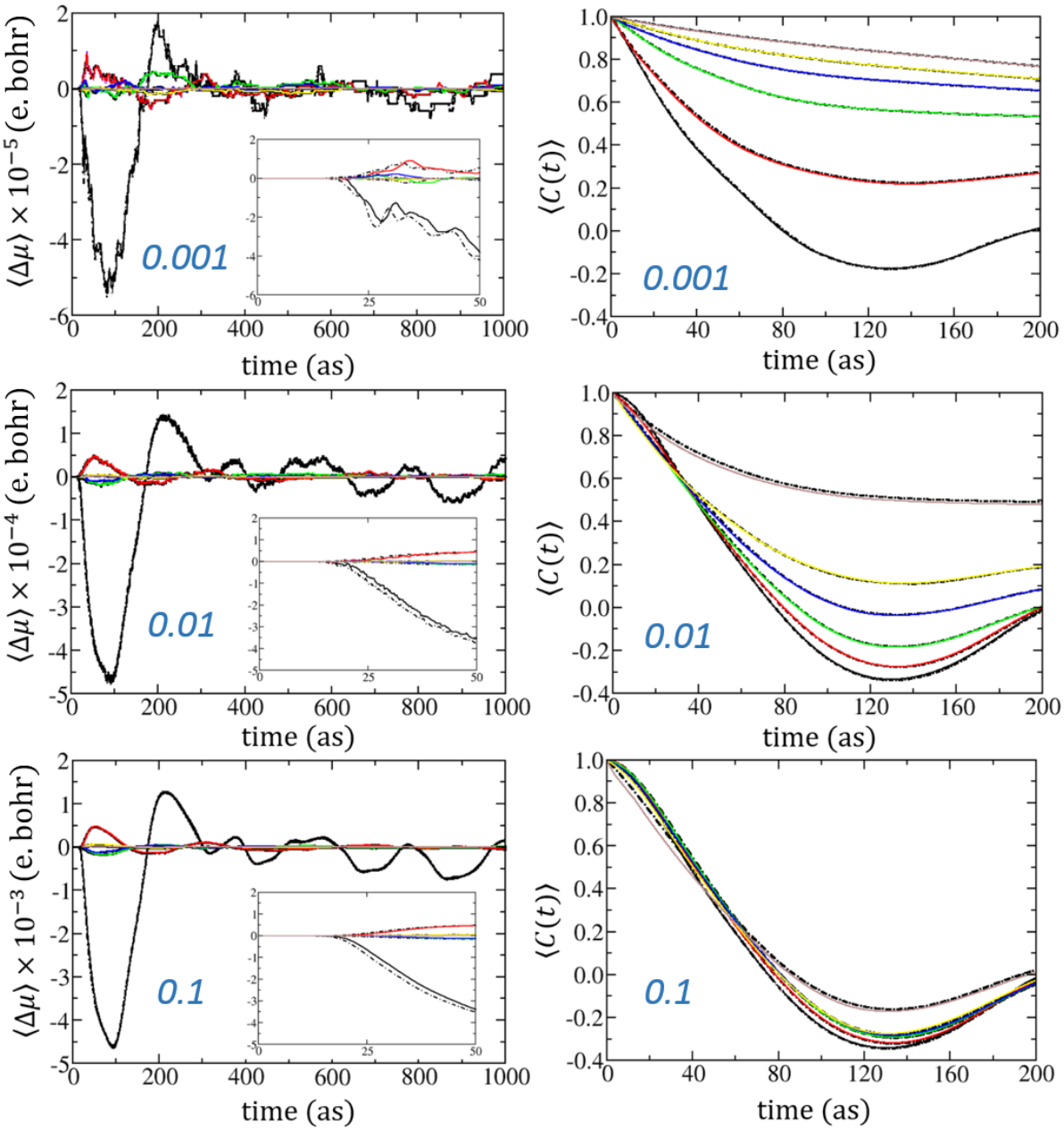


Figure 10. Left: variations of the average induced dipole moment by hydration layers ⟨Δμ(t)⟩ = ⟨μ(t) − μ(0)⟩ (Inset: zoom on 0-50 as). Note the change of scales for each graph. Right: normalized autocorrelation function of the water molecule induced dipole moments averaged by hydration. The upper, middle and lower graphs correspond to initial perturbing electric field of strength 0.001, 0.01 and 0.1 Ha/e.bohr. respectively (atomic units). Color code: 0 to 3 Å (black), 3 to 6 Å (red), 6 to 9 Å (green), 9 to 12 Å (blue), 12 to 15 Å (orange) and finally beyond 15 Å (brown). The dashed line corresponds to RT-TDDFT/MMpol at infinite speed of light (no retardation).

CONCLUSION

In this chapter, we have devised a method to simulate electron dynamics in heterogeneous environments. A hybrid QM/MMpol approach was implemented in deMon2k where the region-of-special-interest is simulated with RT-TDDFT and the environment is approximated with polarizable MM. We carefully justified the model for electronic induction.

The coupling between RT-TDDFT and MMpol is realized with a mixed stationary-non-stationary scheme. This choice assumes that the MM dipoles completely relax at each RT-TDDFT step. Our tests show that it is a valid approximation for sufficiently small time steps.

Response mechanisms of a polarizable environment of a solute peptide submitted to an external perturbation were investigated. The dissipation of energy between the QM region and the environment is caught. The complexity of response mechanisms of the environment was revealed. The same analyses were carried out with retardation effect. Results shown that retardation effects can be safely neglected. This is due to the fact that electric fields generated by the molecule or its environment do not fluctuate sufficiently rapidly to create noticeable retardation effects.

With this implementation, electron dynamics in complex molecular systems like those encountered in biology are now accessible. One possibility is to reduce the cost of induction by implementing more efficient algorithms to compute the potential created by the induced dipoles. The sudden instability of electrodynamics when using a slightly large time step with QM/MMpol schemes should be studied in more detail. GPU could be very good choice to accelerate the calculation of induce dipole of each atom side. More advanced polarizable force fields like AMOEBA could be used to describe the MMpol part.

REFERENCES

1. Jacopo, T.; Roberto, C.; Benedetta, M., Medium effects on the properties of chemical systems: An overview of recent formulations in the polarizable continuum model (PCM). *Int. J. Quantum Chem* **1999,** *75* (4 - 5), 783.
2. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227.
3. Morzan, U. N.; Ramírez, F. F.; Oviedo, M. B.; Sánchez, C. G.; Scherlis, D. A.; Lebrero, M. C. G., Electron dynamics in complex environments with real-time time dependent density functional theory in a QM-MM framework. *J. Chem. Phys.* **2014,** *140* (16), 164105.
4. (a) Corni, S.; Pipolo, S.; Cammi, R., Equation of Motion for the Solvent Polarization Apparent Charges in the Polarizable Continuum Model: Application to Real-Time TDDFT. *J. Phys. Chem. A* **2015,** *119*

(21), 5405; (b) Ding, F.; Lingerfelt, D. B.; Mennucci, B.; Li, X., Time-dependent non-equilibrium dielectric response in QM/continuum approaches. *J. Chem. Phys.* **2015,** *142* (3), 034120.
5. Dinh, P. M.; Reinhard, P. G.; Suraud, E., Dynamics of clusters and molecules in contact with an environment. *Phys. Rep.* **2010,** *485* (2–3), 43.
6. Donati, G.; Wildman, A.; Caprasecca, S.; Lingerfelt, D. B.; Lipparini, F.; Mennucci, B.; Li, X., Coupling Real-Time Time-Dependent Density Functional Theory with Polarizable Force Field. *J. Phys. Chem. lett.* **2017,** *8* (21), 5283.
7. Mei, Y.; Simmonett, A. C.; Pickard, F. C.; DiStasio, R. A.; Brooks, B. R.; Shao, Y., Numerical Study on the Partitioning of the Molecular Polarizability into Fluctuating Charge and Induced Atomic Dipole Contributions. *J. Phys. Chem. A* **2015,** *119* (22), 5865.
8. Lamoureux, G.; Roux, B. t., Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003,** *119* (6), 3025.
9. (a) Souaille, M.; Loirat, H.; Borgis, D.; Gaigeot, M. P., MDVRY: a polarizable classical molecular dynamics package for biomolecules. *Comput. Phys. Commun.* **2009,** *180* (2), 276; (b) Tabacchi, G.; Mundy, C. J.; Hutter, J.; Parrinello, M., Classical polarizable force fields parametrized from ab initio calculations. *J. Chem. Phys.* **2002,** *117* (4), 1416; (c) Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009,** *124* (1), 11.
10. (a) Krause, P.; Schlegel, H. B., Strong-field ionization rates of linear polyenes simulated with time-dependent configuration interaction with an absorbing potential. *J. Chem. Phys.* **2014,** *141* (17), 174104; (b) Krause, P.; Sonk, J. A.; Schlegel, H. B., Strong field ionization rates simulated with time-dependent configuration interaction and an absorbing potential. *J. Chem. Phys.* **2014,** *140* (17), 174113; (c) Parise, A.; Alvarez-Ibarra, A.; Wu, X.; Zhao, X.; Pilmé, J.; Lande, A. d. l., Quantum Chemical Topology of the Electron Localization Function in the Field of Attosecond Electron Dynamics. *J. Phys. Chem. Lett.* **2018**, 844.
11. (a) Nagano, R.; Yabana, K.; Tazawa, T.; Abe, Y., Application of the time-dependent local density approximation to collision between a highly charged ion and an atom. *J. Phys. B: At., Mol. Opt. Phys.* **1999,** *32* (4), L65; (b) Covington, C.; Hartig, K.; Russakoff, A.; Kulpins, R.; Varga, K., Time-dependent density-functional-theory investigation of the collisions of protons and α particles with uracil and adenine. *Phys. Rev. A* **2017,** *95* (5), 052701.
12. Li, X.; Tully, J. C.; Schlegel, H. B.; Frisch, M. J., Ab initio Ehrenfest dynamics. *J. Chem. Phys.* **2005,** *123* (8), 084106.
13. Salahub, D.; Noskov, S.; Lev, B.; Zhang, R.; Ngo, V.; Goursot, A.; Calaminici, P.; Köster, A.; Alvarez-Ibarra, A.; Mejía-Rodríguez, D.; Řezáč, J.; Cailliez, F.; de la Lande, A., QM/MM Calculations with deMon2k. *Molecules* **2015,** *20* (3), 4780.
14. Ponder, J. *TINKER - Software Tools for Molecular Design*, 7.1.2.
15. OPLS Force Fields. In *Encyclopedia of Computational Chemistry*.
16. Brooks, B. R.; III, C. L. B.; Jr, A. D. M.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545.
17. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157.
18. Caldwell, J. W.; Kollman, P. A., Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. *J. Phys. Chem.* **1995,** *99* (16), 6208.
19. Köster, A. M.; Reveles, J. U.; del Campo, J. M., Calculation of exchange-correlation potentials with auxiliary function densities. *J. Chem. Phys.* **2004,** *121* (8), 3417.

20. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P., *Numerical Recipes in Fortran 77 and 90 (Second Edition)*. Cambridge University Press: Cambridge 2001.
21. Dunlap, B. I.; Rösch, N.; Trickey, S. B., Variational fitting methods for electronic structure calculations. *Mol. Phys.* **2010,** *108* (21-23), 3167.
22. Nguyen, P. D.; Ding, F.; Fischer, S. A.; Liang, W.; Li, X., Solvated First-Principles Excited-State Charge-Transfer Dynamics with Time-Dependent Polarizable Continuum Model and Solvent Dielectric Relaxation. *J. Phys. Chem. Lett.* **2012,** *3* (19), 2898.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Simulating Electron Dynamics in Polarizable Environments

Xiaojing Wu,[†] Jean-Marie Teuler,[†] Fabien Cailliez,[†] Carine Clavaguéra,[†] Dennis R. Salahub,[‡,§] and Aurélien de la Lande*[,†]

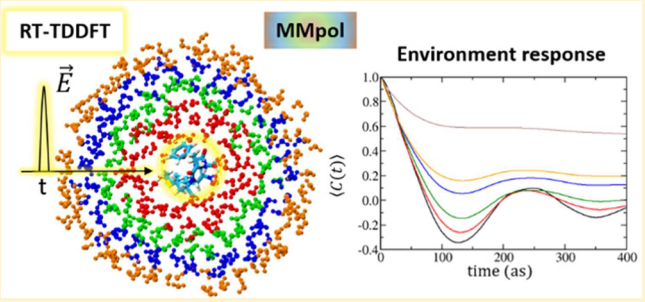
[†]Laboratoire de Chimie Physique, CNRS - Université Paris Sud, Université Paris-Saclay, 15 avenue Jean Perrin, 91405 Orsay CEDEX, France

[‡]Department of Chemistry, Centre for Molecular Simulation, Institute for Quantum Science and Technology and Quantum Alberta, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

[§]College of Chemistry and Chemical Engineering, Henan University of Technology, No. 100, Lian Hua Street, High-Tech Development Zone, Zhengzhou 450001, P. R. China

Supporting Information

ABSTRACT: We propose a methodology for simulating attosecond electron dynamics in large molecular systems. Our approach is based on the combination of real time time-dependent-density-functional theory (RT-TDDFT) and polarizable Molecular Mechanics (MMpol) with the point-charge-dipole model of electrostatic induction. We implemented this methodology in the software deMon2k that relies heavily on auxiliary fitted densities. In the context of RT-TDDFT/MMpol simulations, fitted densities allow the cost of the calculations to be reduced drastically on three fronts: (i) the Kohn−Sham potential, (ii) the electric field created by the (fluctuating) electron cloud which is needed in the QM/MM interaction, and (iii) the analysis of the fluctuating electron density on-the-fly. We determine conditions under which fitted densities can be used without jeopardizing the reliability of the simulations. Very encouraging results are found both for stationary and time-dependent calculations. We report absorption spectra of a dye molecule in the gas phase, in nonpolarizable water, and in polarizable water. Finally, we use the method to analyze the distance-dependent response of the environment of a peptide perturbed by an electric field. Different response mechanisms are identified. It is shown that the induction on MM sites allows excess energy to dissipate from the QM region to the environment. In this regard, the first hydration shell plays an essential role in absorbing energy. The methodology presented herein opens the possibility of simulating radiation-induced electronic phenomena in complex and extended molecular systems.



I. INTRODUCTION

Recent years have seen a growing interest in the electron dynamics taking place in molecules when they are subjected to an external perturbation. This interest has been stimulated by progress in attosecond spectroscopy that now gives access to details on electron dynamics. The realm of subfemtosecond electron dynamics involves fascinating processes such as ultrafast charge migration,[1] Auger decays, and Intra Coulomb Decays.[2−5] These are not driven by nuclear dynamics but instead by electron correlation and energy redistribution.[6] This nascent research field has led to new debated concepts like attosecond chemistry, a possible promise of which would be the possibility to control chemical reactions by the control of electronic motion.[7] Electron dynamics is also important in the description of ultrafast nonadiabatic molecular dynamics. The relaxation pathways within molecules electronically excited or ionized by a photon or a high-energy particle are particularly rich and complex. They involve coupled electron−nuclear dynamics.[8]

On the computational side, much effort has been spent to devise simulation algorithms of electron dynamics. In the family of wave function approaches the TD-HF (Time-dependent-Hartree−Fock),[9] TD-CI (Time-dependent-configuration interaction),[10,11] or the TD-MCSCF (Time-dependent multiconfiguational self consistent field)[12,13] methods have been developed. Another popular approach for simulating electron dynamics relies on time-dependent density functional theory (TDDFT). This approach is frequently referred to as real-time TDDFT (RT-TDDFT) to distinguish it from the linear response (LR-TDDFT) formalism. The latter relies on perturbation theory to simulate UV−visible absorption spectra.[14] Although not exempt from intrinsic limitations like the self-interaction-error,[15,16] a noticeable advantage of TDDFT is its excellent computational cost/accuracy ratio. TDDFT can be applied to molecular systems comprised of hundreds of atoms. TDDFT finds its root in the seminal work of Runge and Gross.[17] Under the Kohn−Sham framework that refers to a fictitious reference system of noninteracting electrons, the coupled time-dependent KS equations describe

Received: March 10, 2017
Published: July 24, 2017

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

the time evolution of the KS molecular orbitals, hence the dynamics of the electron density of the real system. RT-TDDFT has been used to calculate static and dynamic polarizabilities and hyperpolarizabilities of molecules,[18] to simulate UV−visible spectra of molecules[19,20] and of nanoparticles,[21] to simulate core-level near-edge X-ray absorption spectra,[22,23] to simulate photoelectron emission spectra,[24,25] electron conductance in electronic junctions,[26,27] photoinduced electron transfer,[28,29] magnetization dynamics in inorganic complexes,[30] attosecond dynamics following X-ray photoionization of gas molecules,[3] or charge migration following radiolysis of water.[5,31] RT-TDDFT has also been coupled to mean-field (Ehrenfest) nuclear dynamics to simulate nonadiabatic processes[32−34] with many interesting applications, for example to simulate the ultrafast dynamics of photoexcited metal complexes or in optimal control of chemical reactions.[8]

Curiously enough, most implementations have been designed for molecular systems in the gas phase and not in contact with environments. If the systems of interest are periodic, periodic boundary conditions can be used to simulate infinite systems,[35] but in many cases, systems are not periodic and alternatives must be found. A challenge is to account for the electronic response of the environment due to the changes in the electronic structure of the molecule and vice versa. The environment may be homogeneous (solutions) in which case a polarizable dielectric continuum (PCM) can be used. The environment may also be heterogeneous, as for example for extended biosystems (DNA, proteins, lipid membranes), nanoclusters, or interfaces. In such cases hybrid QM/MMpol (i.e., using polarizable force fields)[36] constitutes a method of choice to retain the atomistic details of the environment at moderate computational cost. In the linear response formalism, coupling between TDDFT and either PCM or polarizable QM/MMpol has been devised.[37−39] For explicit propagation in time of the TDDFT equations, a further challenge is to account, by definition, for the time-dependence of the environment's response. Remote atoms should take longer times to respond than closer ones, for instance. Li and coworkers developed a combined RT-TDDFT/PCM[40,41] method, with applications to charge transfer dynamics in bulk heterojunction models.[42] In their PCM model the dielectric constant of the environment was made time-dependent even though the PCM was made stationary with the evolving potential created by the QM region. Corni et al.[43,44] as well as Ding et al.[45] later described a more general approach of RT-TDDFT/PCM calculations where both the PCM and QM region were propagated in time. Regarding hybrid RT-TDDFT/MMpol approaches Dinh et al. reported a few years ago a coupling between RT-TDDFT and a polarizable force field (FF).[46] Induction was introduced by distinguishing core from valence electrons on MM atoms. The average position of the core electrons + nucleus and the average position of valence electrons had the possibility to be different depending on the electrical environment, thereby creating induced dipoles. The authors reported insightful applications to sodium clusters deposited on metal surfaces[46] with detailed analyses of their optical properties. In the computational setup of Dinh et al. the RT-TDDFT engine itself relies on a grid-based implementation of DFT. This is quite different from the algorithms employed in the community of quantum chemists that generally rely on local basis sets (Gaussian or Slater atomic orbitals). There is thus a clear need to develop hybrid RT-TDDFT/MMpol schemes for the modeling of electron dynamics with local basis sets in extended molecular systems.

The structure of the present article is as follows. First, we report an implementation of polarizable QM/MMpol based on the charge-induced dipole model[36] of electronic induction in the software deMon2k.[47] Second, we describe our implementation of RT-TDDFT and its coupling with polarizable MM. In both modules density fitting techniques are used to reduce the computational cost drastically.[48,49] In section III we carefully test the reliability of substituting the Kohn−Sham density by the auxiliary density for propagating the electron dynamics or for calculating the QM/MMpol coupling interactions. Very encouraging results are obtained. We investigate in section IV the time-dependent electronic response of molecules in vacuum and in solution.

II. METHODOLOGY

II.1. Auxiliary Kohn−Sham Density Functional Theory.

We start the Methodology section by recalling the general DFT framework implemented in deMon2k. This program solves the Kohn−Sham DFT equations with KS molecular orbitals (MO) $\psi_i(\boldsymbol{r})$ represented as linear combinations of Gaussian-type atomic orbitals[50] (LCGTAO). For simplicity, we will consider only closed-shell molecules, but we mention that the methodologies presented in this work have been adapted to the open-shell case too.

$$\psi_i(\boldsymbol{r}) = \sum_{\sigma=1} c_{\sigma i}(t)\sigma(\boldsymbol{r}) \tag{1}$$

$$\rho(t, \boldsymbol{r}) = \sum_{\sigma,\tau} P_{\sigma\tau}(t)\sigma(\boldsymbol{r})\tau(\boldsymbol{r}) \tag{2}$$

$$P_{\sigma\tau}(t) = 2\sum_{i}^{N_{MO}/2} c^{*}_{\sigma i}(t)c_{\tau i}(t) \tag{3}$$

Greek letters are used both as indices and as AO function names. The MO coefficients ($c_{\sigma i}$), hence the density matrix ($P_{\sigma\tau}$) and the electron density (ρ), depend on time. When solving the stationary KS equations, this dependence would by definition vanish, but we keep here the more general RT-TDDFT formulation. Note that the MO coefficients are complex numbers in real time propagation. deMon2k relies heavily on the variational density fitting method originally introduced by Dunlap[48] to avoid the calculation of four-center electron repulsion integrals (ERIS). The fitted densities ($\tilde{\rho}$) are expressed as linear combinations of auxiliary basis functions k: $\tilde{\rho}(\boldsymbol{r}) = \sum_k x_k k(\boldsymbol{r})$. For computational efficiency the auxiliary basis functions k are Hermite Gaussian polynomials that are grouped by functions sharing the same exponents.[50] With this auxiliary density, the electronic energy expression for an isolated molecule reads

$$E(t) = \sum_{\sigma,\tau} P_{\sigma\tau}(t)H_{\sigma\tau} + \sum_{\sigma,\tau}\sum_{k} P_{\sigma\tau}(t)\langle\sigma\tau||k\rangle x_k(t) - \frac{1}{2}\sum_{k,l} x_k(t)x_l(t)\langle k||l\rangle + E_{xc}[\rho] \tag{4}$$

The symbol || stands for the coulomb operator ($1/|r_1 - r_2|$). As evident in eq 4, no four-centers ERIS are needed but only two- and three-centers ERIS. $H_{\sigma\tau}$ are the matrix elements of the core Hamiltonian, encompassing the kinetic energy and the electron-nuclei attraction. The Kohn−Sham potential matrix

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

elements for an isolated molecule are obtained by differentiating the total energy with respect to the density matrix elements:

$$H^{\mathrm{isol}}_{\sigma\tau} \equiv \frac{\partial E}{\partial P_{\sigma\tau}} = H_{\sigma\tau} + \sum_k \langle \sigma\tau \| k \rangle x_k(t) + \frac{\partial E_{\mathrm{xc}}[\rho]}{\partial P_{\sigma\tau}} \tag{5}$$

deMon2k offers also the possibility to use the fitted density in the calculation of the exchange-correlation (XC) energy, in which case $E_{\mathrm{xc}}[\rho]$ is replaced by $E_{\mathrm{xc}}[\tilde{\rho}]$ in eq 4 and eq 5.[51] Now the KS potential does not depend explicitly on the KS density but only on the auxiliary density. We refer to this framework as Auxiliary DFT.[52]

II.2. Polarizable QM/MM in deMon2k. *Model of Electronic Induction.* There are various ways to carry out DFT/MM calculations with deMon2k, and we refer the interested reader to a recent review describing these alternatives.[53] In the present work we focus on the so-called in-deMon2k QM/MM by which both QM, here (TD)DFT, and MM calculations are carried out by deMon2k, without using program interfaces.[53] We expect advantages in terms of data passing management between the DFT and MM modules. The objective of the present work is to upgrade the pre-existing QM/MM method to QM/MMpol. Electrostatic induction can be introduced in classical force fields in different ways.[36,54−56] We note, for example, that the Drude polarizable force field has recently shown great promise for ions interacting with protein models, for which additive fixed-charge force fields come up short.[57,58] We have chosen here to consider the point-charge dipole model by which induction is simulated by the introduction of induced dipoles ($\boldsymbol{\mu}_i$, note that vectors are written in bold) on every polarizable MM site i.[36,37] Each induced dipole is determined from the electric field $\boldsymbol{F}_i$ at the MM atom position. $\boldsymbol{F}_i$ stems from the electric field created by other MM permanent charges ($\boldsymbol{F}_i^{(0)}$) and by other MM induced dipoles ($\boldsymbol{F}_i^{\mathrm{ind}}$). In QM/MMpol calculations one further adds the electric field created by the QM region ($\boldsymbol{F}_i^{\mathrm{QM}}$), that is by the atomic nuclear charges ($\boldsymbol{F}_i^{\mathrm{ZQM}}$) and by the electron density ($\boldsymbol{F}_i^{\rho}$). We also introduced the possibility of adding an external electric field ($\boldsymbol{F}_i^{\mathrm{ext}}$) to mimic for instance the interaction with the electric part of an electromagnetic wave. The mathematical expressions for the various contributions are given by

$$\boldsymbol{\mu}_i = \alpha_i \boldsymbol{F}_i = \alpha_i \left(\boldsymbol{F}_i^{(0)} + \boldsymbol{F}_i^{\mathrm{ind}} + \boldsymbol{F}_i^{\mathrm{QM}} + \boldsymbol{F}_i^{\mathrm{ext}} \right) \tag{6}$$

$$\boldsymbol{F}_i^{(0)} = \sum_{\substack{j \in \mathrm{MM} \\ j \neq i}} \frac{q_j}{r_{ij}^3} \boldsymbol{r}_{ij} \tag{7}$$

$$\boldsymbol{F}_i^{\mathrm{ind}} = - \sum_{\substack{j \in \mathrm{MM} \\ j \neq i}} \boldsymbol{T}_{ij} \mu_i \tag{8}$$

$$\boldsymbol{T}_{ij} = \frac{1}{r_{ij}^3} \boldsymbol{I} - \frac{3}{r_{ij}^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \tag{9}$$

$$\boldsymbol{F}_i^{\mathrm{QM}} = \boldsymbol{F}_i^{\mathrm{ZQM}} + \boldsymbol{F}_i^{\rho} = \sum_{k \in \mathrm{QM}} \frac{Z_k}{r_{ik}^3} \boldsymbol{r}_{ik} - \int \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}_i - \boldsymbol{r}|^3} (\boldsymbol{r}_i - \boldsymbol{r}) \mathrm{d}r \tag{10}$$

where α_i is the polarizability of MM atom i. It is assumed to be isotropic. q_j is the charge of MM atom j; $\boldsymbol{r}_{ij}$ is the vector between atoms i and j; $\boldsymbol{T}_{ij}$ is the dipole−dipole interaction tensor, and $\boldsymbol{I}$ is the identity matrix. Z_k is the nuclear charge of QM nucleus k. The total induction energy is comprised of three terms that reflect the interaction between the MM induced dipoles with (i) the MM permanent charges ($E^{\mathrm{ind}}_{\mu-\mathrm{qMM}}$), (ii) the atomic nuclei of the QM atoms ($E^{\mathrm{ind}}_{\mu-\mathrm{ZQM}}$), and (iii) the electron cloud ($E^{\mathrm{ind}}_{\mu-\rho}$).

$$E^{\mathrm{ind}}_{\mathrm{tot}} = E^{\mathrm{ind}}_{\mu-\mathrm{qMM}} + E^{\mathrm{ind}}_{\mu-\mathrm{ZQM}} + E^{\mathrm{ind}}_{\mu-\rho} \tag{11}$$

Contributions (i) and (ii) are calculated as $-1/2 \sum_{i \in \mathrm{MM}} \boldsymbol{\mu}_i \cdot \boldsymbol{F}_i^{(0)}$ and $-1/2 \sum_{i \in \mathrm{MM}} \boldsymbol{\mu}_i \cdot \boldsymbol{F}_i^{\mathrm{ZQM}}$, respectively. The last term depends on the electron density and is given by

$$\begin{aligned} E^{\mathrm{ind}}_{\mu-\rho} &= \frac{1}{2} \sum_{i \in \mathrm{MM}} \int \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}_i - \boldsymbol{r}|^3} \boldsymbol{\mu}_i \cdot (\boldsymbol{r}_i - \boldsymbol{r}) \mathrm{d}r \\ &= \frac{1}{2} \sum_{i \in \mathrm{MM}} \sum_{\sigma,\tau} P_{\sigma\tau} \left\langle \sigma \left| \frac{\boldsymbol{\mu}_i (\boldsymbol{r}_i - \boldsymbol{r})}{|\boldsymbol{r}_i - \boldsymbol{r}|^3} \right| \tau \right\rangle \end{aligned} \tag{12}$$

QM/MMpol calculations must capture the interdependence between the electron cloud and the polarizable environment. For stationary DFT calculations this is done by regularly updating the induced dipoles that depend on $\boldsymbol{F}_i^{\mathrm{QM}}$ and the electrostatic potential created by the induced dipoles to be included in the KS Hamiltonian. We use a similar algorithm for RT-TDDFT propagation. We will come back to this methodological point at the end of section II. Note that besides $E^{\mathrm{ind}}_{\mu-\rho}$, the interaction between the QM and MM atoms also includes the electrostatic energy between the electron cloud and the MM permanent charges (q_i).

$$E^{\mathrm{perm}}_{\mathrm{q}-\rho} = \sum_{i \in \mathrm{MM}} \int \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}_i - \boldsymbol{r}|} q_i \cdot \mathrm{d}\boldsymbol{r} = \sum_{i \in \mathrm{MM}} \sum_{\sigma,\tau} q_i P_{\sigma\tau} \left\langle \sigma \left| \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}|} \right| \tau \right\rangle \tag{13}$$

Within this electrostatic embedding scheme, the potential created by permanent charges and induced dipoles on MM sites are obtained by differentiation of the respective interaction energies with respect to the electronic density. The calculation of electrostatic integrals ($\left\langle \sigma \left| \frac{1}{r} \right| \tau \right\rangle$) required in eq 13 was optimized by Alvarez et al.[53,59] To take advantage of these algorithmic developments we represent each induced dipole $\boldsymbol{\mu}_i$ by two charges of opposite sign ($\pm\delta_i$) separated by 0.5 bohr and centered around the MM atom positions.[60−62] This way the potential created by the induced dipoles ($H^{\mu\rho}_{\sigma\tau}$) are included in the KS potential via a set of point charges, the calculation of which is performed efficiently in deMon2k.[53,59]

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

$$H_{\sigma\tau}^{\mu\rho} \equiv \frac{\partial E_{\mu-\rho}^{\mathrm{ind}}}{\partial P_{\sigma\tau}} = \sum_{i\in \mathrm{MM}} \left\langle \sigma \left| \frac{\boldsymbol{\mu}_i}{|\boldsymbol{r}_i - \boldsymbol{r}|^3} \right| \tau \right\rangle$$
$$= \sum_{i\in \mathrm{MM}} \left[\delta_i \left\langle \sigma \left| \frac{1}{\left| \boldsymbol{r}_i - \boldsymbol{r} + 0.5\frac{\boldsymbol{\mu}_i}{\|\boldsymbol{\mu}_i\|} \right|} \right| \tau \right\rangle - \delta_i \left\langle \sigma \left| \frac{1}{\left| \boldsymbol{r}_i - \boldsymbol{r} - 0.5\frac{\boldsymbol{\mu}_i}{\|\boldsymbol{\mu}_i\|} \right|} \right| \tau \right\rangle \right] \quad (14)$$

As for any QM/MM scheme, a critical point of DFT/MMpol and RT-TDDFT/MMpol calculations is to set the boundaries between MM and QM regions. Setting boundaries across polar groups may deteriorate the efficiency of the hybrid energies and the derived properties. The choice of QM/MM partitioning is the responsibility of the user. Our QM/MMpol implementation does not rely on interfaces between QM and MM software, and both DFT and MM calculations are done within deMon2k. This is a critical advantage to reach efficient DFT/MMpol calculations. Indeed, passing information between the DFT and MM branches of the same program is rapid compared to I/O operations. The electric field and the induced dipoles are vectors that can be stored in random access memory (RAM), and one can easily restart convergence of induced dipoles at every new SCF cycle or RT-TDDFT time step from previous steps. Finally, our in-deMon2k QM/MMpol uses a direct algorithm similar to the calculation of electron repulsion integrals (ERIS).[50,63] At every SCF cycle or RT-TDDFT step n, instead of calculating the QM electric field $\boldsymbol{F}_i^{\mathrm{QM},(n)}$ from the current density matrix $P^{(n)}$, we increment it from $\boldsymbol{F}_i^{\mathrm{QM},(n-1)}$ and the difference density $\Delta P = P^{(n)} - P^{(n-1)}$, namely $\boldsymbol{F}_i^{\mathrm{QM},(n)} = \boldsymbol{F}_i^{\mathrm{QM},(n-1)} - \sum_{\mu,\nu} \Delta P_{\mu\nu} \left\langle \mu \left| \frac{\boldsymbol{r}}{r^3} \right| \nu \right\rangle$. This procedure has the advantage that one can screen many terms of the sum if $\Delta P_{\mu\nu}$ is below a predefined threshold. This helps to decrease the computational time in SCF calculations when the density is close to convergence or in RT-TDDFT simulations when the density evolves slowly. Direct SCF procedures have been used by other groups in the context of QM/MMpol calculations.[64,65]

Polarization Catastrophe. A well-known pitfall of polarizable force fields is the risk of "polarization catastrophe" the origin of which has been exposed by Thole.[56,66,67] This term defines a divergence of the polarization energy that happens when adjacent dipoles align on the same line in head-to-tail configurations. Most polarizable MM implementations avoid the polarization catastrophe by damping the electric fields at short distance. For MM atoms bonded in 1-2, 1-3, or 1-4 positions to a given polarizable MM site, the electronic field can be simply ignored. In deMon2k, choice is given to the user to set up these parameters. For nonbonded atoms electric damping is achieved by the modification of the dipole interaction tensor with two distance-dependent screening functions (f_e and f_t)

$$\boldsymbol{T}_{ij} = \frac{f_e}{r_{ij}^3}\boldsymbol{I} - \frac{3f_t}{r_{ij}^5}\begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \quad (15)$$

Three alternatives of the screening function have been implemented in deMon2k following previous proposals reported in the literature. One is the linear scheme[67]

$$\upsilon = r_{ij}/s \text{ with } s = a(\alpha_i\alpha_j)^{1/6} \quad (16)$$

$$f_e = \begin{cases} 1.0 & \text{if } r_{ij} > s \\ 4\upsilon^3 - 3\upsilon^4 & \text{if } r_{ij} < s \end{cases} \quad (17)$$

$$f_t = \begin{cases} 1.0 & \text{if } r_{ij} > s \\ \upsilon^4 & \text{if } r_{ij} < s \end{cases} \quad (18)$$

another is the exponential scheme

$$\upsilon = ar_{ij}/(\alpha_i\alpha_j)^{1/6} \quad (19)$$

$$f_e = 1 - \left(\frac{\upsilon^2}{2} + \upsilon + 1\right)\exp(-\upsilon) \quad (20)$$

$$f_t = 1 - \left(\frac{\upsilon^3}{6} + \frac{\upsilon^2}{2} + \upsilon + 1\right)\exp(-\upsilon) \quad (21)$$

while the third one is the Tinker-exponential form

$$u = r_{ij}/(\alpha_i\alpha_j)^{1/6} \quad (22)$$

$$f_e = 1 - \exp(-au^3) \quad (23)$$

$$f_t = 1 - (1 + au^3)\exp(-au^3) \quad (24)$$

In these expressions a is a unitless parameter that depends on the force field. The higher the value of a, the faster the field damping with distance. In the water model of the AMOEBA force field a is, for example, set to 0.39.[68]

On the Use of Fitted Densities. We have tested various alternatives for estimating $\boldsymbol{F}_i^{\mathrm{QM},}$. The most correct way is to calculate it from the Kohn−Sham density ρ (eqs 2 and 10). Since $\boldsymbol{F}_i^{\mathrm{QM},}$ needs to be evaluated on every MM site at every SCF cycle and at every RT-TDDFT step, this task can become computationally expensive. A tempting alternative is to replace ρ by the auxiliary density $\tilde{\rho}$. Because the number of auxiliary basis functions is typically four to five times lower than the number of products of atomic basis functions, substituting ρ by $\tilde{\rho}$ is expected to drastically reduce the cost of calculation of $\boldsymbol{F}_i^{\mathrm{QM},}$. Such a substitution is, however, not guaranteed to yield reliable results. This is because auxiliary fitted densities are not designed to reproduce ρ but to provide auxiliary densities from which approximate electronic repulsion interactions can be computed with reduced computational cost. That said, we recently showed that electrostatic multipoles obtained either

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

from the KS density or from the fitted density are very similar, provided sufficiently flexible auxiliary basis sets are used.[69] This is an encouraging result. Indeed, we may expect that if the intrinsic multipoles on QM atoms extracted from $\tilde{\rho}$ are similar to those extracted from ρ, so will be the electric fields generated by $\tilde{\rho}$ and ρ. We will thus test the accuracy of stationary and time-dependent DFT/MMpol calculations when replacing ρ by $\tilde{\rho}$.

II.3. Electron Dynamics Equations-of-Motion. We now move to the description of our RT-TDDFT implementation. It is largely based on algorithmic developments reported previously by other groups in the past decade.[9,20,26,70] We thus refer the reader to the original publications. We insist here on the specificities of the implementation in deMon2k and on the novel features that we have introduced, notably the coupling between RT-TDDFT and the QM/MMpol just described. Runge and Gross developed the many body wave function TD Schrödinger equation into the single-particle TD density Kohn−Sham (TDKS) equation with an effective Hamiltonian $H(t)$ uniquely described by the TD electron density $\rho(t)$[17]

$$i\frac{\partial \psi_i(t)}{\partial t} = H[\rho(t)]\psi_i(t) \tag{25}$$

where H is the time-dependent Kohn−Sham operator which is a functional of the charge density. It includes the KS potential of the isolated molecule, the matrix elements of which are given by eq 5, and the interaction potentials of the electron cloud with external electric fields H^{ext}. In QM/MMpol the KS operator also includes the perturbation from the MM charges ($H^{q\rho}$) and MM induced dipoles ($H^{\mu\rho}$).

$$H(t) = H^{\text{isol}}(t) + H^{q\rho} + H^{\mu\rho}(t) + H^{\text{ext}}(t) \tag{26}$$

H^{isol} includes the contribution from the XC potential. We make the adiabatic approximation and consider only the spatial dependence of the XC potential, neglecting its temporal nonlocality. Eq 25 can be recast using the density matrix $P(t)$ into a Liouville-von Neumann type of equation, which reads in the case of a nonorthogonal basis set

$$iS\frac{\partial P(t)}{\partial t}S = H(t)P(t)S - SP(t)H(t) \tag{27}$$

where S is the overlap matrix in the atomic orbital basis set. As in refs 9 and 20 we transform H and P to the orthogonal MO basis leading to H' and P'. Since the MO are orthogonal eq 27 simplifies to

$$i\frac{\partial P'(t)}{\partial t} = [H'(t), P'(t)] \tag{28}$$

We will use primes to denote matrices in the molecular orbital (MO) basis and no primes to denote matrices in the atomic orbital (AO) basis. The formal solution of eq 28 can be expressed as

$$P'(t) = U(t, t_0)P'(t_0)U^{\dagger}(t, t_0) \tag{29}$$

where U is the evolution operator, which can be discretized into small time steps Δt,

$$U(t, t_0) = \prod_i^{n-1} U(t_i + \Delta t, t_i) \tag{30}$$

$$U(t + \Delta t, t) = \mathcal{T}\exp\{-i\int_t^{t+\Delta t} H'(\tau)\mathrm{d}\tau\} \tag{31}$$

and $\mathcal{T}$ is the time-ordering operator, ensuring that operators associated with later times always appear to the left of those associated with earlier times. Many schemes have been proposed to evaluate the propagator in RT-TDDFT, and we refer the reader to recent reviews describing the physical conditions that propagators should fulfill.[70,71] We have implemented in deMon2k the Euler and second-order Magnus propagators.

Euler Propagation. To solve eq 28 by applying Lagrange's Mean Value Theorem we obtain

$$P'(t + \Delta t) = P'(t) - i[H'(t), P'(t)]*\Delta t \tag{32}$$

The propagation of the density matrix requires only the value of the density matrix and the Kohn−Sham matrix at the current time. These are easy to obtain; however, this propagation scheme does not guarantee the preservation of the norm of the KS wave function which can lead to divergence of the electronic propagation. We found the Euler propagation to be unstable in most of our applications, and it will not be considered any further in this article.

Magnus Propagation. A convenient solution to eq 31 is given by a Magnus expansion[72]

$$\mathcal{T}\exp\{-i\int_t^{t+\Delta t} H'(\tau)d\tau\} = e^{W} = e^{\Omega_1+\Omega_2+\ldots} \tag{33}$$

where $\{\Omega_i\}$ is a series of nested commutator integrals:

$$\Omega_1(t + \Delta t, t) = -i\int_t^{t+\Delta t} H'(\tau)\mathrm{d}\tau \tag{34}$$

$$\Omega_2(t + \Delta t, t) = -i\int_t^{t+\Delta t} \mathrm{d}\tau_1 \int_t^{\tau_1} \mathrm{d}\tau_2[H'(\tau_1), H'(\tau_2)] \tag{35}$$

Stopping at second order: $W = \Omega_1$, this integral can be evaluated using a quadrature formula:[26]

$$\Omega_1(t + \Delta t, t) \simeq -iH'\left(t + \frac{\Delta t}{2}\right)*\Delta t \tag{36}$$

This is equivalent to the well-known split-operator method.[70]

Three algorithms have been implemented to calculate the matrix exponential e^W entering eq 33. The first one is based on the diagonalization of the W matrix

$$e^W = Ue^wU^{\dagger} \text{ with } WU = wU \tag{37}$$

another is based on a Taylor expansion of the exponential

$$e^W = \sum_{n=0}^{k} \frac{1}{n!}W^n \tag{38}$$

while the third one is based on the Baker-Campbell-Hausdorff (BCH) scheme.[20,73]

$$P'(t + \Delta t) = P'(t) + \frac{1}{1!}[W, P'(t)] + \frac{1}{2!}[W, [W, P'(t)]] + \frac{1}{3!}[W, [W, [W, P'(t)]]] + \ldots \tag{39}$$

Note that the latter scheme assumes the Kohn−Sham matrix is Hermitian. Other methods based on polynomial Chebychev expansion or Krylov subspace projections have been considered by other groups to evaluate the matrix exponents.[70,74]

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

Propagation with the Magnus scheme requires the knowledge of the KS matrix at a later time $H'\left(t + \frac{\Delta t}{2}\right)$, which is unknown. Two methods have been implemented. In the iterative algorithm[70] $H'\left(t + \frac{\Delta t}{2}\right)$ is first extrapolated from the knowledge of H' at earlier times. P' is then propagated from t to $t + \Delta t$ by the Magnus propagator, and the resulting density matrix is used to build the KS potential at $t + \Delta t$. A new KS potential at time $t + \frac{\Delta t}{2}$ is interpolated from the potential at t and $t + \Delta t$. The propagation of P' is repeated with this new KS potential. The iterations are continued until convergence. This is a robust but time-consuming procedure. An alternative is the two-step predictor-corrector scheme proposed by Van Voorhis and coworkers.[26]

Application of an External Electric Field. We continue this section with the mathematical definitions of the external electric fields that can be applied in deMon2k. One option is to apply a Gaussian shaped pulse

$$\boldsymbol{F}(t) = F_{\max}\cdot\exp[(-(t - t_0)^2)/2{u_F}^2]\cos(\omega t)\hat{\boldsymbol{d}} \tag{40}$$

where t_0 is the center of the pulse, u_F is the pulse width, ω is the field pulsation, $\hat{\mathbf{d}} = (\hat{x},\hat{y},\hat{z})$ is the polarization vector, and $F_{\max}$ is the maximum field strength.

$$\boldsymbol{F}(t) = \begin{cases} \dfrac{t}{t_0}F_{\max}\cos(\omega t)\hat{\boldsymbol{d}} & (t < t_0) \\ F_{\max}\cos(\omega t)\hat{\boldsymbol{d}} & (t \geq t_0) \end{cases} \tag{41}$$

A drawback of the Gaussian shaped pulse or of the linear ramp is the possible introduction of spurious static field effects which arise if the zero pulse area condition (ZPAC, $\int_{t=0}^{t_{\max}}\boldsymbol{F}(t)dt = 0$) is not fulfilled.[75,76] Care must be taken to avoid such effects, for example by setting the center of the Gaussian pulse sufficiently far from the initial time. In fact Gaussian pulses are not convenient in practical applications because of the shallow decay of Gaussian functions which require long simulation times to ensure the ZPAC. To alleviate this inconvenience some authors proposed the use of squared sinusoidal functions.[75,77] We implemented the following one in deMon2k

$$\boldsymbol{F}(t) = F_{\max}\cdot\sin^2\left[\frac{\pi t}{T_{\text{pulse}}}\right]\cdot\theta(t)\cdot\theta(T_{\text{pulse}} - t)\cdot\cos(\omega t)\hat{\boldsymbol{d}} \tag{42}$$

where T_{pulse} is the duration of the pulse, and θ is the Heaviside function. Finally we also implemented an infinitely narrow kick in the first step of the RT-TDDFT simulation. As illustrated below kick perturbations are useful to simulate absorption spectra.

The applied field excites the molecular system through the coupling with the electrostatic dipole. The corresponding potential term is added to the KS matrix.

$$E^{\text{app}}(t) = -\boldsymbol{\mu}(\boldsymbol{t})\cdot\boldsymbol{F}(\boldsymbol{t}) \tag{43}$$

$$\boldsymbol{\mu}(\boldsymbol{t}) = \sum_A Z_A\boldsymbol{R}_A - \sum_{\sigma,\tau}P_{\sigma\tau}\langle\sigma|\boldsymbol{r}|\tau\rangle \tag{44}$$

$$H^{\text{ext}}_{\sigma\tau} \equiv \frac{\partial E^{\text{ext}}}{\partial P_{\sigma\tau}} = \langle\sigma|\boldsymbol{r}|\tau\rangle\boldsymbol{F} \tag{45}$$

Several analysis tools of the electron dynamics have been implemented. They will be introduced in the following applications sections as needed.

On the Coupling between the Electron and the Induced Dipole Intrinsic Dynamics. The full hybrid QM/MMpol Hamiltonian is given by eq 46

$$H^{\text{QM/MM}}(t) = H^{\text{QM}}(t) + H^{\text{embed}}(t) + H^{\text{MM}}(t) \tag{46}$$

where H^{QM} is given by eq 5, and H^{embed} collects the coupling between the QM and MM regions (embedding energy) which includes the interaction between the electrons and QM nuclei with the permanent and induced dipoles on MM sites. Finally $H^{\text{MM}}(t)$ is the energy of the MM part computed with molecular mechanics force field. The latter term also holds a dependence in time because of the interaction between the time-dependent MM induced dipoles and the MM permanent charges. If one is interested in solving the time-independent KS equations to determine the stationary states of the system of interest, a common procedure is to relax the MM induced dipoles at every SCF cycle. The MM dipoles are then injected in the next SCF cycle to calculate a new embedding potential. The convergence threshold for converging the MM dipole moments is tightened along with the SCF convergence to reach, at global convergence, a user-defined value, typically 10^{-8} to 10^{-10} D. On the other hand, if one is interested in the time-dependent solutions of the KS equations, more subtle algorithms are needed because of the time dependence of each term of eq 46. In principle one needs to set up the coupled equations of motion for the overall system. This is not a trivial task because of the composite quantum-classical nature of the system. One may think of coupling RT-TDDFT for the electron cloud to a fictitious dynamics of the MM induced dipoles, in the spirit of what is done for molecular dynamics simulations with MMpol.[78−80] Here we consider a simpler scheme in which we make the assumption that the MM dipoles completely relax at each RT-TDDFT step. In other words, we look for the stationary polarization state of the environment along with the nonstationary propagation of the electron cloud. For sufficiently small time steps this mixed stationary-nonstationary scheme is certainly a valid approximation to the real dynamics. As shown in the Validation section we found this approximation to be acceptable for standard electron dynamics simulations with 1 as time steps or below.

III. PERFORMANCE AND VALIDATION

III.1. RT-TDDFT Propagation with Density Fitting. As explained in the Methodology section our implementation of RT-TDDFT relies on the use of fitted densities. In particular, one has the choice in deMon2k of using either the KS or the fitted density to calculate the (time-dependent) XC energies and potentials. These two approaches are referred to as BASIS and AUXIS, respectively. Using the fitted densities usually induces a reduction of the computational cost by a factor of 10, which is clearly advantageous, yet it remains to be tested if $\tilde{\rho}$ can be used safely in RT-TDDFT propagation. As test cases, we consider two molecules: carbon monoxide and cysteine, a sulfur containing amino acid taken in the nonzwitterionic form. The propagation has been run using the Magnus propagator with an integration time step of 1 as and diagonalization (eq 37). The simulations have been carried out with the TZVP-FIP2[81] basis set and the PBE functional.[82] Auxiliary basis sets are generated by an automatic procedure implemented in deMon2k that

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

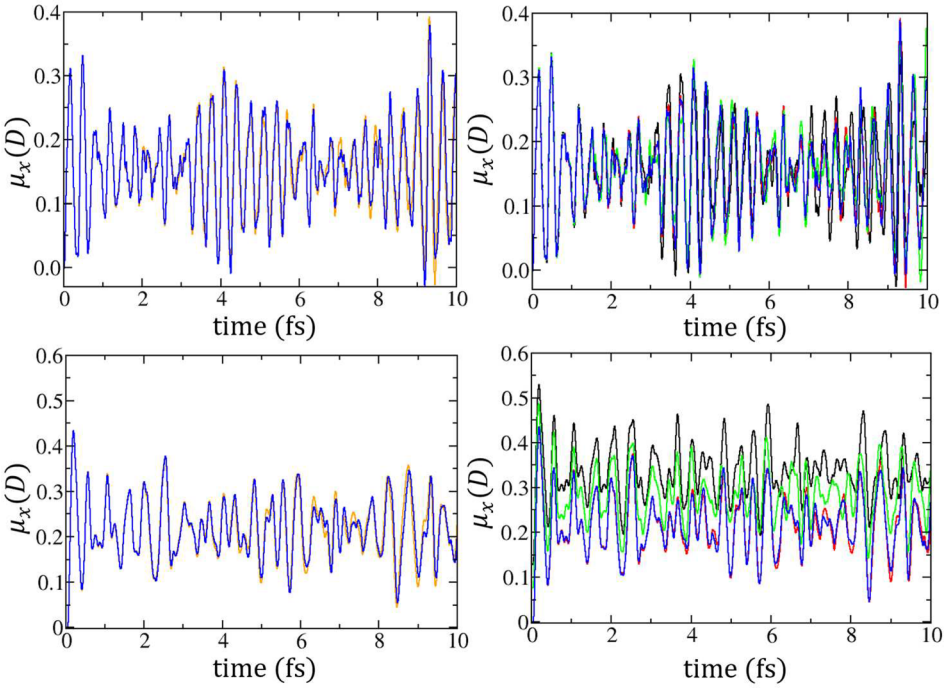


Figure 1. Influence of the accuracy of auxiliary fitted density on electron dynamics for CO_2 (top) and cysteine (bottom). The x-component of the molecular dipole moment is represented. Left: comparison of RT-TDDFT simulations using either the KS density (BASIS), in orange, or the auxiliary density (AUXIS), in blue, to integrate the XC potential in RT-TDDFT propagations. The GEN-A2* auxiliary basis set is used. Right: the simulations use the AUXIS approach with GEN-A2 (black), GEN-A3 (green), GEN-A2* (blue), and GEN-A3*(red) auxiliary basis sets.

depends on the atomic orbital basis set. The GEN-An auxiliary function sets contain groups of auxiliary functions with *s* and *spd* angular momenta. The index *n* determines the number of auxiliary function sets, i.e. the number of these sets increases with increasing *n*. We have considered the GEN-A2 and GEN-A3 auxiliary function sets, as well as the GEN-A2* and GEN-A3* that are supplemented by *f* and *g* auxiliary functions. As a general rule of thumb the larger the auxiliary basis set the more accurate the DFT-based energies and properties. An adaptive grid of accuracy 10^{-7} Ha has been used to integrate the XC potential and energies.[51] In Figure 1 we report the fluctuations of the *x*-component of the dipole moment of the two molecules when subjected to a constant electric field of intensity 0.01 au along the *z*-axis. Similar conclusions looking at the *y*- or *z*-components of the induced dipoles can be drawn.

On the left-hand side of Figure 1 we analyze the sensitivity of the propagation to the electron density used to calculate the XC potential, using the GEN-A2* auxiliary basis set. During the 10 fs of the propagation we find no important differences between the BASIS (orange) and AUXIS (blue) approaches. The simulation corresponding to the graphs on the right-hand side has been obtained with the AUXIS approach but with different auxiliary basis sets. For CO_2 all the simulations give similar electronic evolution. The electronic response of CO_2 is well captured by each auxiliary basis set. For cysteine the results are clearly more contrasted. Taking GEN-A3* as the reference auxiliary basis set, we find that the simulations with GEN-A2 and GEN-A3 are clearly different. On the other hand, with GEN-A2* the simulation nicely reproduces the evolution of the induced dipole moment. This is an encouraging result. Provided sufficiently flexible auxiliary basis sets are chosen, one can rely on density fitting techniques to reduce the computational cost of the RT-TDDFT propagations similar to what is done in stationary auxiliary DFT or auxiliary perturbation theory calculations. Our RT-TDDFT implementation thus takes advantage fully of the optimized density fitting algorithms already implemented in deMon2k. We will come back to the code performance at the end of section III.

III.2. RT-TDDFT To Calculate Static Polarizabilities. We continue the Validation section with the calculation of polarizabilities of molecules. Static and dynamic polarizabilities, as well as hyperpolarizabilities, can be calculated with standard DFT either by finite-field methods,[83,84] by the coupled-perturbed KS approach,[85,86] or by the auxiliary density perturbation theory.[87−89] These are actually recommended approaches for computing these properties at modest computational cost. Here they are used to test the validity of our RT-TDDFT module. After converging the stationary ground state of the molecule in the absence of an external field, an electric field $\boldsymbol{F}_i^{\text{ext}}$ is applied, and the response of the electron density is simulated by RT-TDDFT. The resulting induced dipole is related to the applied electric field vector via the polarizability tensor. The external electric field may either be constant or time-dependent, giving access to static or dynamic polarizabilities, respectively. Focusing here on the static case, the polarizability tensor elements are given by[83]

$$\alpha_{ij}F_j = \frac{2}{3}[\bar{\mu}_i(F_j) - \bar{\mu}_i(-F_j)] - \frac{1}{12}[\bar{\mu}_i(2F_j) - \bar{\mu}_i(-2F_j)] \tag{47}$$

where $\bar{\mu}_i(F_j)$ denotes the average contribution along $i(= x, y, z)$ of the induced dipole when an external field F_j has been applied along *j*. The uncertainty in the knowledge of $\bar{\mu}_i$ is given by $u(\bar{\mu}_i) = \sigma_i/\sqrt{NS_{\text{eff}}}$ where σ_i is the standard deviation of the sample of dipole moments calculated along the RT-TDDFT propagation. *N* is the number of terms in the sample, and S_{eff} is the statistical chain efficiency. The latter has been evaluated with the Coda package of the R project for statistical

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

Table 1. Static Polarizability Tensors (bohr^3) Computed with RT-TDDFT and with a Finite Field Method

molecule		finite field difference			RT-TDDFT		
benzene	α_{ij}	X	Y	Z	X	Y	Z
	X	83.5	0.0	0.0	83.6 ± 0.6	0.0 ± 0.0	0.0 ± 0.0
	Y	0.0	83.5	0.0	0.0 ± 0.6	83.6 ± 0.6	0.0 ± 0.0
	Z	0.0	0.0	44.5	0.0 ± 0.0	0.0 ± 0.6	44.3 ± 0.1
phenol	X	97.1	0.0	1.9	98.1 ± 0.4	0.0 ± 0.6	2.3 ± 0.0
	Y	0.0	46.8	0.0	0.0 ± 0.0	46.8 ± 0.0	0.0 ± 0.0
	Z	2.0	0.0	87.1	2.0 ± 0.1	0.0 ± 0.0	87.2 ± 0.2
cysteine	X	78.7	13.3	−8.0	78.6 ± 0.0	13.4 ± 0.0	−8.1 ± 0.0
	Y	13.3	72.6	−2.4	13.4 ± 0.0	72.4 ± 0.0	−2.4 ± 0.0
	Z	−8.0	−2.4	85.2	−8.0 ± 0.0	−2.3 ± 0.0	85.0 ± 0.0

computing.[90,91] The calculations have been done with the PBE functional,[82] the TZVP-FIP1 basis set,[81] which has been optimized for electric properties calculations, and the GEN-A2* auxiliary basis set. Fields F_j of 0.01 au have been applied. This value is also used by default in the finite field method in deMon2k.[84] The RT-TDDFT propagations have been run for 10 fs with an integration time step of 1 as. The propagator-corrector Magnus scheme with diagonalization has been used. As can be seen in Table 1 the static polarizability tensors calculated from simulation of the electron density subjected to a perturbation by RT-TDDFT match nicely those obtained by the finite field method. The agreement between both approaches validates our implementation.

III.3. Absorption Spectra in the Gas Phase. A further validation of our RT-TDDFT implementation is now sought by comparing molecular absorption spectra calculated by LR- and by RT-TDDFT. Indeed, the electronic spectrum of a molecule is encoded in the evolution of the molecular dipole simulated by RT-TDDFT after a molecule is perturbed by an infinitely narrow electric field. The Fourier transform of the dipole signal gives access to the polarizability tensor in the frequency domain which in turn yields the dipole strength function (absorption spectrum). More details about this procedure can be found in many recent publications.[19−21] Coumarin has been chosen here as a test case because it is a solvatochromic dye, the dipole moment of which is strongly modified in the excited state corresponding to the first electronic absorption band. Calculations have been done with the DZVP/GEN-A2* combination of atomic and auxiliary basis sets and with the PBE XC functional. The LR-TDDFT absorption spectrum has been simulated by assigning a Lorentzian function of width 0.25 eV centered at each excited state energy and with an amplitude proportional to the oscillator strength of the transition. The first 493 singlet excited states have been included in the construction of the spectrum, spanning an energy window 18 eV wide. The LR-TDDFT calculation has been carried out with deMon2k within the framework of Auxiliary Density Perturbation Theory (ADPT).[92] For the RT-TDDFT spectrum three 15 fs simulations have been carried out, each with a different orientation of the initial kicking electric field. The strength of the field was set to 0.005 au. A time step of 1 as was used in the propagation (using the PC Magnus propagator and the diagonalization technique). To construct the spectrum the molecular dipole was damped by an exponential function ($e^{-t/\tau}$ with τ = 180 au) to broaden the absorption peaks in the RT-TDDFT spectrum. In both the LR- and RT-TDDFT approaches fitted densities are employed to evaluate the Coulomb and the XC integrals. The spectra are depicted in Figure 2. The lowest energy transition with significant oscillator strength is found at 4.12 eV (300 nm). This excitation energy is close to the first band of the absorption spectra of coumarin in isopentane (310 nm).[93] As illustrated in Figure 3, this electronic excitation mainly corresponds to $\pi\pi^*$ transitions from the HOMO−2 and HOMO toward the LUMO with coefficients of 0.53 and 0.29, respectively.

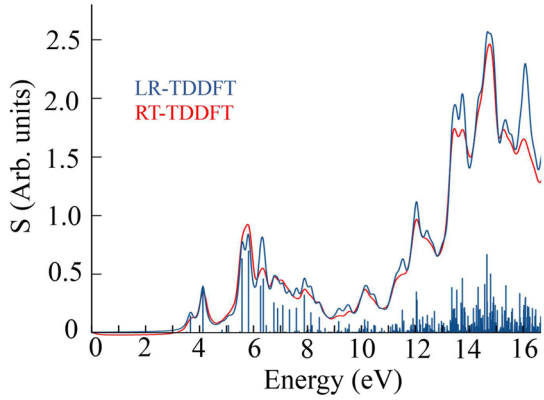


Figure 2. Absorption spectra of coumarin in the gas phase obtained from Linear-Response (blue) or real-time (red) TDDFT calculations. The LR-TDDFT excitation energies are indicated by vertical lines the heights of which are proportional to the oscillator strength of the transition. The convoluted spectra have been obtained from Lorentzian functions assigned to each excitation. Both spectra have been normalized on the absorption value at 4.12 eV.

The agreement between the LR- and RT-TDDFT spectra is excellent over the entire range of energies. Both the positions of the maximum and the relative amplitude obtained match with the two types of TDDFT implementations. Our results validate the use of the Auxiliary DFT framework[52] for constructing absorption spectra from RT-TDDFT simulations. As usual though, the choice of the auxiliary basis is critical and must include polarization functions (GEN-An*) to reach good accuracy.

III.4. QM/MMpol Framework in Stationary and Non-stationary Cases. *(DFT+Density Fitting)/MMpol: Stationary Calculations.* We now present the first results of QM/MMpol calculations with the in-deMon2k QM/MM module. To this end we consider a peptide (Tyr-Gly-Gly-Phe-Met) treated by DFT immersed in a box of 4,030 polarizable POL3[96] water molecules. The full system was previously equilibrated by classical MD simulations (data not shown). Five geometries

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

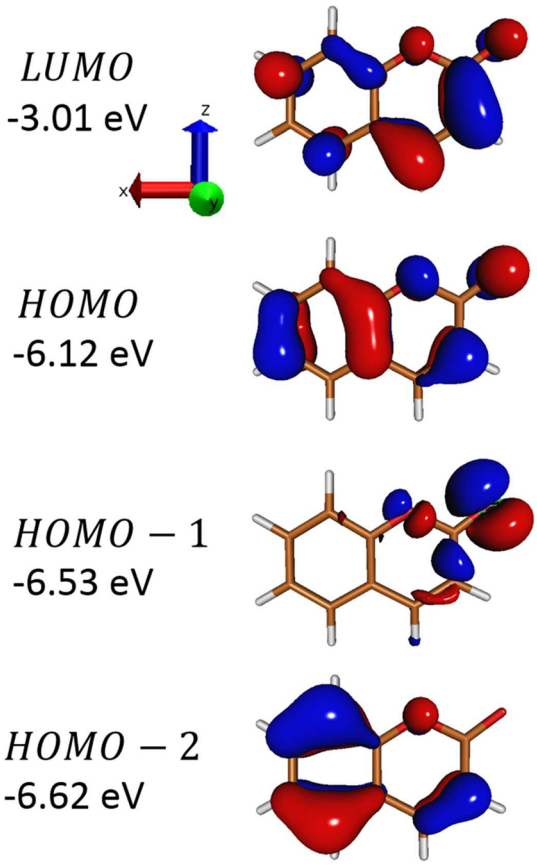


Figure 3. Frontier KS molecular orbitals for an isolated coumarin molecule calculated with the PBE XC functional. The isosurface corresponds to values of 0.05 bohr$^{-3/2}$. Picture made with gmolden[94] and VMD.[95]

extracted from the classical MD simulations have been calculated at the QM/MMpol level. For the latter we have used the PBE functional and the DZVP-GGA atomic basis set in combination with the GEN-A2, GEN-A2*, or GEN-A3* auxiliary sets. The fitted density has been used to calculate both the Coulomb and XC potentials.[52] A grid of high accuracy is used to integrate the XC contributions (10^{-7} Ha). The induced dipoles have been updated at every SCF cycle by an iterative procedure until the Root-Mean-Square between two successive cycles is below 10^{-9} D. To avoid the polarization catastrophe we have used the Tinker-exponential forms of field attenuation (eqs 22−24). For the calculation of the electric field created by the electron density (eq 10) we consider two options. The first one is to use the KS density; the other is to use the auxiliary fitted density, respectively referred to as FBASIS and FAUXIS.

Let us first consider the QM/MMpol-FBASIS scheme. The timings reported in Table 2 indicate that including induction significantly increases the cost of the calculation. It represents 65% of the total time spent in the SCF module with GEN-A2*. This is for the most part the calculation of the electric field created by the electron density on the MM sites that leads to this timing. Thirty seconds are needed on average for each SCF cycle. For comparison, a nonpolarizable DFT/MM calculation on the same system requires 12 s per SCF cycle. There is indeed a clear cost to the inclusion of induction. When switching to the less accurate GEN-A2 auxiliary set, the computational cost is decreased to 21 s/SCF cycle, but more SCF cycles are needed to reach convergence so that 85% of the time is spent in the induction modules. In an attempt to reduce this supplementary cost, we consider the QM/MMpol-FAUXIS scheme. With GEN-A2* the cost drops to 16.1 s per SCF cycle, which is a just a little more than a nonpolarizable DFT/MM (12 s). This reduction comes from the calculation of the electric field created by the electron density that has dropped to almost zero. The remaining time for induction is spent in the induced dipole iterations and in the calculation of the induced-dipole electrostatic potential. To assess whether the FAUXIS is reliable we report in Table 2 the different contributions to the polarization energy (eq 11). For each contribution, the Root-Mean-Square-Deviations are calculated taking the FBASIS/GEN-A2* as reference. At SCF convergence we find that the RMSD of induction energies ($E^{\mathrm{ind}}_{\mu-\mathrm{ZQM}} + E^{\mathrm{ind}}_{\mu-\rho} + E^{\mathrm{ind}}_{\mu-\mathrm{qMM}}$) differs by less than 1.5 kcal/mol between the FBASIS and FAUXIS approaches with GEN-A2*. A similar value is found for FAUXIS/GEN-A3*. We can thus consider the FAUXIS/GEN-A2* combination to be an excellent approximation of FBASIS/GEN-A2*, without further need to go to GEN-A3*. On the other hand, the comparison is less encouraging with GEN-A2. In this case, the RMSD of the induction energy between the

Table 2. Induction Energies and Timings of Stationary QM/MMpol Calculations

	QM/MMpol (FBASIS)		QM/MMpol (FAUXIS)		
	GEN-A2	GEN-A2*	GEN-A2	GEN-A2*	GEN-A3*
induction energies (kcal/mol)[a]					
RMSD($E^{\mathrm{ind}}_{\mu-\mathrm{qMM}}$)	0.03	0.00	0.04	0.02	0.02
RMSD($E^{\mathrm{ind}}_{\mu-\mathrm{ZQM}}$)	2.74	0.00	26.94	1.49	1.56
RMSD($E^{\mathrm{ind}}_{\mu-\rho}$)	6.35	0.00	55.61	2.97	3.06
RMSD($E^{\mathrm{ind}}_{\mathrm{tot}}$)	3.64	0.00	28.70	1.46	1.50
timings (s)					
QM electric field	1060 (61%)	657 (43%)	3 (≈0%)	3 (≈0%)	5 (≈0%)
MM dipole iterations	236 (14%)	196 (13%)	246 (36%)	163 (18%)	154 (13%)
dipole embedding	167 (10%)	133 (8.8%)	169 (25%)	106 (12%)	121 (10%)
total induction	1463 (85%)	986 (65%)	418 (61%)	272 (30%)	280 (23%)
SCF (s/cycle)[b]	20.9	29.6	8.4	16.1	20.0

[a]We report the RMSDs over five conformers of the peptide taking the FBASIS/GEN-A2* as the reference method. [b]Total time spent in SCF divided by the number of SCF cycles to reach convergence. For the timing, the numbers within brackets represent the percentage of the total time spent to compute the given contribution. Calculations have been performed with Intel Xeon E5649 (2.53 GHz) 6 core CPUs with 24 GB RAM per core. Jobs were run in parallel with 48 processors.

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

FAUXIS and FBASIS approaches is 29 kcal/mol. This is a rather large value: the fitted density can be used safely *in lieu* of the KS density in QM/MMpol calculations as long as sufficiently flexible auxiliary basis sets are used (GEN-A2* or larger). In such cases the computational advantage of the FAUXIS approach is significant. We make two final remarks. First, rather expectedly, the induction terms that are the most sensitive to the choice of the auxiliary basis are those between the MM induced dipoles and the QM region. The induction energy between the dipoles and the MM charges ($E^{ind}_{\mu-qMM}$) is always the same. Second, we note that the SCF process converges more rapidly with larger auxiliary basis sets, somewhat lowering the increase in computational cost due to the greater number of integrals to compute.

(DFT+Density Fitting)/MMpol: Nonstationary Calculations. We now examine the sensitivity of the RT-TDDFT/MMpol simulation to the method chosen to calculate the electric field generated by the QM region. After SCF convergence the electronic density of the peptide is perturbed by a Gaussian shaped electric field centered at 10 as and with standard deviation 1 as. Three field strengths have been tested: 0.001, 0.01, and 0.1 au. The simulations have been run for 1 fs with the propagator-corrector Magnus scheme and a time-step of 1 as. The BCH formula (eq 39) has been used to calculate the exponential of the complex matrices. Thirty terms have been used in the expansion of eq 39. We report in Figure 4 the evolution of the peptide dipole moment (top) and the difference of polarization energy with respect to their values at the initial time. Both the FAUXIS and FBASIS schemes are tested. Each simulation has been carried out with either the GEN-A2 or GEN-A2* auxiliary basis set. As expected the stronger the intensity of the electric field perturbing the electron density at the beginning of the propagation, the larger the response of the system. This can be seen on the evolution of the peptide dipole moment, that exhibits larger amplitudes with the stronger field (0.1 au), and also on the fluctuations of the polarization energy. With the weakest field the subsequent variations of the polarization energy are small (of the order of 10^{-2} kcal/mol). The agreement between the FAUXIS and FBASIS approaches is always very satisfactory. With the GEN-A2* auxiliary basis set the results between the FAUXIS and FBASIS simulations are even indistinguishable whatever the strength of the electric pulse that perturbs the system.

Coupling between Electron and MM Dipoles in RT-TDDFT/MMpol Simulations. As explained in the Methodology section a central aspect of the present RT-TDDFT/MMpol implementation is the assumption that the MM induced dipoles respond instantaneously to the electronic motion taking place in the QM region. Technically, this means that the MM dipoles are fully converged at every RT-TDDFT step. We tested the suitability of this strategy by repeating the previous simulations of the solvated peptide with shorter time steps of 0.75, 0.5, 0.25, and 0.1 as. For short enough time steps the decoupling approximation is certainly valid. With a time step of 0.1 as we found, indeed, that the MM induced dipoles evolve very smoothly. Figure 5 depicts the differences of QM/MMpol total energy, of polarization energy, and of embedding energy as a function of time taking the 0.1 as time-step simulation as reference. The initial perturbing field was set to 0.001 au. Clearly, the larger the time step the larger the difference. For the total energy and for the embedding energy the maximum error is of the order of a few thousandths of a kcal/mol with a 1 as time step. It is an order of magnitude smaller for the

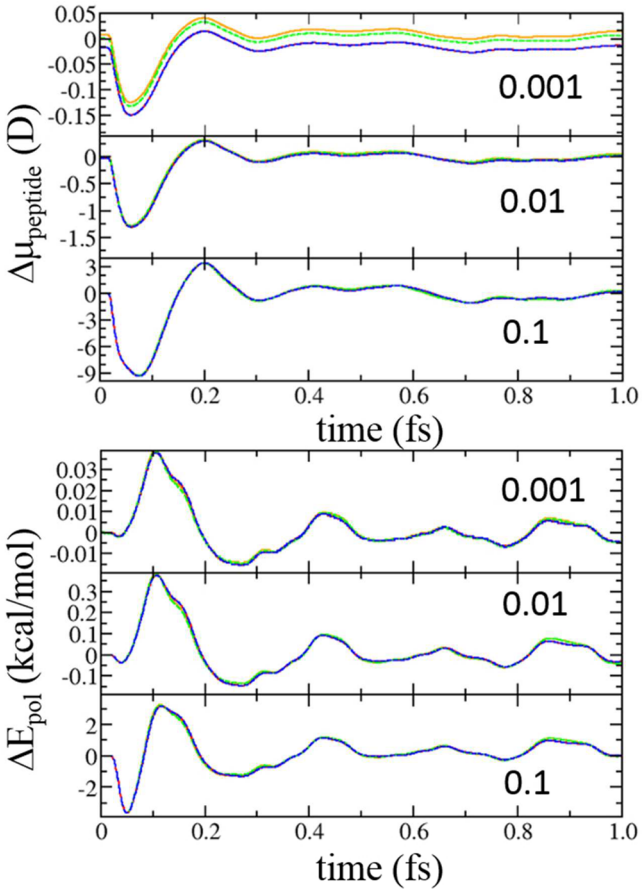


Figure 4. Evolution of the peptide dipole moment (top) and of the polarization energy of POL3 water molecules (bottom) after an initial perturbation of the peptide electron density by a short electric pulse. Three intensities of the pulse have been applied: 0.001, 0.01, and 0.1 au. Simulations have been run with two types of auxiliary basis sets (GEN-A2 and GEN-A2*) to calculate the Coulomb and XC contribution to the KS potential. In addition to calculating the electric field created by the QM region, that contributes to determine the MM induced dipoles, we used either the KS density (FBASIS, dashed lines) or the auxiliary density (FAUXIS, full line). Blue: FBASIS/GEN-A2*, red: FAUXIS/GEN-A2*, green: FBASIS/GEN-A2, and orange: FAUXIS/GEN-A2.

polarization energy. These values are rather small compared to the variations of the total energy in these simulations caused by the initial perturbation with the external electric field (around 0.06 kcal/mol). Interestingly the energy errors fluctuate around zero. This suggests that the simulations with time steps larger than 1 as eventually depart from the reference trajectory but do not diverge from it. We also found that the peptide dipole as well as the water dipoles of the first solvation layer (i.e., those mainly impacted by the electron dynamics taking place on the peptide) were within $1.0E^{-5}$ D from those of the reference trajectory. This is a very small value. In simulations in which the initial perturbing electric field strength was increased to 0.01 au the same trends are obtained albeit with a factor of 10 in the amplitudes of the errors (Figure S1). This again seems acceptable in view of the overall total energy change (6 kcal/mol). For an even stronger perturbing field (0.1 au) the errors in total, polarization energy, and embedding energies are of the order of a kcal/mol, a tenth of a kcal/mol, and a hundredth of a kcal/mol, respectively (Figure S2). These values are quite high but again much smaller than the fluctuations of the total energy

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

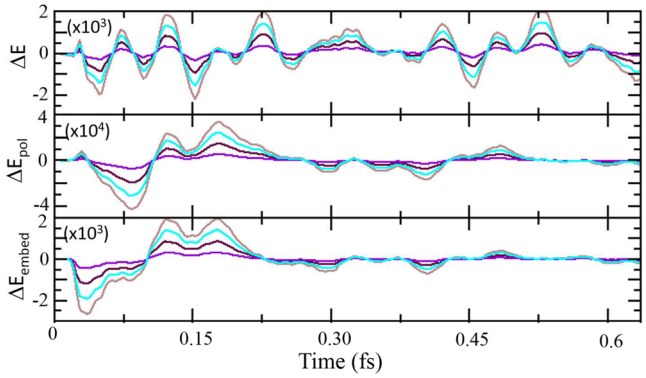


Figure 5. Error estimations of the MMpol-stationary-RT-TDDFT scheme. The 0.1 as time-step simulation is taken as a reference, and the data for larger time steps are given with respect to the reference: light brown, 1 as; cyan, 0.75 as; maroon, 0.5 as; violet, 0.25 as. All the energies are given in kcal/mol scaled by coefficients given in each graph. ΔE is the total QM/MM energy, ΔE_{pol} is the total polarization energy, and ΔE_{embed} is the embedding energy (permanent charges and induced dipole).

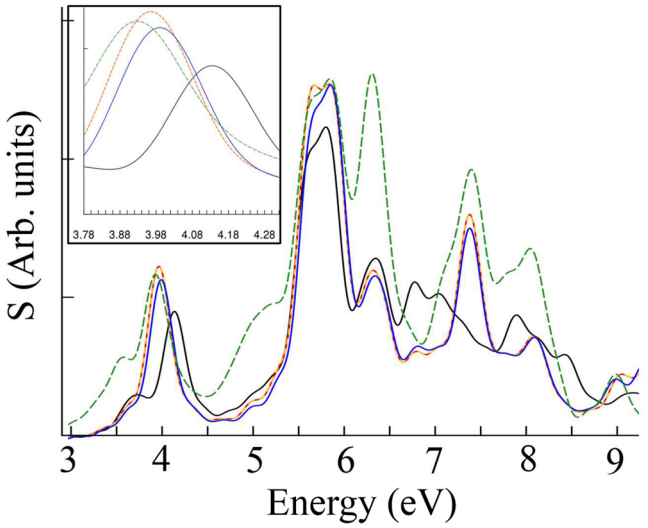


Figure 6. Absorption spectra of coumarin in the gas phase (black) and in the condensed phase (other curves) calculated from RT-TDDFT. Blue: spectrum obtained with the TIP3P model for water (nonpolarizable), yellow and red: spectra obtained with the polarizable POL3 water model and within the FAUXIS/GEN-A2* and FBASIS/GEN-A2* schemes, respectively, green and dashed: reference spectrum RT-TDDFT/MMpol (POL3) for which both the coumarin molecule and its first hydration level were treated at the quantum level. In the latter case we extracted the Hirshfeld intrinsic dipole of coumarin to remove absorption bands due to the QM water molecules. Inset: zoom on the 3.28−4.78 eV range.

of the molecule (around 50 kcal/mol). We note that such electric fields are extremely strong and would trigger nonlinear effects like ionization. Altogether these tests justify the nonstationary/stationary coupling scheme between RT-TDDFT for the QM part and stationary MMpol for the environment although one should be careful to adapt the propagation time step to the amplitude of the electronic fluctuations that take place in the QM region. The most suitable time step might depend on the particular system of interest.

In principle, though, there should be a time step beyond which the decoupling between electrons and MM dipoles ceases to be valid. When we increased the time step (2 or 5 as), the electronic propagation was not stable anymore and diverged in a few steps. RT-TDDFT propagations are usually very sensitive to discontinuities that may arise in the time-dependent KS potential. Hence, a plausible explanation for the numerical instabilities observed in RT-TDDFT/MMpol simulations for the largest time steps may be that they arise from potential discontinuities caused by significant variations of MM induced dipoles between two propagation steps. Interestingly, sudden instability of electron dynamics propagation may thus well be a sign of the breakdown of the decoupling hypothesis between the electron cloud dynamics and induced MM dipoles. Further work will be needed to examine this point in more detail.

III.5. Absorption Spectra from RT-TDDFT/MMpol Simulations. We report in Figure 6 the absorption spectrum for the coumarin molecule solvated by a 30 Å radius sphere of water molecules. The system was equilibrated in a previous step by a classical MD simulation (data not shown) followed by a few-steps geometry optimization of the coumarin molecule before the RT-TDDFT/MMpol simulations. This partial optimization was intended to avoid too large distortions of the molecule, which would cause unreasonable displacements of the electronic absorption bands. We followed a similar protocol as for the gas phase case (see section III.3) to build the spectra. In most simulations the QM region encompasses the coumarin molecule, while the water environment is described by the force field. However, in order to define a reference spectrum we have also carried out RT-TDDFT/MMpol simulations including the first hydration layer of coumarin in the QM region (12 water molecules).

As seen in Figure 6 the environment has a significant effect on the spectrum. This is especially noticeable for the band around 4 eV that corresponds to a transition having charge transfer character. The center of this band is red-sifted by 0.22 eV when comparing the gas phase and the reference spectra (see inset). The nonpolarizable TIP3P model leads to a red-shift of 0.15 eV, while the polarizable RT-TDDFT/MMpol (POL3) spectrum exhibits a more pronounced red-shift of 0.18 eV, that is closer to the reference. Electrostatic induction thus permits a slight improvement on the position of the most displaced absorption bands. We also remark that the absorption spectra calculated with polarizable water combined with either the FBASIS or the FAUXIS option (see above) are indistinguishable. This is a further element showing that as long as sufficiently flexible auxiliary basis sets are chosen (GEN-A2* here), density fitting techniques can be safely used in TDDFT/MMpol calculations.

III.6. Numerical Stability and Performances. We conclude section III with some notes on the numerical stabilities of RT-TDDFT simulations and on the computational performance of the implementation in deMon2k. It is known that simulating electron dynamics by RT-TDDFT can be difficult in terms of numerical stability. This numerical stability is highly system-dependent, and the best propagation scheme has to be sought for each new molecular system of interest. The list of parameters impacting the numerical stability encompasses not only those directly related to the propagation schemes (propagator, matrix exponentiation method, size of time-step, ...) but also other more general DFT parameters: the quality of atomic basis set, of auxiliary basis set, the grid quality for XC potential numerical integration, ... In particular, we

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

found the initial conditions of the electronic propagation to be extremely important. For instance, in most of our calculations the preliminary stationary SCF needs to be converged with tolerance criteria below 10^{-10} Ha for the total electronic energy and 10^{-7} for the charge density error (respectively defined by the TOL and CDF options of the SCFTYPE keyword in deMon2k). These convergence thresholds are tight compared to those customarily used in stationary DFT calculations. We also found the method for determining the fitted density coefficients to be important. To solve the sets of inhomogeneous systems of linear equations associated with density fitting, one can either use an analytical[97] or numerical approach.[98] The former generally led to more stable RT-TDDFT simulations and has been used throughout.

In Figure 7 we report the computational timing to carry out a 1.5 fs RT-TDDFT/MMpol electron dynamics simulation on the solvated peptide just described. Following our previous conclusions, this simulation has been run with the AUXIS and FAUXIS approaches with the DZVP-GGA/GEN-A2* combination of basis sets. We have used the predictor-corrector Magnus propagator with the BCH expansion (30 terms) and a time step of 1 as. A grid of accuracy 10^{-7} Ha has been used for the XC contributions. The time-dependent auxiliary density was integrated at every RT-TDDFT step to extract intrinsic atom multipoles (charges, dipole, quadrupole) according to the Hirshfeld scheme. With GEN-A2* multipoles extracted from $\tilde{\rho}$ are very close to those extracted from ρ.[69] The simulation took 12 h on 48 processors with the message passing interface protocol.

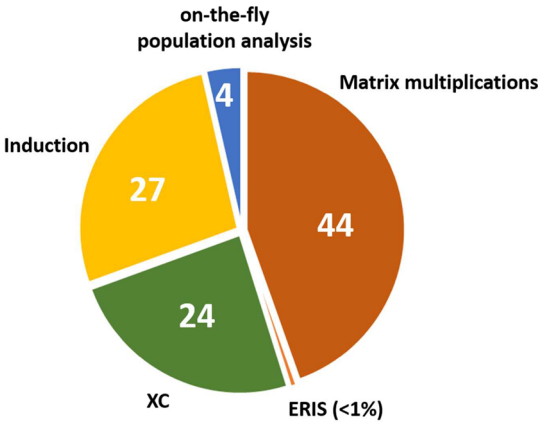


Figure 7. Computational timing for 1.5 fs of RT-TDDFT/MMpol electron simulations. Only the most time-consuming tasks are indicated. The numbers are the percentages of time spent in each type of task.

The most time-consuming part of the simulation corresponds to the matrix multiplications (of which there are almost 400 000). These are needed (i) to transform the density and KS matrices between the MO and AO representation (ii) in the BCH approximation that involves nested commutators. In second position is the cost of including induction stemming from the embedding of the QM region by induced dipoles (8%) and by the convergence of MM induced dipoles by the iterative procedure (18%). This situation could be improved in the future by adopting more advanced simulation algorithms that avoid the iterative procedure used here to converge the MM induced dipoles.[99] The XC potential represents 24% of the overall computational cost, while the calculation of the Coulomb contribution is almost negligible. This remarkable performance is possible thanks to the mixed scheme implemented in deMon2k to store short-range ERIS in RAM and to compute long-range ERIS by double asymptotic expansions.[100] Finally we note that the repeated analysis of the time-dependent electron density induced only a small supplementary cost to the calculation.[69] In summary, we think our RT-TDDFT/MMpol is efficient enough to tackle the simulation of electron dynamics in large molecular systems.

IV. ELECTRONIC POLARIZATION DYNAMICS

IV.1. Electron Dynamics within an Isolated Molecule.

In this subsection we analyze the electron dynamics taking place when a coumarin molecule in the gas phase is submitted to a monochromatic laser field. We choose a wavelength that corresponds to an excitation energy of 4.1547 eV (see section III.3). We have propagated the electron density for 25 fs with a time step of 10 as. A Gaussian shaped electric field centered at 9.5 fs and full-width at half-maximum of 3.53 fs (u_F = 1.5 fs) was applied (eq 40). The electric field orientation was set in the molecular plane with an intensity of 10^{-4} au (51 mV/nm). The frequency ω of the field was set to correspond to an excitation energy of ΔE = 0.1528 Ha. To follow the evolution of the electron density we have computed the molecular time-dependent dipole moment and the time-dependent intrinsic atomic charges q_i and dipole moments ($\boldsymbol{\mu}_X$). To define the latter three distinct population schemes have been tested, namely the Hirshfeld,[101] Becke,[102] and Voronoi Deformation

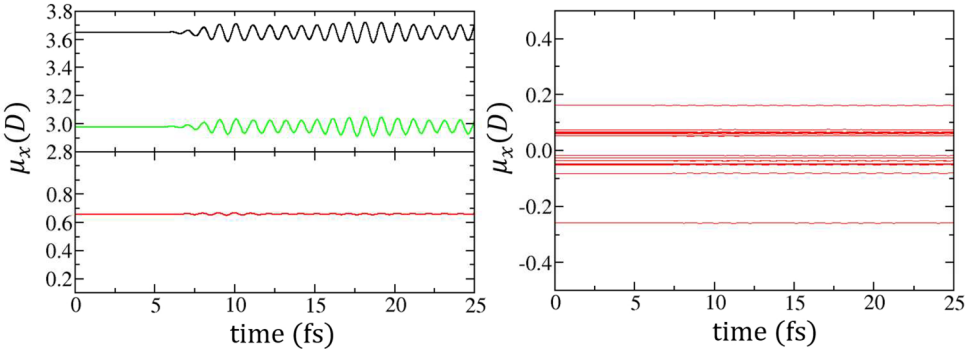


Figure 8. Evolution of dipole moments (x component) during RT-TDDFT propagation with the Hirshfeld scheme. Left: the molecular dipole moment is shown in black, the atomic-charges-derived dipole is shown in green, and the sum of intrinsic atomic dipole moments is shown in red. Right: intrinsic atomic dipole moments are shown in red for each coumarin atom.

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

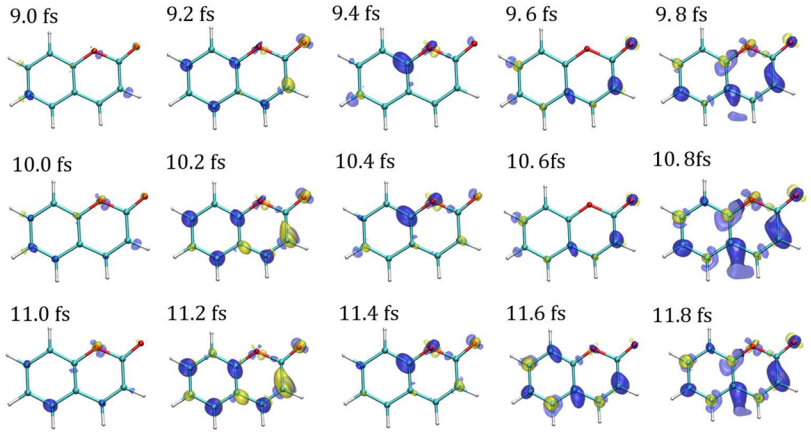


Figure 9. Time-dependent deformation of the electron density taking the initial time as reference. The isosurface corresponds to values of 10^{-5} bohr^{-3} in absolute value. The yellow surfaces correspond to positive values (accumulation of electron density), while the violet ones correspond to negative values (depletion of electron density). Each line corresponds to one Rabi oscillation period.

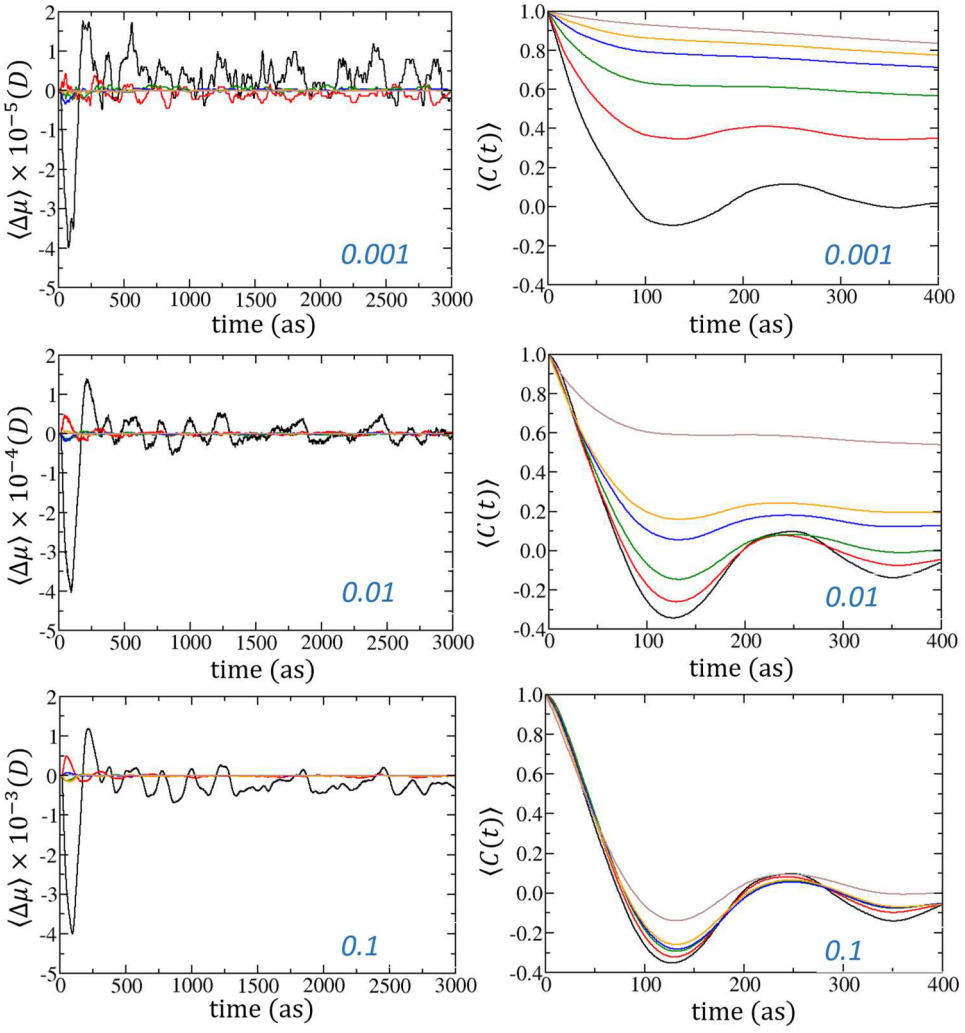


Figure 10. Left: variations of the average induced dipole moment by hydration layers $\langle \Delta\mu(t) \rangle = \langle \mu(t) - \mu(0) \rangle$. Note the change of scales for each graph. Right: normalized autocorrelation function of the water molecule induced dipole moments averaged by hydration layers for three values of the electric field affecting the peptide at the beginning of the RT-TDDFT/MMpol simulation. Color code: water molecules between 0 and 3 Å in black, 3−6 Å in red, 6−9 Å in green, 9−12 Å in blue, 12−15 Å in orange, and beyond 15 Å in brown. The uncertainties on the curves are around 0.06.

Density[103] schemes. Only the Hirshfeld results are shown in the main text, while the others are given in the SI. Because repetitive population analyses along the RT-TDDFT simulations may become time-consuming, we have extracted atomic charges and dipoles from the analysis of the auxiliary density ($\tilde{\rho}$) which is cheaper than from the KS density.[69] We have

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

previously shown this alternative to be reliable as soon as the GEN-A2* auxiliary basis (or larger) is chosen, as is the case in this application.

The set of time-dependent atomic charges (q_X) permits the definition of a total charge-derived dipole moment $\boldsymbol{\mu}^q = \sum_X q_X \boldsymbol{r}_X$. The variations in time of $\boldsymbol{\mu}^q$ provide information on charge transfer between atoms in the course of the simulation. On the other hand, the sum of the intrinsic atomic dipoles $\boldsymbol{\mu}^{\mathrm{pol}} = \sum_X \boldsymbol{\mu}_X$ provides information on the internal polarization of the atoms during the simulation. The sum of $\boldsymbol{\mu}^q$ and $\boldsymbol{\mu}^{\mathrm{pol}}$ gives rise to the full molecular dipole moment. As seen in Figure 8 the molecular dipole starts to oscillate at around 7 fs as a result of the application of the electric field. Regular oscillations are seen in subsequent times, the period of which (0.99 fs) corresponds well to the applied electric field energy (4.1547 eV). The molecular system thus undergoes Rabi oscillations between the ground and the targeted excited state. We observe soft beatings on a few fs that correspond to the Gaussian shaped pulse. Looking at the decomposition of the dipole moment we find that the oscillations are essentially due to $\boldsymbol{\mu}^q$, i.e. they are caused by charge transfers between atoms. The sum of the internal polarization $\boldsymbol{\mu}_X$ ($\boldsymbol{\mu}^{\mathrm{pol}}$) fluctuates much less, as well as each intrinsic dipole moment (Figure 8, right). Of course, the separation between $\boldsymbol{\mu}^q$ and $\boldsymbol{\mu}^{\mathrm{pol}}$ is arbitrary and depends on the chosen population scheme. With Voronoi Deformation Density, we found very similar results as for Hirshfeld (Figure S3). On the other hand, with the Becke partitioning scheme a very different separation between polarization and charge transfer is obtained (Figure S4). Actually the Becke scheme is not recommended for extracting electrostatic multipoles because it may produce "nonchemical" charges in many cases (e.g. hydrogens always have charges around −0.5e). It should probably be avoided for analyzing time-dependent electron density. That said, we found fluctuations of the total dipole moment come from fluctuations of the charge-derived dipole moments whatever the chosen partitioning scheme.

The isosurfaces of the deformation electron density shown in Figure 9 render a pictorial view of ultrafast dynamics over three Rabi oscillations. For this simple excitation process involving two electronic states the overall shape of the deformation density can be rationalized looking at the shape of the MOs involved in the process. Especially the population of the LUMO can be identified looking at Figure 9.

IV.2. Dynamics of the Response of the Environment. In this section, we analyze the electronic response of the environment of a central molecule after perturbation of the latter by an external electric field. To this end we take the same system as in subsection III.4, namely a methionine enkephalin solvated in a box of POL3 water molecules. After tight SCF convergence, the central peptide was perturbed by a Gaussian shaped electric pulse centered at 20 as with 3 as width. The field strength was set to either 0.001, 0.01, or 0.1 au. Note the latter corresponds to a very strong intensity. Following our conclusions from the Validation section we use the AUXIS and FAUXIS approaches in combination with the GEN-A2* auxiliary basis. The simulations were conducted for 3 fs with a time step of 3 as using the predictor-corrector-Magnus/BCH propagator. We report in Figure 10 the variation of the induced dipoles on MM atoms with respect to the initial time ($\Delta\mu(t) = \mu(t) - \mu(0)$) and their normalized autocorrelation functions ($C(t)$, ACF). Both quantities are averaged by hydration layers as indicated by the angular brackets $\langle ... \rangle$.

We start by considering the upper graphs that correspond to perturbing field strength of 0.001 au. As expected, the longer the distance between the water molecules and the peptide, the smaller the impact on the induced dipoles. The first hydration layer is the one that experiences the highest variations of induced dipoles. As evident from the black curve in Figure 10, top-left, the average induced dipoles undergo damped oscillations. These are caused by energy dissipation in the MM environment, which is possible thanks to the use of a polarizable FF. Dissipation is very pronounced for the first hydration layer, but it is also seen for the outer hydration layers. The induced dipoles for molecules pertaining to the inner hydration layer completely lose correlation within a few tens of as, while beyond 15 Å, the average ACF remains close to 0.8 at 200 as. The characteristic response time is distance dependent. This characteristic time increases for each successive hydration shell. Some of the averaged ACF become negative which is to be related to the oscillatory nature of the variations of $\langle \Delta\mu \rangle$. We finally remark that the response of MM induced dipoles not only is fully instantaneous but also exhibits relaxation components over hundreds of attoseconds.

When the strength of the initial perturbing field is increased to 0.01 au, the amplitude of oscillation of the average induced dipoles is larger by a factor of 10. This is true for each hydration shell. When the field strength is further increased to 0.1 au, a further increase of response amplitudes is observed for $\langle \Delta\mu \rangle$. The ACFs exhibit more complex evolutions with the increase of perturbing field strength. For the weakest perturbing field strength (0.001 au) we already mentioned that the response was distance dependent (top-right). For a perturbing field strength of 0.01 au the response of induced dipoles is not distance dependent within the first 50 as, apart from water molecules situated beyond 15 Å (middle-right). Only after this time a scattering of the average ACFs becomes apparent. Finally, for a perturbing field strength of 0.1 au all the average ACFs but one (again for water molecules situated beyond 15 Å) are almost superimposed (bottom-right). The response mechanism of MM induced dipoles is therefore not distance dependent at all within 15 Å. All these results reflect subtle response mechanisms that deserve a more detailed analysis.

We recall that the induced dipoles are determined by the electric field created by the other MM atoms ($\boldsymbol{F}_i^{(0)} + \boldsymbol{F}_i^{\mathrm{ind}}$) and by the QM region ($\boldsymbol{F}_i^{\mathrm{QM}}$) (eqs 6−10). In the present RT-TDDFT simulations only $\boldsymbol{F}_i^{\mathrm{ind}}$ and $\boldsymbol{F}_i^{\mathrm{QM}}$ can account for the variations of the MM induced dipoles since the nuclei are fixed. We also recall that we employ here a mixed nonstationary/stationary RT-TDDFT/MMpol scheme (see section III). Accordingly the response of the MM induced dipoles caused by variations of $\boldsymbol{F}_i^{\mathrm{QM}}$ is expected to be enhanced in our scheme compared to what it would be in fully dynamical simulations. Nonetheless, the average $\langle \Delta\mu(t) \rangle$ and the associated ACF extracted from a 0.1 as time-step simulation were found to be very similar to the graphs shown in Figure 10, thereby indicating that the artificial enhancement of the dipole relaxation due to the RT-TDDFT/MMpol coupling scheme is moderate. The response of MM dipoles should be less pronounced as the distance r increases because of the decay of $\boldsymbol{F}_i^{\mathrm{QM}}$ with distance. On the other hand, the response caused by variation of $\boldsymbol{F}_i^{\mathrm{ind}}$ should be associated with a certain delay, because it requires the other induced dipoles to be affected. For example, the induced dipoles of water molecules situated between 6 and 9 Å will be affected by induced dipoles of innermost hydration waters only when their induced dipoles

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

have varied. The prevalence of one mechanism over the other should depend on the relative strength of $\boldsymbol{F}_i^{\mathrm{ind}}$ and $\boldsymbol{F}_i^{\mathrm{QM}}$.

For the inner hydration layer (<3 Å) $C(t)$ is almost identical whatever the initial perturbation field. It decays to 0.3 in around 50 as (although the variations of induced dipole amplitudes are different for each perturbing electric field). For this hydration layer the source of variation of MM induced dipoles is primarily $\boldsymbol{F}_i^{\mathrm{QM}}$. It is the time-dependent field created by the electron cloud of the peptide that determines the response of MM induced dipoles. The oscillations of the MM induced dipoles essentially follow that of the peptide dipole moment (Figure S5). For the outer hydration layers, the response mechanism depends on the relative importance of $\boldsymbol{F}_i^{\mathrm{ind}}$ and $\boldsymbol{F}_i^{\mathrm{QM}}$, the latter being itself dependent on the perturbing field strength. For the strongest perturbing field (0.1 au) the response mechanism of the MM induced dipoles is completely imposed by $\boldsymbol{F}_i^{\mathrm{QM}}$ for all hydration shells (except for water molecules beyond 15 Å). This explains why the average ACFs are almost superimposed. The amplitude of the response decays with distance, but the speed at which the induced dipoles vary is the same. In this regime $\boldsymbol{F}_i^{\mathrm{QM}} \gg \boldsymbol{F}_i^{\mathrm{ind}}$ so that $\boldsymbol{F}_i^{\mathrm{QM}}$ imposes the response mechanism: the MM induced dipoles within 15 Å follow the variations of the peptide dipole moment. For the intermediate perturbing field (0.01 au) $\boldsymbol{F}_i^{\mathrm{QM}}$ dominates the response mechanism for the shorter distances (<9 Å) and for short times (<50 as). At longer distances or after a certain time, 50 as here, $\boldsymbol{F}_i^{\mathrm{ind}}$ becomes more important and starts to introduce a distance dependence in the response delay of the MM induced dipoles.

We finally remark that including nuclear motion in the simulation protocol would further make the response mechanisms more complex by allowing $\boldsymbol{F}_i^{(0)}$ to be time-dependent. This would enable simulating the reorientation polarization of the environment molecule. We leave this possibility for future work.

V. CONCLUSION

In this paper, we have reported the implementation in deMon2k of an original combination of RT-TDDFT and a polarizable force field based on the point charge-dipole model of induction. Our implementation is made efficient thanks to the use of fitted densities. Although we have focused here on applications combining both methods, RT-TDDFT or QM/MMpol calculations can of course be run independently. Electron dynamics in complex molecular systems like those encountered in biology are now accessible. Applications will be reported in the future.

Our simulations have revealed the complexity of response mechanisms of the environment of a solute submitted to an external perturbation. We have focused on electronic induction, which corresponds to the optical dielectric constant. Inclusion of nuclear motion would further complexify even more of these mechanisms by making $\boldsymbol{F}_i^{(0)}$ nonconstant. In general cases, the response mechanisms will probably be strongly system and process dependent, but we expect the RT-TDDFT/MMpol scheme introduced in the present work to be capable of covering wide ranges of processes involving significant electron dynamics.

The present work may be extended along various lines. Improvements of the QM/MMpol engine could be sought to reduce the cost of induction, or we could couple RT-TDDFT to more advanced polarizable force fields like AMOEBA. Recent results in the context of linear-response TDDFT coupled to AMOEBA are promising.[104] The RT-TDDFT propagation itself could be made more efficient, for example by using adaptive time-steps.[105] An obvious extension would be to couple RT-TDDFT/MMpol to molecular dynamics in the so-called Ehrenfest scheme to simulate coupled electron−nuclear dynamics. Simulations based on the exact factorization of the electronic-nuclear wave function as introduced by Gross and co-workers may also constitute an interesting road to follow.[106] Introducing nuclear dynamics would open new energy dissipation channels that are not included in the present schemes. Efforts along the aforementioned lines are underway in our laboratories.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00251.

> Error estimates of the nonstationary/stationary RT-TDDFT/MMpol scheme with initial perturbing fields of 0.01 and 0.1 au time-dependent atomic charges and dipole moment of the coumarin molecule (section IV.1) with the Becke and VDD partition scheme; correlation between the time-dependent dipole moment on metenkephalin and the average induced dipole moment over each hydration layer (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: aurelien.de-la-lande@u-psud.fr.

ORCID

Carine Clavaguéra: 0000-0001-5531-2333

Aurélien de la Lande: 0000-0003-0745-4171

Funding

We are grateful to Compute Canada/WestGrid for generous computational resources. This work was performed using HPC resources from GENCI [CCRT/CINES/IDRIS] (Grant 2015-2016, project number x2016076913). This work was supported by the ANR programs *Investissements d'Avenir* of LabEx PALM (Physique, Atome, Lumière, Matière) (Grant ANR-10-LABX-0039-PALM, PALM projects PHOTOHEME and DYNA-TOP). D.R.S. is grateful to NSERC/Canada for ongoing Discovery Grants.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Javier Carmona-Espíndola, Andreas Köster, and Gerald Geudtner for providing us with a developer version of deMon2k implementing LR-TDDFT in the framework of Auxiliary Density Perturbation Theory. We kindly thank Niloufar Shafizadeh, Michèle Desouter, Benoit Soep, and David Lauvergnat for numerous illuminating discussions in the course of this project.

■ REFERENCES

(1) Cederbaum, L. S.; Zobeley, J. Ultrafast Charge Migration by Electron Correlation. *Chem. Phys. Lett.* **1999**, *307*, 205−210.

(2) Cederbaum, L. S.; Zobeley, J.; Tarantelli, F. Giant Intermolecular Decay and Fragmentation of Clusters. *Phys. Rev. Lett.* **1997**, *79*, 4778−4781.

(3) Calegari, F.; Ayuso, D.; Trabattoni, A.; Belshaw, L.; De Camillis, S.; Anumula, S.; Frassetto, F.; Poletto, L.; Palacios, A.; Decleva, P.;

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

Greenwood, J. B.; Martín, F.; Nisoli, M. Ultrafast Electron Dynamics in Phenylalanine Initiated by Attosecond Pulses. *Science* **2014**, *346*, 336−339.
(4) Thürmer, S.; Ončák, M.; Ottosson, N.; Seidel, R.; Hergenhahn, U.; Bradforth, S. E.; Slavíček, P.; Winter, B. On the nature and origin of dicationic, charge-separated species formed in liquid water on X-ray irradiation. *Nat. Chem.* **2013**, *5*, 590−596.
(5) Wang, F.; Schmidhammer, U.; de la Lande, A.; Mostafavi, M. Ultra-fast Charge Migration Competes with Proton Transfer in the Early Chemistry of H_2O^{o+}. *Phys. Chem. Chem. Phys.* **2017**, *19*, 2894−2899.
(6) Kuleff, A. K.; Cederbaum, L. S. Ultrafast correlation-driven electron dynamics. *J. Phys. B: At., Mol. Opt. Phys.* **2014**, *47*, 124002.
(7) Lepine, F.; Ivanov, M. Y.; Vrakking, M. J. J. Attosecond molecular dynamics: fact or fiction? *Nat. Photonics* **2014**, *8*, 195−204.
(8) Tavernelli, I. Nonadiabatic Molecular Dynamics Simulations: Synergies between Theory and Experiments. *Acc. Chem. Res.* **2015**, *48*, 792−800.
(9) Li, X.; Smith, S. M.; Markevitch, A. N.; Romanov, D. A.; Levis, R. J.; Schlegel, H. B. A time-dependent Hartree-Fock approach for studying the electronic optical response of molecules in intense fields. *Phys. Chem. Chem. Phys.* **2005**, *7*, 233−239.
(10) Klamroth, T. Laser-driven electron transfer through metal-insulator-metal contacts: Time-dependent configuration interaction singles calculations for a jellium model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *68*, 245421.
(11) Rohringer, N.; Gordon, A.; Santra, R. Configuration-interaction-based time-dependent orbital approach for ab initio treatment of electronic dynamics in a strong optical laser field. *Phys. Rev. A: At., Mol., Opt. Phys.* **2006**, *74*, 043420.
(12) Kato, T.; Kono, H. Time-dependent multiconfiguration theory for electronic dynamics of molecules in an intense laser field. *Chem. Phys. Lett.* **2004**, *392*, 533−540.
(13) Zanghellini, J.; Kitzler, M.; Brabec, T.; Scrinzi, A. Testing the multi-configuration time-dependent Hartree−Fock method. *J. Phys. B: At., Mol. Opt. Phys.* **2004**, *37*, 763.
(14) Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules. In *Recent Advances in Density Functional Methods*; World Scientific: 1995; pp 155−192.
(15) Tsuneda, T.; Hirao, K. Self-interaction corrections in density functional theory. *J. Chem. Phys.* **2014**, *140*, 18A513.
(16) Hofmann, D.; Kümmel, S. Self-interaction correction in a real-time Kohn-Sham scheme: Access to difficult excitations in time-dependent density functional theory. *J. Chem. Phys.* **2012**, *137*, 064117.
(17) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.
(18) Takimoto, Y.; Vila, F. D.; Rehr, J. J. Real-time time-dependent density functional theory approach for frequency-dependent nonlinear optical response in photonic molecules. *J. Chem. Phys.* **2007**, *127*, 154114.
(19) Nguyen, T. S.; Parkhill, J. Nonadiabatic Dynamics for Electrons at Second-Order: Real-Time TDDFT and OSCF2. *J. Chem. Theory Comput.* **2015**, *11*, 2918−2924.
(20) Lopata, K.; Govind, N. Modeling Fast Electron Dynamics with Real-Time Time-Dependent Density Functional Theory: Application to Small Molecules and Chromophores. *J. Chem. Theory Comput.* **2011**, *7*, 1344−1355.
(21) Sun, J.; Song, J.; Zhao, Y.; Liang, W.-Z. Real-time propagation of the reduced one-electron density matrix in atom-centered Gaussian orbitals: Application to absorption spectra of silicon clusters. *J. Chem. Phys.* **2007**, *127*, 234107.
(22) Lopata, K.; Govind, N. Near and Above Ionization Electronic Excitations with Non-Hermitian Real-Time Time-Dependent Density Functional Theory. *J. Chem. Theory Comput.* **2013**, *9*, 4939−4946.
(23) Lopata, K.; Van Kuiken, B. E.; Khalil, M.; Govind, N. Linear-Response and Real-Time Time-Dependent Density Functional Theory Studies of Core-Level Near-Edge X-Ray Absorption. *J. Chem. Theory Comput.* **2012**, *8*, 3284−3292.
(24) Mitrić, R.; Werner, U.; Bonačić-Koutecký, V. Nonadiabatic dynamics and simulation of time resolved photoelectron spectra within time-dependent density functional theory: Ultrafast photoswitching in benzylideneaniline. *J. Chem. Phys.* **2008**, *129*, 164118.
(25) Vincendon, M.; Dinh, P. M.; Romaniello, P.; Reinhard, P.-G.; Suraud, É. Photoelectron spectra from full time dependent self-interaction correction. *Eur. Phys. J. D* **2013**, *67*, 97.
(26) Cheng, C.-L.; Evans, J. S.; Van Voorhis, T. Simulating molecular conductance using real-time density functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *74*, 155112.
(27) Schaffhauser, P.; Kümmel, S. Using time-dependent density functional theory in real time for calculating electronic transport. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 035115.
(28) Chen, H.; Ratner, M. A.; Schatz, G. C. Time-Dependent Theory of the Rate of Photo-induced Electron Transfer. *J. Phys. Chem. C* **2011**, *115*, 18810−18821.
(29) Chapman, C. T.; Liang, W.; Li, X. Ultrafast Coherent Electron−Hole Separation Dynamics in a Fullerene Derivative. *J. Phys. Chem. Lett.* **2011**, *2*, 1189−1192.
(30) Peralta, J. E.; Hod, O.; Scuseria, G. E. Magnetization Dynamics from Time-Dependent Noncollinear Spin Density Functional Theory Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 3661−3668.
(31) Tavernelli, I.; Gaigeot, M.-P.; Vuilleumier, R.; Stia, C.; Hervé du Penhoat, M.-A.; Politis, M.-F. Time-Dependent Density Functional Theory Molecular Dynamics Simulations of Liquid Water Radiolysis. *ChemPhysChem* **2008**, *9*, 2099−2103.
(32) Tavernelli, I.; Röhrig, U. F.; Rothlisberger, U. Molecular dynamics in electronically excited states using time-dependent density functional theory. *Mol. Phys.* **2005**, *103*, 963−981.
(33) Li, X.; Tully, J. C.; Schlegel, H. B.; Frisch, M. J. Ab initio Ehrenfest dynamics. *J. Chem. Phys.* **2005**, *123*, 084106.
(34) Kolesov, G.; Grånäs, O.; Hoyt, R.; Vinichenko, D.; Kaxiras, E. Real-Time TD-DFT with Classical Ion Dynamics: Methodology and Applications. *J. Chem. Theory Comput.* **2016**, *12*, 466−476.
(35) Gaigeot, M. P.; Lopez-Tarifa, P.; Martin, F.; Alcami, M.; Vuilleumier, R.; Tavernelli, I.; Hervé du Penhoat, M. A.; Politis, M. F. Theoretical investigation of the ultrafast dissociation of ionised biomolecules immersed in water: Direct and indirect effects. *Mutat. Res., Rev. Mutat. Res.* **2010**, *704*, 45−53.
(36) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.
(37) Curutchet, C.; Muñoz-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.
(38) Monari, A.; Rivail, J.-L.; Assfeld, X. Theoretical Modeling of Large Molecular Systems. Advances in the Local Self Consistent Field Method for Mixed Quantum Mechanics/Molecular Mechanics Calculations. *Acc. Chem. Res.* **2013**, *46*, 596−603.
(39) Jurinovich, S.; Pescitelli, G.; Di Bari, L.; Mennucci, B. A TDDFT/MMPol/PCM model for the simulation of exciton-coupled circular dichroism spectra. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16407−16418.
(40) Liang, W.; Chapman, C. T.; Ding, F.; Li, X. Modeling Ultrafast Solvated Electronic Dynamics Using Time-Dependent Density Functional Theory and Polarizable Continuum Model. *J. Phys. Chem. A* **2012**, *116*, 1884−1890.
(41) Nguyen, P. D.; Ding, F.; Fischer, S. A.; Liang, W.; Li, X. Solvated First-Principles Excited-State Charge-Transfer Dynamics with Time-Dependent Polarizable Continuum Model and Solvent Dielectric Relaxation. *J. Phys. Chem. Lett.* **2012**, *3*, 2898−2904.
(42) Chapman, C. T.; Liang, W.; Li, X. Solvent Effects on Intramolecular Charge Transfer Dynamics in a Fullerene Derivative. *J. Phys. Chem. A* **2013**, *117*, 2687−2691.
(43) Corni, S.; Pipolo, S.; Cammi, R. Equation of Motion for the Solvent Polarization Apparent Charges in the Polarizable Continuum Model: Application to Real-Time TDDFT. *J. Phys. Chem. A* **2015**, *119*, 5405−5416.

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

(44) Delgado, A.; Corni, S.; Pittalis, S.; Rozzi, C. A. Modeling solvation effects in real-space and real-time within density functional approaches. *J. Chem. Phys.* **2015**, *143*, 144111.
(45) Ding, F.; Lingerfelt, D. B.; Mennucci, B.; Li, X. Time-dependent non-equilibrium dielectric response in QM/continuum approaches. *J. Chem. Phys.* **2015**, *142*, 034120.
(46) Dinh, P. M.; Reinhard, P. G.; Suraud, E. Dynamics of clusters and molecules in contact with an environment. *Phys. Rep.* **2010**, *485*, 43−107.
(47) Köster, A. M.; Geudtner, G.; Alvarez-Ibarra, A.; Calaminici, P.; Casida, M. E.; Carmona-Espindola, J.; Dominguez, V.; Flores-Moreno, R.; Gamboa, G. U.; Goursot, A.; Heine, T.; Ipatov, A.; de la Lande, A.; Janetzko, F.; del Campo, J.-M.; Mejia-Rodriguez, D.; Reveles, J.; Vasquez-Perez, J.; Vela, A.; Zuniga-Gutierrez, B.; Salahub, D. R. *deMon2k Version 5*; Mexico City, Mexico, 2016.
(48) Mintmire, J. W.; Dunlap, B. I. Fitting the Coulomb potential variationally in linear-combination-of-atomic-orbitals density-functional calculations. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 88−95.
(49) Dunlap, B. I.; Rösch, N.; Trickey, S. B. Variational fitting methods for electronic structure calculations. *Mol. Phys.* **2010**, *108*, 3167−3180.
(50) Köster, A. M. Hermite Gaussian auxiliary functions for the variational fitting of the Coulomb potential in density functional methods. *J. Chem. Phys.* **2003**, *118*, 9943−9951.
(51) Köster, A. M.; Flores-Moreno, R.; Reveles, J. U. Efficient and reliable numerical integration of exchange-correlation energies and potentials. *J. Chem. Phys.* **2004**, *121*, 681−690.
(52) Köster, A. M.; Reveles, J. U.; del Campo, J. M. Calculation of exchange-correlation potentials with auxiliary function densities. *J. Chem. Phys.* **2004**, *121*, 3417−3424.
(53) Salahub, D.; Noskov, S.; Lev, B.; Zhang, R.; Ngo, V.; Goursot, A.; Calaminici, P.; Köster, A.; Alvarez-Ibarra, A.; Mejía-Rodríguez, D.; Řezáč, J.; Cailliez, F.; de la Lande, A. QM/MM Calculations with deMon2k. *Molecules* **2015**, *20*, 4780.
(54) Lamoureux, G.; Roux, B. t., Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(55) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141−6156.
(56) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.
(57) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of Ion−Protein Interactions Using the Drude Polarizable Force-Field. *J. Phys. Chem. B* **2015**, *119*, 9401−9416.
(58) Ngo, V.; da Silva, M. C.; Kubillus, M.; Li, H.; Roux, B.; Elstner, M.; Cui, Q.; Salahub, D. R.; Noskov, S. Y. Quantum Effects in Cation Interactions with First and Second Coordination Shell Ligands in Metalloproteins. *J. Chem. Theory Comput.* **2015**, *11*, 4992−5001.
(59) Alvarez-Ibarra, A.; Köster, A. M.; Zhang, R.; Salahub, D. R. Asymptotic Expansion for Electrostatic Embedding Integrals in QM/MM Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 4232−4238.
(60) Gao, Q.; Yokojima, S.; Fedorov, D. G.; Kitaura, K.; Sakurai, M.; Nakamura, S. Octahedral point-charge model and its application to fragment molecular orbital calculations of chemical shifts. *Chem. Phys. Lett.* **2014**, *593*, 165−173.
(61) Devereux, M.; Raghunathan, S.; Fedorov, D. G.; Meuwly, M. A Novel, Computationally Efficient Multipolar Model Employing Distributed Charges for Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4229−4241.
(62) Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Andrés Cisneros, G. ICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *J. Comput. Chem.* **2016**, *37*, 1019−1029.
(63) Almlöf, J.; Faegri, K.; Korsell, K. Principles for a direct SCF approach to LICAO−MOab-initio calculations. *J. Comput. Chem.* **1982**, *3*, 385−399.
(64) Dupuis, M.; Aida, M.; Kawashima, Y.; Hirao, K. A polarizable mixed Hamiltonian model of electronic structure for micro-solvated excited states. I. Energy and gradients formulation and application to formaldehyde (1A2). *J. Chem. Phys.* **2002**, *117*, 1242−1255.
(65) Lu, Z.; Zhang, Y. Interfacing ab Initio Quantum Mechanical Method with Classical Drude Osillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *J. Chem. Theory Comput.* **2008**, *4*, 1237−1248.
(66) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.
(67) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.
(68) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(69) de la Lande, A.; Clavaguéra, C.; Köster, A. On the accuracy of population analyses based on fitted densities. *J. Mol. Model.* **2017**, *23*, 99.
(70) Castro, A.; Marques, M. A. L.; Rubio, A. Propagators for the time-dependent Kohn−Sham equations. *J. Chem. Phys.* **2004**, *121*, 3425−3433.
(71) Bandrauk, A. D.; Lu, H. Exponential propagators (integrator) for the time-dependent Schrödinger equation. *J. Theor. Comput. Chem.* **2013**, *12*, 1340001.
(72) Magnus, W. On the exponential solution of differential equations for a linear operator. *Commun. Pure. Appl. Math* **1954**, *7*, 649−673.
(73) Gilmore, R. Baker-Campbell-Hausdorff formulas. *J. Math. Phys.* **1974**, *15*, 2090−2092.
(74) Williams-Young, D.; Goings, J. J.; Li, X. Accelerating Real-Time Time-Dependent Density Functional Theory with a Nonrecursive Chebyshev Expansion of the Quantum Propagator. *J. Chem. Theory Comput.* **2016**, *12*, 5333−5338.
(75) Penka, E. F.; Couture-Bienvenue, E.; Bandrauk, A. D. Ionization and harmonic generation in CO and H_2CO and their cations with ultrashort intense laser pulses with time-dependent density-functional theory. *Phys. Rev. A: At., Mol., Opt. Phys.* **2014**, *89*, 023414.
(76) Brabec, T.; Krausz, F. Intense few-cycle laser fields: Frontiers of nonlinear optics. *Rev. Mod. Phys.* **2000**, *72*, 545−591.
(77) Wopperer, P.; Dinh, P. M.; Reinhard, P. G.; Suraud, E. Electrons as probes of dynamics in molecules and clusters: A contribution from Time Dependent Density Functional Theory. *Phys. Rep.* **2015**, *562*, 1−68.
(78) Souaille, M.; Loirat, H.; Borgis, D.; Gaigeot, M. P. MDVRY: a polarizable classical molecular dynamics package for biomolecules. *Comput. Phys. Commun.* **2009**, *180*, 276−301.
(79) Tabacchi, G.; Mundy, C. J.; Hutter, J.; Parrinello, M. Classical polarizable force fields parametrized from ab initio calculations. *J. Chem. Phys.* **2002**, *117*, 1416−1433.
(80) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11−28.
(81) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995−2016.
(82) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(83) Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. Calculation of the nonlinear optical properties of molecules. *J. Comput. Chem.* **1990**, *11*, 82−87.
(84) Calaminici, P.; Jug, K.; Köster, A. M. Density functional calculations of molecular polarizabilities and hyperpolarizabilities. *J. Chem. Phys.* **1998**, *109*, 7756−7763.
(85) Sophy, K. B.; Pal, S. Density functional response approach for the linear and nonlinear electric properties of molecules. *J. Chem. Phys.* **2003**, *118*, 10861−10866.
(86) Sophy, K. B.; Calaminici, P.; Pal, S. Density Functional Static Dipole Polarizability and First-Hyperpolarizability Calculations of Nan (n = 2, 4, 6, 8) Clusters Using an Approximate CPKS Method and its

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

Comparison with MP2 Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 716−727.
(87) Flores-Moreno, R.; Köster, A. M. Auxiliary density perturbation theory. *J. Chem. Phys.* **2008**, *128*, 134105.
(88) Carmona-Espíndola, J.; Flores-Moreno, R.; Köster, A. M. Time-dependent auxiliary density perturbation theory. *J. Chem. Phys.* **2010**, *133*, 084102.
(89) Carmona-Espíndola, J.; Flores-Moreno, R.; Köster, A. M. Static and dynamic first hyperpolarizabilities from time-dependent auxiliary density perturbation theory. *Int. J. Quantum Chem.* **2012**, *112*, 3461−3471.
(90) R Development Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2012.
(91) Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **2006**, *6*, 7−11.
(92) Carmona-Espíndola, J.; Köster, A. M. Photoabsorption spectra from time-dependent auxiliary density functional theory. *Can. J. Chem.* **2013**, *91*, 795−803.
(93) Song, P.-S.; Gordon, W. H. Spectroscopic study of the excited states of coumarin. *J. Phys. Chem.* **1970**, *74*, 4234−4240.
(94) Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures*. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123−134.
(95) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(96) Caldwell, J. W.; Kollman, P. A. Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208−6219.
(97) Köster, A. M.; del Campo, J. M.; Janetzko, F.; Zuniga-Gutierrez, B. A MinMax self-consistent-field approach for auxiliary density functional theory. *J. Chem. Phys.* **2009**, *130*, 114106.
(98) Domínguez-Soria, V. D.; Geudtner, G.; Morales, J. L.; Calaminici, P.; Köster, A. M. Robust and efficient density fitting. *J. Chem. Phys.* **2009**, *131*, 124102.
(99) Lagardère, L.; Lipparini, F.; Polack, É.; Stamm, B.; Cancès, É.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II. Toward Massively Parallel Computations Using Smooth Particle Mesh Ewald. *J. Chem. Theory Comput.* **2015**, *11*, 2589−2599.
(100) Alvarez-Ibarra, A.; Köster, A. M. A new mixed self-consistent field procedure. *Mol. Phys.* **2015**, *113*, 3128−3140.
(101) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoret. Chim. Acta* **1977**, *44*, 129−138.
(102) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547−2553.
(103) Fonseca Guerra, C.; Handgraaf, J.-W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *J. Comput. Chem.* **2004**, *25*, 189−210.
(104) Loco, D.; Polack, É.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations. *J. Chem. Theory Comput.* **2016**, *12*, 3654−3661.
(105) Liang, W.; Chapman, C. T.; Li, X. Efficient first-principles electronic dynamics. *J. Chem. Phys.* **2011**, *134*, 184102.
(106) Agostini, F.; Abedi, A.; Suzuki, Y.; Min, S. K.; Maitra, N. T.; Gross, E. K. U. The exact forces on classical nuclei in non-adiabatic charge transfer. *J. Chem. Phys.* **2015**, *142*, 084303.

DOI: 10.1021/acs.jctc.7b00251
J. Chem. Theory Comput. 2017, 13, 3985−4002

SIMULATING ELECTRON DYNAMICS IN POLARIZABLE ENVIRONMENTS

Xiaojing Wu[1], Jean-Marie Teuler[1], Fabien Cailliez[1], Carine Clavaguéra[1],

Dennis R. Salahub[2,3], Aurélien de la Lande[1]*

1: Laboratoire de Chimie Physique, CNRS - Université Paris Sud, Université Paris-Saclay. 15 avenue Jean Perrin, 91405, Orsay CEDEX, France.

2: Department of Chemistry, Centre for Molecular Simulation, Institute for Quantum Science and Technology and Quantum Alberta, University of Calgary, 2500 University Drive N. W., Calgary, Alberta, Canada T2N 1N4.

3: College of Chemistry and Chemical Engineering, Henan University of Technology, No 100, Lian Hua Street, High-Tech Development zone, Zhengzhou 450001, P. R. China

*To whom correspondence should be addressed: aurelien.de-la-lande@u-psud.fr

SUPPLEMENTARY INFORMATION

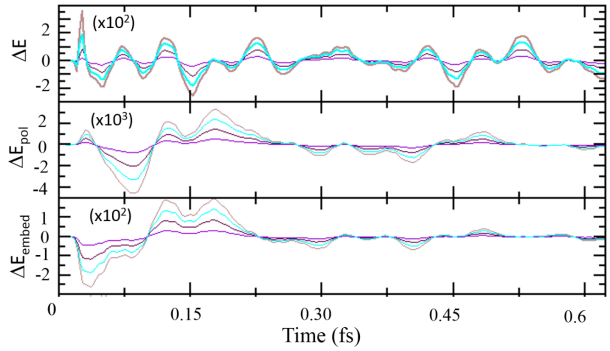


Figure S1. Error estimations of the MMpol-stationary-RT-TDDFT scheme. The 0.1as time-step simulation is taken as a reference and the data for larger time steps are given with respect to the reference. Light brown: 1 as, cyan : 0.75 as, marron: 0.5 as, violet: 0.25 as. All the energies are given in kcal/mol scaled by coefficients given on each graph. ΔE is the total QM/MM energy, ΔE_{pol} is the total polarization energy and ΔE_{embed} is the embedding energy (permanent charges and induced dipole). The peptide is initially perturbed by a electric field of strength 0.01a.u.

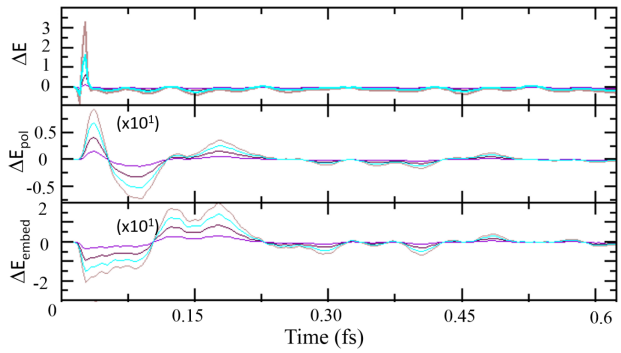


Figure S2. Error estimations of the MMpol-stationary-RT-TDDFT scheme. The 0.1as time-step simulation is taken as a reference and the data for larger time steps are given with respect to the reference. Light brown: 1 as, cyan : 0.75 as, marron: 0.5 as, violet: 0.25 as. All the energies are given in kcal/mol scaled by coefficients given on each graph. ΔE is the total QM/MM energy, ΔE_{pol} is the total polarization energy and ΔE_{embed} is the embedding energy (permanent charges and induced dipole). The peptide is initially perturbed by a electric field of strength 0.1a.u.

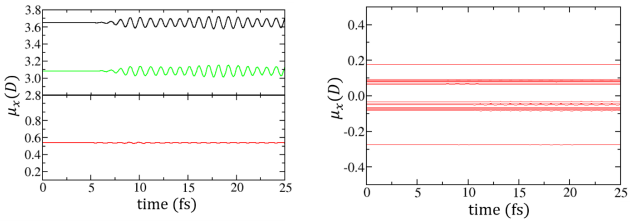


Figure S3: Evolution of dipole moments (x component) during RT-TDDFT propagation with the Voronoi Deformation Density scheme. Left: the molecular dipole moment is shown in black, the dipole derived from atomic charges is shown in green and the sum of intrinsic atomic dipole moments is shown in red. Right: intrinsic atomic dipole moments are shown in red for each coumarin atom.

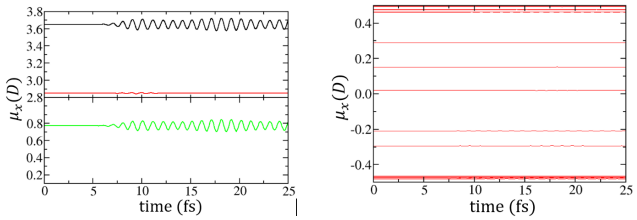


Figure S4: Evolution of dipole moments (x component) during RT-TDDFT propagation with the Becke scheme. Left: the molecular dipole moment is shown in black, the dipole derived from atomic charges is shown in green and the sum of intrinsic atomic dipole moments is shown in red. Right: intrinsic atomic dipole moments are shown in red for each coumarin atom.

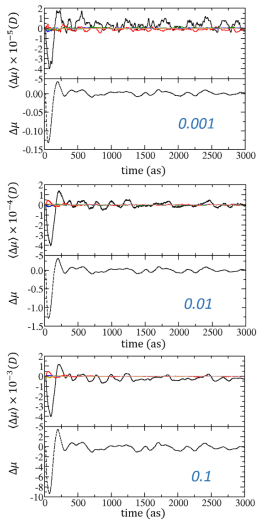


Figure S5. Correlation between the variations of the induced dipole moments on water molecules averaged by hydration shells $\langle \Delta\mu(t) \rangle = \langle \mu(t) - \mu(0) \rangle$ (upper-part of each graph) and the variation of dipole moment of the peptide (Lower-part of each graph). The three main graphs correspond to three perturbing field strengths of 0.001a.u (Top-Lefts), 0.01 (Bottom-Left) and 0.1 a.u. (Top-Right); see main text for details.

GENERAL CONCLUSION

In this PhD thesis, we have reported progress along two lines. In the first project, the objective was to improve the reliability of computational approaches dedicated to the simulation of redox potentials of heme proteins. Redox potentials can be evaluated with the linear response approximation, which is the equivalent of the Marcus framework at the microscopic level. A QM+MM method was chosen to compute the vertical energy gap between diabatic states thereby leading to redox free energies of reduction. We focused on improving the electrostatic description of the MM part in two directions. One was to take into account polarization effects by induction. The other was to go beyond the point charge description by using multipolar descriptions associated with explicit polarization.

In Chapter 1, as workhorses, we considered six cytochromes. Simulations with non-polarizable FF provide globally the correct trend of redox potential values compared to experiments. Nevertheless, for differences of redox potential between proteins of the order of 50 mV, our simulations fail to reach this level of resolution.

In Chapter 2 we made an important step toward by devising the first AMOEBA sets of parameters for the heme cofactor in its ferrous and ferric forms. The results of interaction between heme and water solvent shells were found to be very encouraging. AMOEBA results reach the accuracy of QM results and outperform that of non-polarizable FF results. Combined with the recent implementation of AMOEBA in TINKER-HP program, these parameters open the possibility in a near future for evaluating redox potentials, and more generally, owing to the ubiquity of heme proteins, for addressing other important properties. Last but not least, all these advanced approaches can be used to simulate electron transfer rate in heme proteins. In all these possible applications the enhanced computational cost of AMOEBA, even in the TINKER-HP program, will likely require to imagine combined schemes between non-polarizable forces fields and AMOEBA. The former to explore faster the conformational spaces of proteins, the latter to obtain better electrostatic interaction.

In the second project, our objective was to develop new methods for investigating ultrafast electron dynamics in extended biomolecules. In Chapter 3 we reported the first implemented Real-time Time-Dependent DFT in deMon2k. Thanks to the use of density fitting techniques, one can address electron dynamics simulation in systems comprised of hundreds of atoms. A wide range of processes can be tackled

such as perturbations by electric fields or inelastic collisions with charged particles. Various propagation algorithms and analysis tools were implemented.

In Chapter 4, we reported an original combination of RT-TDDFT and polarizable MM force field adapted to biomolecules. An efficient and robust implementation of this method has been realized in deMon2k software. Density fitting techniques allow again to reduce the computational cost in the RT-TDDFT propagations and QM/MM coupling. The code is ready to simulate electron dynamics in extended biomolecules. This method has been applied to understand the complexity of response mechanisms of environment of a solute peptide submitted to an external perturbation.

ANNEXES

ANNEX I

Table I. DFT derived charge set for Cyt. b type Heme (PFe-ImMe$_2$) in both oxidized and reduced states in CHARMM format.

Cyt.b5	Atom	PFe^{II}-$ImMe_2$	PFe^{III}-$ImMe_2$
	CB	-0.5329	-0.5551
	HB1	0.1719	0.1832
	HB2	0.1503	0.168
	ND1	-0.4409	-0.4294
	HD1	0.353	0.37
HSD	CG	0.274	0.3044
	CE1	-0.0843	-0.0836
	HE1	0.1294	0.1358
	NE2	-0.5227	-0.5315
	CD2	0.2721	0.292
	HD2	0.0968	0.1057
	CB	-0.5332	-0.5554
	HB1	0.1498	0.1679
	HB2	0.1721	0.1835
	ND1	-0.4412	-0.4299
	HD1	0.3528	0.3702
HSD	CG	0.2752	0.3065
	CE1	0.2711	0.2911
	HE1	0.0979	0.1083
	NE2	-0.5216	-0.5305
	CD2	-0.0852	-0.0854
	HD2	0.1316	0.1395
	FE	1.6873	1.7946
	NA	-0.6092	-0.625
	NB	-0.6062	-0.6202
	NC	-0.628	-0.6444
	ND	-0.6341	-0.65
	C1A	0.3146	0.3184
HEME	C2A	-0.0872	-0.0439
	C3A	-0.0117	0.0084
	C4A	0.2937	0.3122
	C1B	0.2845	0.3049
	C2B	0.031	0.0418
	C3B	-0.138	-0.0902
	C4B	0.3283	0.3284
	C1C	0.292	0.2981
	C2C	0.0436	0.0703
	C3C	-0.1547	-0.1232
	C4C	0.3232	0.3361
	C1D	0.292	0.309
	C2D	-0.004	0.0179
	C3D	-0.0945	-0.0544
	C4D	0.3198	0.3242
	CHA	-0.4064	-0.3912
	HA	0.1458	0.1626
	CHB	-0.4036	-0.3975
	HB	0.1427	0.1609
	CHC	-0.3891	-0.3741
	HC	0.1401	0.1568
	CHD	-0.4277	-0.4201
	HD	0.147	0.1643
	CMA	-0.4525	-0.4751
	HMA1	0.1438	0.1624
	HMA2	0.1388	0.1576
	HMA3	0.1376	0.1491
	CAA	-0.2352	-0.2118
	HAA1	0.1115	0.1208
	HAA2	0.1212	0.1374
	CBA	-0.28	-0.28
	HBA1	0.1621	0.1641
	HBA2	0.1598	0.156
	CGA	0.62	0.62
	O1A	-0.76	-0.76
	O2A	-0.76	-0.76
	CMB	-0.4664	-0.487
	HMB1	0.1386	0.1548
	HMB2	0.1352	0.148
	HMB3	0.1521	0.1682
	CAB	-0.0362	-0.0725
	HAB	0.1013	0.116
	CBB	-0.3583	-0.2999
	HBB1	0.13	0.1454
	HBB2	0.1419	0.1484
	CMC	-0.4672	-0.4929
	HMC1	0.143	0.159
	HMC2	0.1379	0.1507
	HMC3	0.15	0.1688
	CAC	-0.0355	-0.0667

	HAC	0.0977	0.1119
	CBC	-0.3642	-0.3098
	HBC1	0.1286	0.1452
	HBC2	0.1399	0.1477
	CMD	-0.4719	-0.495
	HMD1	0.1398	0.1538
	HMD2	0.137	0.1515
	HMD3	0.154	0.1723
	CAD	-0.2415	-0.2146
	HAD1	0.1263	0.1425
	HAD2	0.1153	0.1244
	CBD	-0.28	-0.28
	HBD1	0.157	0.1691
	HBD2	0.1613	0.1602
	CGD	0.62	0.62
	O1D	-0.76	-0.76
	O2D	-0.76	-0.76

Table II. DFT derived charge set for Cyt. c type Heme (PFe-ImMe-EMS-MMS_2) in both oxidized and reduced states in CHARMM format.

Cyt. c551	Atom	PFe^{II}-ImMe-EMS-MMS_2	PFe^{III}-ImMe-EMS-MMS_2
CYS	CB	-0.4042	-0.3668
	HB1	0.1526	0.1706
	HB2	0.1354	0.1478
	SG	0.0349	0.0552
	CB	-0.3965	-0.361
	HB1	0.1547	0.1671
	HB2	0.1362	0.1442
	SG	0.0147	0.0429
HSD	CB	-0.5272	-0.5461
	HB1	0.1496	0.1657
	HB2	0.1691	0.1794
	ND1	-0.4337	-0.4236
	HD1	0.3557	0.3707
	CG	0.2698	0.3003
	CE1	0.2633	0.2676
	HE1	0.099	0.1097
	NE2	-0.4779	-0.4801
	CD2	-0.0957	-0.1038
	HD2	0.1365	0.1481
MET	CB	-0.4124	-0.4285
	HB1	0.1239	0.1422
	HB2	0.1386	0.1397
	CG	-0.2239	-0.2242
	HG1	0.1344	0.1382
	HG2	0.1115	0.1321
	SD	0.0954	0.1359
	CE	-0.523	-0.5299
	HE1	0.1551	0.1752
	HE2	0.1606	0.1718
	HE3	0.1721	0.184
HEME	FE	1.5964	1.7165
	NA	-0.642	-0.6649
	NB	-0.6252	-0.6494
	NC	-0.6307	-0.6536
	ND	-0.6446	-0.6672
	C1A	0.2939	0.3108
	C2A	-0.0937	-0.0495
	C3A	-0.0049	0.0043
	C4A	0.2909	0.3146
	C1B	0.2849	0.3083
	C2B	0.025	0.0472
	C3B	-0.1516	-0.101
	C4B	0.3063	0.3156
	C1C	0.2805	0.2998
	C2C	0.0249	0.0476
	C3C	-0.1566	-0.1407
	C4C	0.3194	0.3373
	C1D	0.2939	0.3127
	C2D	0.0009	0.0216
	C3D	-0.0863	-0.039
	C4D	0.3048	0.2989
	CHA	-0.4144	-0.3941
	HA	0.1459	0.1664
	CHB	-0.3838	-0.3804
	HB	0.1513	0.167
	CHC	-0.3893	-0.3867
	HC	0.1354	0.153
	CHD	-0.3913	-0.3854
	HD	0.1468	0.1623
	CMA	-0.4545	-0.4745
	HMA1	0.1388	0.1607
	HMA2	0.1409	0.1534
	HMA3	0.1409	0.1526
	CAA	-0.1737	-0.2005

	HAA1	0.1166	0.1241
	HAA2	0.1175	0.1367
	CBA	-0.28	-0.28
	HBA1	0.1591	0.1598
	HBA2	0.1662	0.1536
	CGA	0.62	0.62
	O1A	-0.76	-0.76
	O2A	-0.76	-0.76
	CMB	-0.4688	-0.4896
	HMB1	0.136	0.1604
	HMB2	0.1415	0.1533
	HMB3	0.1392	0.1629
	CAB1	0.02	-0.0129
	HAB1	0.0883	0.1142
	CBB1	-0.4763	-0.4772
	HBB11	0.1347	0.1518
	HBB21	0.1432	0.1437
	HBB31	0.143	0.1478
	CMC	-0.4599	-0.4825
	HMC1	0.145	0.1595
	HMC2	0.1342	0.1509
	HMC3	0.1434	0.1601
	CAC1	0.0191	0.0036
	HAC1	0.1021	0.1056
	CBC1	-0.4752	-0.4796
	HBC11	0.1346	0.1487
	HBC21	0.1384	0.147
	HBC31	0.1401	0.1437
	CMD	-0.463	-0.4798
	HMD1	0.1418	0.1543
	HMD2	0.1452	0.1617
	HMD3	0.1501	0.1574
	CAD	-0.1767	-0.2105
	HAD1	0.1235	0.1342
	HAD2	0.1081	0.1354
	CBD	-0.28	-0.28
	HBD1	0.1377	0.1562
	HBD2	0.1635	0.133
	CGD	0.62	0.62
	O1D	-0.76	-0.76
	O2D	-0.76	-0.76

ANNEX II

Transformation between Atomic orbitals (AO) and Molecular orbital (MO) representation:

AO basis are a set of normalized but not orthogonal functions $\{\mu(\boldsymbol{r})\}$, S is the overlap matrix.

$$S_{\mu v} = \langle \mu | v \rangle \quad (1)$$

It is always possible to find a transformation matrix such that transformed set of function $\{\mu(\boldsymbol{r})'\}$ are orthonormal.

$$\mu(\boldsymbol{r})' = \sum_{v} X_{\mu v} v(\boldsymbol{r}) \quad (2)$$

$$X^{\dagger} S X = 1 \quad (3)$$

Since S is Hermitian, it can be diagonalized by a unitary matrix U, where s is a diagonal matrix of the eigenvalues of S.

$$U^{\dagger} S U = s \quad (4)$$

One can use the symmetric orthogonalization by using the inverse square root of S for X

$$X \equiv S^{-1/2} = U s^{-1/2} U^{\dagger} \quad (5)$$

However, if there is linear dependence or near linear dependence in the basis set, then some of the eigenvalues will approaches zero and will involve dividing by quantities that are nearly zero. Thus we have chosen to use canonical orthogonalization.

$$X = U s^{-1/2} \quad (6)$$

$$X^{\dagger} S X = (U s^{-1/2})^{\dagger} S U s^{-1/2} \quad (7)$$

We see that X is also an orthogonalizing transformation matrix. If any s_i approaches to zero (≤10^{-4}), we can truncated and then got matrix $\bar{X}$. If we have N atomic orbitals with m linear dependencies, we eliminated these m columns with too small values. We got an N*(N-m) matrix. Converting the KS matrix from AO basis to MO basis is then straightforward:

$$H' = X^{\dagger} H X \quad (8)$$

H (AO basis) is an N*N matrix, while H' (MO basis) is a smaller (N-m)*(N-m) matrix.

It is slightly more complicated to convert density matrix P from AO to MO by the fact that X is not square and cannot be easily inverted. Here we use left and right inverses.

$$X_L^{-1} = (X^\dagger X)^{-1} X^\dagger \quad (9)$$

$$(X^\dagger)_R^{-1} = X(X^\dagger X)^{-1} \quad (10)$$

We get

$$P' = (X^\dagger X)^{-1} X^\dagger P X (X^\dagger X)^{-1} \quad (11)$$

Because $(X^\dagger X)^{-1} = s$

$$P' = s X^\dagger P X s \quad (12)$$

$$P = X^\dagger P' X \quad (13)$$

In summary, the transformation of KS matrix H (AO basis) to H' (MO basis) can be performed with Eq. 8. The transformation between density matrix P to P' can be performed with Eqs. 12 and 13.

Titre : Contribution au Développement de Simulations Numériques des Dynamiques Electroniques et Moléculaires Pour des Biomolécules Environnées.

Mots clés : Simulations numériques, potentiel d'oxydoréduction, hémoprotéines, champs de forces polarisables, dynamique électronique attoseconde

Résumé : Cette thèse porte sur deux projets visant au développement de nouvelles approches pour simuler les dynamiques moléculaire et électronique avec application à des biomolécules étendues.

Dans la première partie nous cherchons à améliorer significativement la précision des simulations des propriétés rédox des protéines. Dans ce contexte, l'objectif est de recourir à de champ de force reposant sur une description multipolaire des interactions électrostatiques (AMOEBA) pour estimer les potentiels redox d'hémoprotéines. Nous avons dérivé des paramètres pour AMOEBA afin de décrire précisément les interactions électrostatiques avec l'hème. Une amélioration très encourageante est obtenue par rapport aux champs de forces standard.

Le second projet vise à développer de nouvelles méthodes pour étudier la dynamique des électrons dans des biomolécules à l'échelle attoseconde en incluant les effets d'environnement. Nous avons conçu un couplage original entre la théorie de la fonctionnelle de la en temps réel densité dépendant du temps (RT-TDDFT) et un modèle de mécanique moléculaire polarisable (MMpol). Une implémentation efficace et robuste de cette méthode a été réalisée dans le logiciel deMon2k. L'utilisation de techniques d'ajustements de densités électroniques auxiliaires permet de réduire drastiquement le coût de calcul des propagations RT-TDDFT/MMpol. La méthode est appliquée à l'analyse de la dissipation d'énergie dans l'environnement d'un peptide excité par un impulsion laser.

Title : Contribution to the Development of Advanced Approaches for Electron and Molecular Dynamics Simulations in Extended Biomolecules

Keywords : Numerical simulations, redox potential, hemoproteins, polarizable force fields, attosecond electron dynamics

Abstract : This thesis involves two projects devoted to the development of advanced approaches for simulating molecular and electron dynamics in extended biomolecules.

The first project aims at significantly improving the accuracy of redox potentials of proteins by numerical simulations. A sophisticated force field relying on a multipolar description of electrostartic interactions (AMOEBA) is used to perform molecular dynamics simulations onheme proteins. We derived parameters for AMOEBA to accurately describe electrostatic interactions with hemein both ferrous and ferric states.Very encouraging improvements are obtained compared to the standard force fields.

The second project aims at developing original approaches for simulating ultrafast electron dynamics in biomolecules in contact to polarizable environments. We devised acombination of Real-time Time-Dependent Density Functional Theory (RT-TDDFT) and polarizable Molecular Mechanics (MMpol). An efficient and robust implementation of this method has been realized in deMon2k software. Density fitting techniques allow to reduce the computational cost of RT-TDDFT/MMpol propagations. The methodology is applied to understand the mechanisms of energy dissipation of a peptide excited by a laser pulse.