PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Polarizable embedding QM/MM: the future gold standard for complex (bio)systems?†

Mattia Bondanza,[a] Michele Nottoli,[a] Lorenzo Cupellini,[a] Filippo Lipparini,[a] Benedetta Mennucci*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Nowadays, hybrid QM/MM approaches are widely used to study (supra)molecular systems embedded in complex biological matrices. However, in their common formulation, mutual interactions between the quantum and classical parts are neglected. To go beyond such a picture, a polarizable embedding can be used. In this perspective, we focus on the induced point dipoles formulation of polarizable QM/MM approaches and we show how efficient and linear scaling implementations have allowed their application to the modeling of complex biosystems. In particular, we discuss their use in the prediction of spectroscopies and in molecular dynamics simulations, including Born-Oppenheimer dynamics, enhanced sampling techniques and nonadiabatic descriptions. We finally suggest the theoretical and computational developments that still need to be achieved to overcome the limitations which have prevented so far a larger diffusion of these methods.

1 Introduction

Hybrid approaches which integrate quantum chemistry and molecular mechanics (MM) have an almost 45 year-long history being their first formulation published by Warshel and Levitt in 1976.[1] During the years many important developments have been proposed. The main ones concern the accuracy and completeness of the coupling between the quantum and the classical subsystem, and the extension to describe properties and processes beyond the energy of a ground state system. In parallel, the same approaches have been largely reformulated in their numerical and computational aspects so to achieve robust and high-performing implementations. All these improvements have allowed QM/MM methods to become one of the most successful and popular strategies to describe complex systems especially of biological interest. The latter in fact represent an optimal field of study for a QM/MM formulation. In enzymatic reactivity or in the many examples of light-activated biological functions, in fact, the process of interest can be clearly localized in a limited part of the whole system while the rest mostly acts as a perturbation through both short-range and specific effects (mostly H-bonds) and longer range electrostatic interactions. As a matter of fact, enzymes and their catalytic activity have been the very first and the most frequent application of QM/MM methods.[1–6] Another important field where QM/MM methods have reported important successes is the interpretation and the simulation of spectroscopic experiments in a variety of complex environments.[7,8] In particular, in recent years QM/MM methods have been successfully applied to the simulation of electronic spectroscopies of biological systems.[9–12] This has been made possible mostly thanks to the extension of QM/MM to the Time Dependent Density Functional Theory (TDDFT) description of excitation processes. The computational efficiency of TDDFT in describing electronic transitions in medium-to-large molecular systems has in fact allowed to simulate the chromophoric unit present in photoresponsive proteins without the need of introducing approximated model systems. Finally, QM/MM methods have been extended to molecular dynamics (MD) simulations either using an adiabatic or a nonadiabatic formulation.[13]

This brief summary clearly shows that QM/MM methods represent an extremely powerful strategy which has permeated all the main fields of application of quantum chemistry and, at the same time, has largely extended the borders of its applicability towards more and more complex systems. Once recognized that, however, we have to add that most of the applications appeared so far are based on a purely electrostatic formulation of the QM-MM coupling. When we say QM/MM, in fact, we usually indicate the so-called electrostatic embedding (EE) formulation of the model. Namely, the QM subsystem "sees" the MM one as a set of fixed point charges (or a fixed multipolar expansion). Within this formulation any mutual polarization effects and/or any nonelectrostatic interactions between the QM and MM sub-

[a] *Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via G. Moruzzi 13, I-56124 Pisa, Italy E-mail: benedetta.mennucci@unipi.it*

† Electronic Supplementary Information (ESI) available: general treatment of the electrostatic interactions for distributed multipoles and detailed analysis of the computational cost of all the involved terms. See DOI: 10.1039/cXCP00000x/

systems is completely neglected. In the case of nonelectrostatic interactions (e.g. dispersion and repulsion), their effects on the energy can be recovered in an uncoupled way by converting the QM system into a set of MM parameters and calculating these interactions through the common classical potentials used in MM force fields. Due to the generally weaker effect of these interactions when compared to electrostatics, this uncoupled formulation is the standard in QM/MM descriptions and only very few attempts to go beyond that have been proposed for water as solvent,[14] and for more complex environments including biological systems.[15–17] On the contrary, various methods have been presented so far to include mutual polarization effects. A strategy is to introduce the effect of polarization using "ab initio" FFs which are fully built on first principles and require no fitted parameters. Very popular examples of this type are the Effective Fragment Potential (EFP) method[18] and the X-Pol strategy.[19] An alternative approach is to use MM but polarizable FFs[20–23] based either on Drude oscillators,[24–29] fluctuating charges[30–34] or induced point dipoles (IPD).[1,35–42] We recall that within the latter framework, the MM subsystem is represented via a set of fixed point charges, and possibly higher order multipoles, and by endowing polarizable sites with atomic polarizabilities.

In this perspective, we will focus on the IPD formulation of polarizable embedding for different reasons. The first is that IPD polarizable FFs have shown to be extremely accurate not only for describing standard solvents but also for biomacromolecules such as proteins.[43,44] Secondly, the use of atomic induced dipoles allows a clearer interpretation of the response of a biological environment with respect to other types of polarizable FFs. Finally, IPD polarizable FFs are a very active research field especially within the group associated to the development of AMOEBA,[45,46] which represents one of the most sophisticated polarizable FFs. In the presentation, we start from the state of the art of the IPD QM/MM formulation, outlining its main theoretical aspects and how they have been translated into efficient and linear scaling implementations for ground state and excited systems. Then we proceed illustrating applications to spectroscopies and dynamic processes thanks to their extension to molecular dynamics techniques. Starting from a Born-Oppenheimer description, we will show that polarizable QM/MM can be combined with enhanced sampling techniques and reformulated for nonadiabatic dynamics. We finally note that both the presentation of the methodological and computational aspects and the selection of possible applications will be biased towards biosystems. Different types of complex embedding have in fact different specificities which in many cases require different solutions.[47,48]

2 An induced point dipole formulation of polarizable embedding

In this section, we provide an overview of the IPD polarizable QM/MM model for the description of molecular systems in their ground and excited states. We then describe how the approach can be used to perform Born-Oppenheimer (BO) molecular dynamics (MD) simulations, with particular emphasis on the computational aspects of such simulations. All formal and mathematical details are kept to a minimum. We refer the reader to the supporting information and to previous work[38,39,49–52] for a more comprehensive description of the theory and implementation.

2.1 The theory

In a polarizable embedding model, each MM atom is endowed with a point charge and a polarizability. In more sophisticated polarizable force fields (FF), such as AMOEBA,[45,46] higher order multipoles are considered as well. As such a generalization does not alter the substance of the discussion while requiring a much more complex and cumbersome notation, we limit our discussion to point charges. A more general formalism and discussion can be found in the Supporting Information. The polarizability allows the MM atoms to respond to an electric field, generated by the charges on the classical atoms and by the QM density, by creating an induced dipole. The QM/MM electrostatic and polarization interaction energy can be written as follows[38,49,53]

$$\mathscr{E}^{\mathrm{QM/MM}} = \frac{1}{2}\sum_i q_i V_i^{\mathrm{MM}} + \sum_i q_i V_i^{\mathrm{QM}} + \frac{1}{2}\sum_i \left(\alpha_i^{-1}\mu_i^2 + \sum_{j\neq i} \mu_i T_{ij} \mu_j \right) - \sum_i \mu_i (E_i^{\mathrm{MM}} + E_i^{\mathrm{QM}}). \quad (1)$$

In eq. 1, the first term is the interaction among the MM charges, where V_i^{MM} is the potential generated by all other charges at site i, the second term is the interaction between the MM charges and the QM density, V_i^{QM} being the QM potential at site i. These two terms are common to standard EE QM/MM. The third and fourth terms are the self-interaction of the induced dipoles, which can be understood as the work needed to induce the dipole itself, and the repulsion between all induced dipoles, where T_{ij} is the dipole-dipole interaction tensor. Finally, the last two terms are the interaction energy of the induced dipoles with the electric fields produced by the MM charges and QM density, respectively. The last four terms do not appear in EE QM/MM and are characteristic of polarizable embedding models. The energy in eq. 1 is a variational functional[53–55] of the induced dipoles themselves. The minimum of the energy corresponds to the situation where the IPDs maximize the favorable interaction energy with the MM charges and QM density, while at the same time minimizing the repulsion among themselves. The equations for the induced dipoles are obtained by differentiating eq. 1 and read[38,39]

$$\alpha_i^{-1}\mu_i + \sum_{j\neq i} \mathscr{T}_{ij}\mu_j = E_i^{\mathrm{MM}} + E_i^{\mathrm{QM}}. \quad (2)$$

where $\mathscr{T}_{ij}$ is the effective dipole-field tensor that includes a damping function, namely:[56]

$$[\mathscr{T}_{ij}] = \frac{f_e}{|\mathbf{r}_i - \mathbf{r}_j|^3}\mathbf{I} - \frac{3f_t}{|\mathbf{r}_i - \mathbf{r}_j|^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yx \\ zx & zy & z^2 \end{bmatrix} \quad (3)$$

where $\mathbf{I}$ is the unit matrix and f_e and f_t are distance dependent screening functions originally introduced by Thole[57] to avoid the so-called "polarization catastrophe", e.g. the divergence of the Coulomb interaction between two point dipoles when they get

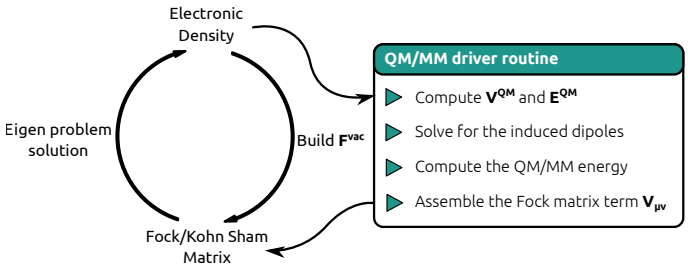

Fig. 1 Schematic representation of the coupling between the SCF algorithm and a polarizable embedding.

too close.

Equations 1 and 2 are the main constituents of a polarizable QM/MM implementation. Assuming the QM subsystem is described at a Self-Consistent Field (SCF) level of theory, such as Hartree-Fock (HF) or Kohn-Sham (KS) Density Functional Theory (DFT), the polarizable QM/MM energy can be computed by assembling the energy in eq. 1 and adding to the Fock or KS matrix the additional contributions, that are the derivatives of eq. 1 with respect to the density matrix. In the atomic orbitals (AO) basis:[38,39,49]

$$V_{\mu\nu}^{\mathrm{pol}} = \sum_i q_i \langle \chi_\mu(\mathbf{r}) | \frac{1}{|\mathbf{r}-\mathbf{r}_i|} | \chi_\nu(\mathbf{r}) \rangle - \sum_i \mu_i \langle \chi_\mu(\mathbf{r}) | \frac{\mathbf{r}-\mathbf{r}_i}{|\mathbf{r}-\mathbf{r}_i|^3} | \chi_\nu(\mathbf{r}) \rangle. \tag{4}$$

The first term in eq. 4 is independent of the QM charge density and can be summed with the one-electron Hamiltonian, as it is commonly done in EE QM/MM. The second term, on the other hand, represents at the same time the main advantage and the main complication introduced by a polarizable model. The IPDs, that are computed by solving the linear system 2, depend on the QM density, as the QM electric field appears in the right-hand side. Therefore, the Fock matrix contribution that stems from the induced dipoles is density dependent. This reflects the fact that the QM density and the IPDs are mutually polarized, i.e., the classical environment is able to adapt to changes in the QM density, which is the main strength of polarizable models. On the other hand, this mutual polarization introduces a non-linearity in the QM/MM interaction, which means that the QM and IPD equations need to be solved iteratively. This is, in practice, not a problem for SCF-based levels of theory, as the SCF procedure is iterative itself. At each SCF iteration, given the density matrix, the QM field has to be assembled. Then, eq. 2 is solved and the dipoles are used to update the Fock or KS matrix. The coupled procedure is schematized in figure 1.

A specific note concerns the integration of a polarizable embedding within a post-HF description for the QM subsystem. The non-linear term introduced by the coupling in the Hamiltonian poses in fact a problem. Mutuating the terminology from Continuum Solvation models, three different coupling schemes are possible.[58–65] The first, called PTE (perturbation to the energy), uses the QM/MM HF orbitals to compute the amplitudes and correlation energy, decoupling therefore correlation from the response of the environment. The second, called PTD (perturbation to the density), computes a correlated relaxed density matrix without contributions from the environment and then uses it to compute the IPD. The last, called PTED[59–61,64,66,67] (perturbation to the energy and density), is the synthesis of PTE and PTD: a correlated density is computed starting from the PTE amplitudes and used to update the response of the environment; the procedure is then iterated until self-consistency of the relaxed density is reached. The PTE scheme has been proven to be consistent with second-order perturbation theory[61] and, given its simplicity, is the most commonly employed strategy[68–70], even though PTED implementations have been presented.[62,63,68,69]

2.2 The implementation

Once presented the main theoretical aspects of the model, some considerations of computational nature are mandatory. The various operations required to perform a polarizable QM/MM calculation can be grouped into three categories: i) computing various one-electron integrals, ii) computing the interaction between the classical electrostatic distributions (charges and dipoles), and iii) solving the polarization equations. Operations in the first group can be performed at a cost that scales as $\mathcal{O}(N_{\mathrm{QM}}N_{\mathrm{MM}})$, where N_{QM} and N_{MM} are the number of QM and MM atoms, respectively. This is in general not a source of major overheads and the difference between electrostatic and polarizable embeddings is simply that field integrals are also required. Therefore, neither major implementation complications nor computational bottlenecks are introduced in this respect. The operations in groups ii) and iii) are, on the other hand, much more delicate from a computational point of view, as they exhibit a formal scaling of $\mathcal{O}(N_{\mathrm{MM}}^2)$, or even of $\mathcal{O}(N_{\mathrm{MM}}^3)$ if the linear system in eq. 2 is solved using standard dense linear-algebra methods. While this burden is shared with EE QM/MM, in the latter case only charge-charge interaction need to be computed. In IPD embeddings, besides having to compute charge-dipole and dipole-dipole interactions, the polarization equations 2 need to be solved at each SCF iteration, severely aggravating the overall computation. Here, it should be mentioned that using an advanced polarizable force field such as AMOEBA[45,71] introduces further complications due to the presence of higher order multipoles (up to quadrupoles) and two sets of induced dipoles. Without going into detail, it suffices to say that a QM/AMOEBA calculation is about twice as expensive as a charge and dipoles polarizable embedding one.

In order to use polarizable QM/MM for large systems, an efficient implementation is therefore of paramount importance. First, the polarization linear system is symmetric and positive definite and can therefore be efficiently solved iteratively using the preconditioned conjugate gradient method.[53] The preconditioner proposed by Wang and Skeel[72] is particularly effective in reducing the number of iterations, which, in our experience, is usually as little as 10-12. Second, one can easily realize that the matrix-vector products needed to solve such a linear system can be recast as the computation of the electric field generated by the induced dipoles, making the distinction between groups ii) and iii) inessential. Electrostatic properties can be computed at a computational cost that scales linearly with the number of MM atoms by using a suitable fast summation technique, such as the Fast Multipole Method (FMM).[73] Recently, we have presented a gen-

eral, FMM based implementation for polarizable QM/MM[50,52] that allows one to compute the interaction between multipoles of, in principle, arbitrary order. Thanks to this implementation, all the computations required by the polarizable model can be performed at a cost that scales as $\mathcal{O}(N_{\mathrm{MM}})$. In reference 52, we showed that systems comprising as many as one million polarizable atoms can be easily treated on standard computer nodes with minimal computational overhead. The same machinery can be extended to the complex electrostatics present in the AMOEBA force field, so that also QM/AMOEBA calculations can be performed efficiently.[46,52,53,72]

To summarize, polarizable QM/MM computations can be affordably performed also for very large systems. However, there is a price to pay. Their implementations are complex and cumbersome and require substantial modifications of a QM code. Furthermore, in order to get a substantial improvement of the performance, the use of complex fast summation techniques is mandatory, further adding to the complexity of the implementation. For our implementation, which aims at maximizing efficiency, we therefore choose a fully integrated approach, using the Gaussian 16 electronic structure program.[74,75] The implementation is entangled with Gaussian both in terms of I/O and memory management, and in terms of code structure. It also makes use of Gaussian's internal FMM library.

At the the cost of reducing to some extent the efficiency, it is possible to achieve a code-independent, modular implementation. The strategies proposed so far can be grouped into two main approaches. The first one resembles as much as possible a fully integrated implementation, but reduces the modifications that one needs to implement in the QM software by creating an external library that performs all the calculations implied by a polarizable scheme. Such a library consists of all the MM specific routines, that deal with input processing, computing MM interactions and inducing fields, solving the polarization equations and handling the contractions between MM quantities and QM integrals. In order to couple such a library to an existing QM package, the user needs to take care of high-level driver routines, that handle the interface between the QM package, mainly calling its internal integrals routines, and the library itself, updating then the energy, the Fock matrix, or other QM quantities for coupling with post-SCF treatments. This strategy is implemented in the CPPE library[76] and is very promising, as it allows one to add the capability to treat a polarizable embedding to any QM software without the need to deal with the cumbersome details of the electrostatics involved. Moreover, the open source nature of the project makes this interface very appealing and could actually help to make polarizable QM/MM much more broadly available in QM software packages. Furthermore, an external library can in principle achieve an efficient implementation and full mutual polarization with a non trivial, but minimal implementation.

A second option is to write software that exploits existing features of QM and MM programs without modifying them, but implementing the QM/MM coupling externally. This is in principle very appealing, however, it implies certain approximations. In particular, the solution to the QM and polarization equations is uncoupled, and therefore no fully self-consistent mutual polarization can be achieved, if not at the price of performing the SCF calculation many times. In this framework, the coupling is handled relying on the fact that most QM package offer electrostatic embedding capabilities based on point charges. The electrostatic field obtained from the MM engine, which can include contributions from both static and induced multipoles, is applied to the QM density by approximating all the MM sources as point charges, the values and coordinates of which are then provided to the QM software for the electrostatic embedding. This is a second approximation introduced by the scheme. For polarizable embedding, a correction to the polarization and interaction energy can be further applied using the field generated by the converged SCF density to recompute the polarization degrees of freedom. The approximations introduced by such a coupling schemes are counterbalanced by a substantially reduced computational cost. The Lichem[77] and Chem Shell[78] implementations of polarizable embedding are representative of this second philosophy.

2.3 The extension to excited states

One of the most attractive features of polarizable QM/MM models is that they are particularly suited for the description of electronic excitations. This can be explained in two equivalent ways. From a physical point of view, an excitation process causes a sizeable rearrangement of the electronic density, which in turn modifies the interaction between the excited molecule and the environment. Polarizable embeddings have the flexibility to respond to such a change, which is lacking in EE schemes. Equivalently, the non-linearity introduced by the former in the Hamiltonian implies an explicit contribution to the electric response function of the molecule, which is absent in EE schemes. In the framework of SCF-based linear response (LR) theory, excitation energies ω and transition densities X, Y are obtained by solving the so called time-dependent SCF equations, given here in Casida's formulation:[79]

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{5}$$

where the A and B matrices are given by

$$\begin{aligned} A_{ia,jb} &= \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + (ia|jb) - c_x(ab|ij) + c_l f^{xc}_{aibj} + \mathscr{V}^{\mathrm{env}}_{iajb} \\ B_{ia,jb} &= (ia|jb) - c_x(aj|ib) + c_l f^{xc}_{aibj} + \mathscr{V}^{\mathrm{env}}_{iajb} \end{aligned} \tag{6}$$

where i, j are the indices of occupied molecular orbitals, and a, b of virtual molecular orbitals, ε_i are the molecular orbital energies, $(ai|bj)$ is a bielectronic integral in the Mulliken notation, and f^{xc}_{aibj} is the exchange–correlation kernel. The coefficients c_x and c_l control the amount of exact exchange for hybrid functionals by interpolating between HF ($c_x = 1$, $c_l = 0$) and pure functionals ($c_x = 0$, $c_l = 1$). The last term is the response of the polarizable embedding,[10,38,39,80]

$$\mathscr{V}^{\mathrm{env}}_{iajb} = -\sum_k E_{ia,k}\mu_{jb,k} \tag{7}$$

$\mu_{jb,k}$ is the induced dipole at site k created by the field arising from orbitals $\phi_j\phi_b$, and $E_{ia,k}$ is the field computed at the k-th site arising from $\phi_i\phi_a$.

Polarizable models are naturally adapted to the LR treatment of excitation and response processes, however, a LR description might not be enough to fully account for the interaction of an excited system with the environment. The LR response, in fact, can be interpreted as the instantaneous response of the MM polarizable sites to the transition density of the electronic excitation. What instead is missing in this formulation is the effect due to the (instantaneous) polarization of the environment to the different electronic charge densities of the ground and excited states. Analogously to what is observed with Continuum Solvation Models,[81,82] this poses the problem of how to introduce a state-specific (SS) polarizable QM/MM response.

Unfortunately, there is no unique way to introduce SS couplings or corrections within a LR theory, nor strong theoretical argument to choose among different schemes. If one is interested in correcting the transition energy for a particular excitation, an approach based on first-order perturbation theory, known as corrected LR (cLR), can be effectively used. This was originally formulated for Continuum Solvation Models[83] but later it has been extended to polarizable QM/MM models.[39,84] However, this does not correct the state's density, in some analogy with what done by PTE scheme to the correlation amplitudes. A possible solution is to iterate the cLR procedure, i.e., recompute the LR amplitude in the presence of a field produced by the environment polarized by the excited-state density, iterating the procedure until self-consistency is reached, in close analogy to what is done in PTED schemes. Such a procedure has been proposed for continuum solvation models[85,86] but it is very expensive from a computational point of view, and introduces terms in the excitation energy that are not consistent with a first-order, LR approach. Even if a state-specific QM method is used, for instance CASSCF, the introduction of a state-specific response term is problematic, especially if state-average formulations have to be used to get reliable results.[87–89]

From this discussion it should come out clearly that the extension of polarizable QM/MM approaches to excitation processes remain challenging and no definitive theoretical solution has been found for the problem. What is clear is that LR and SS responses correspond to different physical contributions to the excited-state energy, and, in principle, both have to be taken into account. On a positive note, we can say that the use of a LR response when combined with cLR corrections is usually more than adequate to correctly predict the excitation energies. In more difficult cases where this strategy is not enough because of the intrinsic limits of the LR QM description in accurately describing the change in the state density, an integration of LR and SS QM methods (mostly CASSCF) has been proposed and successfully applied to various photoresponsive proteins.[90,91]

To conclude the extension of polarizable QM/MM approaches to excitation processes, it is worth recalling a particular class of complex systems that have largely benefited from such a multiscale description, namely multichromophoric systems, in which an electronic excitation can be delocalized over several chromophoric units. A very important example of these systems are the pigment-protein complexes (also called antenna complexes) used in photosynthetic organisms to harvest light and transfer the excitation energy to the reaction centers for further transformation. In this kind of systems, the usual partitioning into a QM and a MM part is not necessarily an optimal choice, as a correct description of the excited state requires extending the QM part to multiple molecules. In order to overcome this limitation, the multichromophoric excited states can be described on the basis of the single chromophores' excited states, in the so-called Frenkel exciton model. In this model, the excited states (excitons) are obtained by diagonalization of the exciton matrix, whose diagonal elements are the excitation energies of the noninteracting chromophores (site energies), whereas the off-diagonal elements, the exciton couplings, represent the interaction between excitations in different chromophores. This model can be coupled to an *ab initio* description of both the site energies and the couplings.[12,92,93] The environment plays a major role in determining the nature and the behavior of excitonic systems: this is particularly true for the antenna complexes where the protein embedding the multichromophoric aggregate is not only necessary to keep the different chromophores in "optimal" relative positions but also to tune the energy and the nature of the excitons. This "tuning" is mainly realized through electrostatic and polarization effects of the protein residues which can significantly affect both site energies and exciton couplings.[94–96] For such complexes, but also for other types of excitonic systems, combining a Frenkel exciton model with a polarizable embedding is a very effective strategy. This combination has been developed and implemented by our group in collaboration with Dr. C. Curutchet in Barcelona using the IPD formulation of the polarizable embedding.[38] In particular, within such a formulation, a new term appears in the exciton coupling definition, which represents the Coulomb interaction between chromophores mediated by the environment polarization. This term generalizes a concept originally introduced within continuum solvation models[97–99] to atomistic models thus allowing a more accurate description for heterogeneous and anisotropic environments like a protein matrix.[100]

3 Molecular dynamics

When the interest is mainly on the simulation of complex biological systems, any model, even if very accurate, is not sufficient if it cannot be extended to describe dynamics. While such an extension is now almost routine for the EE QM/MM approach, only few examples have been proposed so far for polarizable embeddings. The main reasons are two: the extension of polarizable QM/MM methods to the analytical derivatives needed for molecular dynamics simulations is neither automatic nor straightforward, and the computational cost is generally much higher than for the electrostatic embedding counterpart. This latter limitation is being overcome, as we have described in the previous section, and, if the implementations will become available in more and more computational codes, polarizable QM/MM MD simulations will rapidly spread. In this section, we present what is already possible within a Born-Oppenheimer approximation eventually in combination with enhanced sampling techniques.

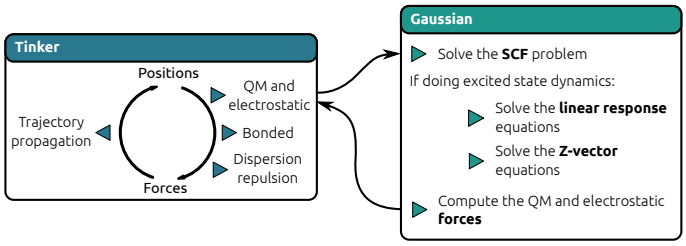

Fig. 2 Example of a coupled QM/MM implementation able to perform BOMD.

3.1 Born-Oppenheimer Molecular dynamics

Polarizable QM/MM Born-Oppenheimer molecular dynamics (BOMD) implementations have been presented both within the Drude oscillator[28] and the IPD formulations of the polarizable embedding.[51,101,102] Here we focus on the implementation that our group has done in collaboration with the group of J.-P. Piquemal in Paris,[51,103] coupling the Gaussian implementation of the IPD QM/MM described in the previous section and the Tinker software.[104,105] The latter handles the propagation of the trajectory, as well as the computation of all the contributions to the energy and forces that do not involve the QM density, the electrostatic distribution and the polarization. In particular, Tinker computes all the Van der Waals interactions, including the QM/MM ones, and all the bonded interactions among MM atoms and, if present, across links between the MM and QM subsystems, that can be handled either using pseudo-bonds[103] or link atoms. Gaussian, instead, deals with the QM subsystem and computes all the electrostatic and polarization contributions to the energy and forces, including both QM/MM and purely MM ones. This allows us to exploit the very efficient FMM-based machinery described in section 2.2, so that we can achieve linear scaling in computational cost with respect to the number of MM atoms. A representation of the steps required to perform BOMD is provided in figure 2.

The IPD QM/MM gradients can be obtained by differentiating eq. 1 with respect to the coordinates and their implementation has been already extensively presented for both ground and excited states[51,52,106,107] and are derived in more detail in the Supporting Information. Here, it suffices to say that they can be assembled by computing contributions of the same kinds discussed for energy computations, namely, electrostatic properties due to the MM charges and dipoles and one-electron integrals, again, at a cost that scales linearly with respect to the number of MM atoms.[52] The latter aspect is of fundamental importance if one wants to use polarizable QM/MM BOMD simulations for modeling all but the smallest systems, as due to the very large number of energy and forces evaluations needed to compute a sufficiently long trajectory, it is paramount that each single force computation is performed in the most efficient way possible.

In our implementation, even for very large systems, the cost is completely dominated by the cost of the QM calculation alone, with the PE part causing overall only a very small increase in computational cost - in the order of a few percents. To give an idea of the efficiency achieved with the herein presented implementation we recall two cases. All the timings were obtained using a single cluster node equipped with two Intel Xeon Gold 5120 processors (28 cores) and 128 GB of memory. The two systems are reported in fig. 3 (i) the BOMD of a small molecule (28 atoms) in an explicit solvent box (13500 atoms). We performed both ground and excited state MD simulation still using AMOEBA as polarizable FF in combination with DFT for the ground and TDDFT for the excited state (ωB97X-D/6-31+G(d), 194 basis functions). With this setup we were able to perform about 2250 steps of MD per day on the ground state and about 500 steps per day on the excited state. (ii) an excited state BOMD simulation of a large chromophore (94 atoms) in an fully polarizable solvated protein environment (5900 atoms). Again, we employed AMOEBA as polarizable FF and TDDFT (CAM-B3LYP/6-31G(d), 734 basis functions) for the QM region. For this system, we were able to perform about 200 steps of MD per day. Considering that 0.5 fs integration step is normally used in these simulations, several ps of trajectory can be easily obtained in both ground and excited states.

A final important note about MD simulations is the definition of boundary conditions. Due to the great efficiency of fast summation techniques, in MM MD this is normally done through periodic boundary conditions (PBC). When using QM methods that exploit atom-centered basis functions, the application of such a strategy is not convenient due to the local character of the wavefunction and alternative approaches have to be devised, which fall in the family of non periodic boundary conditions. Non periodic boundary conditions must provide a mechanical embedding potential to prevent the "evaporation" of the MM molecules, and possibly an electrostatic embedding to reproduce the bulk properties. If only a mechanical embedding is employed, there should be enough MM molecules such that the QM molecule is properly solvated, i.e. its properties should correspond to the ones calculated in bulk. This can be a harder requirement if the QM molecule is charged. An overview of different approaches involving non periodic boundary conditions as well as a historical perspective is given in ref. 108. Among the approaches presented, two are particularly suited for QM/MM MD, and involve the use of an external potential or of a layer of frozen solvent molecules. The frozen layer is a simple solution, available in most MD codes, but it can create some bias at the interface, thus requiring more MM molecules. On the other hand, external potentials provide more reliable boundary conditions, and attractive-repulsive potentials have been used in combination with a spherical cavity.[109] Mechanical embedding potentials have been used in combination with a polarizable continuum model to provide non periodic boundary conditions for EE QM/MM MD.[110–112] The extension of this approach to polarizable MM requires all the components of the system to be mutually polarized which clearly introduces additional computational cost.[24,32,49,50,113]

3.2 Enhanced sampling and free energy methods

When dealing with complex chemical and biological systems, ensemble sampling with MD is slowed down by the presence of high free energy barriers. As such, obtaining a correct sampling with conventional MD (cMD) requires to simulate the system for very long time (up to ms), which is only affordable with specific hard-

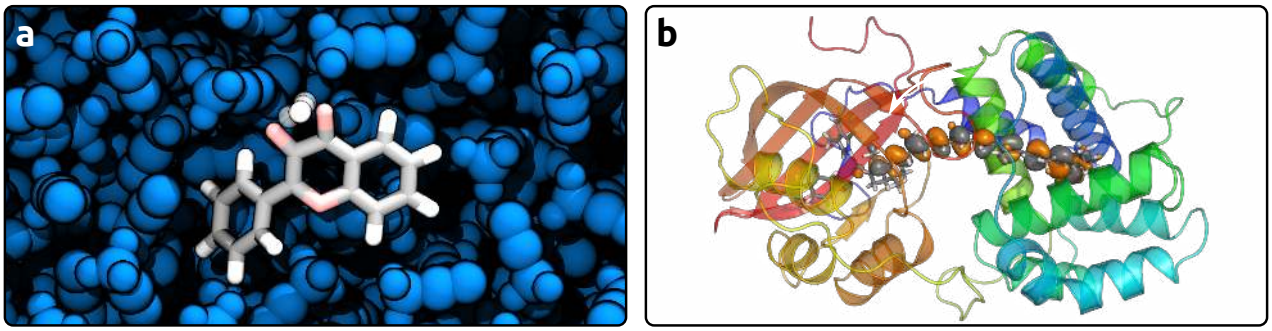

Fig. 3 Examples of systems studied with polarizable QM/MM BOMD. Panel (a) shows shows a QM molecule of 3-hydroxiflavone in MM acetonitrile (drawn in blue). The outcome of the excited state BOMD is an excited state proton transfer, which is highlighted by the white arrow. Panel (b) shows the orange carotenoid protein: the QM part is the carotenoid, the MM part comprises the protein and the solvent molecules (water), which are not drawn for clarity.

ware and with fully classical potentials.[114,115] Numerous methods, called *enhanced sampling* techniques, have been developed in order to face these problems at a much lower computational cost.[116] These methods can be divided into two subcategories: unconstrained methods and collective variable ones.

Among unconstrained methods, replica exchange MD (REMD) exploits simultaneous simulations at different temperatures, or with parametrically scaled Hamiltonians, to cross the high energy barriers of the system.[117] The most widely used formulations allow to preserve the correct canonical sampling of the system in each replica. The main advantage is that no *a priori* knowledge of the free energy surface is needed, and that the massively parallel infrastructures of super-computing centers can be exploited in their full potential. On the other hand, these are rather expensive methods for systems that have a large number of degrees of freedom. Another unconstrained enhanced sampling technique is accelerated MD (aMD),[118,119] and the closely related Gaussian accelerated MD (GaMD).[120] In these techniques the potential is modified in such a way that the barriers are lowered, while conserving the most significant features of the free energy surface (FES). Then, a single MD trajectory is propagated with the new Hamiltonian, and the resulting ensemble is reweighted to match the one that is obtained in the canonical sampling of the original FES.[119]

Collective variable (CV) methods, such as umbrella sampling (US), (well-tempered) metadynamics ((wt-)MetaD),[121–123] and temperature accelerated MD (TAMD)[124] can be considered as a combination of dimensionality reduction techniques with accelerated MD. They leverage an *a priori* knowledge of the free energy landscape through the definition of a CV, a function of the coordinates of the system that can be interpreted as an approximation of a reaction coordinate, and should ideally collect all the slow degrees of freedom of the system. Then a bias is applied only along the CV direction in the phase space, while all the other degrees of freedom are unbiased and are therefore sampled in a cMD framework. To obtain a proper sampling with these methods it is necessary to simulate the system long enough to allow the remaining unbiased degrees of freedom to relax. When this is achieved, the reconstruction of the free energy landscape along the collective variable is obtained. Since the bias is only applied in a one- (or low-) dimensional projection of the phase-space, all these methods can conveniently be applied to highly dimensional systems without particular disadvantages, provided that a good CV is defined.

A detailed discussion of the theoretical aspects of enhanced sampling methods and their applications is beyond the scope of this perspective, and the reader should refer to specific reviews on the topic.[116,119,125] Here, instead, we focus on the extension of enhanced sampling methods to QM/MM Hamiltonians.

As a matter of fact, QM/MM dynamics and enhanced sampling techniques have been developed, and became widely used methods, during the first years of the new millennium. The main objective of these studies was to obtain accurate free energy estimations for chemical reactions in complex environments. Enzymatic reactions are an ideal target for this research: in the early 2000s, QM/MM dynamics with semiempirical (or sometimes DFT) quantum methods and classical non-polarizable force fields were used in combination with umbrella sampling[126] and thermodynamic integration techniques,[127–129] to determine free energy profiles of enzymatic reactions. Non-enzymatic reactions such as Diels-Alder cycloadditions were also studied, but with fewer applications.[130] Moving to more sophisticated enhanced sampling methods, it is worth mentioning the application of transition path sampling (TPS)[131] to the study of lactate dehydrogenase enzymes.[132]

It should be noted that, since the propagation of a QM/MM dynamics requires a great computational effort and only few ps per day of MD can be produced with current methods, the interest was mainly focused on methods, such as US or TPS, where the sampling of the relevant part of the phase space can be performed simultaneously, by exploiting independent trajectories. Despite these CV methods are less flexible than other ones such as MetaD, and a good guess of the CV is essential to obtain good results, they are probably most suited to these studies as they can trivially exploit massively parallel infrastructures to produce a reasonable sampling in relatively low wall-clock time.

To the best of our knowledge, replica exchange methods in combination with QM/MM descriptions were only applied as a proof of concept on model systems.[133,134] The reason for this minor interest in these methods is that, despite being very general

and powerful, when applied to large systems they require a large number of replicas to be propagated for long time. This is often not affordable for QM/MM Hamiltonians with the current computational power; therefore, when a good CV is available, the application of CV methods is by far more convenient. It should be underlined that modification of the classic temperature REMD algorithm, as solute tempering, can significantly reduce the computational cost associated with these calculations as shown by Schwörer *et al.*, but applications to "real-life" cases are still missing.

The implementation of enhanced sampling techniques within a polarizable QM/MM framework does not present significant theoretical issues, because these methods just alter the equation of motion used to propagate the dynamics, and they are not dependent on the potential used. Therefore coupling these methods with polarizable multiscale models is generally straightforward, and only requires the MD engine to support the specific sampling method that one wants to apply. The ingredients needed for the simplest methods (such as harmonic potentials to perform US) are already implemented in almost all MD software, including Tinker,[135] and can be easily exploited to perform free energy calculations. In order to use more complex and flexible CV based methods, it is very convenient to exploit specific codes that provide a general interface to do so. In this regard we mention the powerful Plumed library,[136] which can be added as a plugin in virtually any MD engine with little effort, and is already recognized as a standard in the field. REMD based methods are implemented in some MD engines, and their implementation is in theory quite straightforward. In practice, since this approaches require many trajectory to be propagated in parallel and communicate each other, for an efficient implementation a well organized parallel interface is almost mandatory, making it preferable to exploit the already existing implementations.

Despite the apparent simplicity of integrating enhanced sampling techniques and polarizable QM/MM, the applications are still extremely scarce.[134] This is most probably due to the fact that, with inefficient implementations, the cost of polarizable MM can be quite high. The main approach was therefore to use larger QM regions in combination with cheap QM methods embedded in static MM environments. Nonetheless, new inexpensive and linear-scaling implementations of polarizable embedding for different QM methods as the ones described in the previous section will make these approaches more and more appealing. In particular, the availability of very general polarizable FF, like AMOEBA, should allow the calculation of free energy profiles with a protocol very similar to the ones cited above, but with a more robust description of the environment. This could also be exploited to reduce the region of QM system to the few atoms actually involved in the chemical reaction leaving everything else as a polarizable environment without considerable loss in accuracy.

4 Simulation of spectroscopies

As already commented in the Introduction, EE QM/MM methods are now routinely employed to study several aspects of molecule-environment interactions and their effect on spectroscopy.[7,137] As a polarizable embedding is, in principle, a more complete approach to describe the QM-classical coupling, polarizable QM/MM approaches are expected to represent an even better framework to reproduce spectroscopic properties of molecules embedded in complex environments, and, in particular biosystems.[8] Here we focus on two specific spectroscopies that have been more often investigated with these hybrid methods, namely vibrational and electronic ones.

4.1 Vibrational spectroscopy

The most common approach to vibrational spectroscopy is represented by the normal-mode analysis in the harmonic approximation. This strategy requires to compute and diagonalize the Hessian matrix of the potential energy surface (PES) calculated at a minimum geometry, whereas IR and Raman intensities are obtained respectively through the derivatives of dipole moment and polarizability. Anharmonicity can be included by computing higher-order derivatives of the energy with respect to the atomic positions. The normal-mode approach clearly represents a powerful technique for its simplicity, and allows an easy assignment of vibrational peaks to normal modes of the molecule. This approach has been largely used in combination with electrostatic QM/MM approaches and more recently extended to polarizable QM/MM for describing molecules in solution.[138,139]

However, in complex environments such as proteins, the PES of the system is extremely rugged, and contains numerous local minima. Extensive sampling of these minima might be required in order to obtain accurate vibrational properties, and this task is usually performed through classical MD or Monte Carlo techniques, followed by partial or full optimization of the system.[140] The latter task can be quite tedious given the roughness of the PES, and the optimizations might be difficult to converge. Moreover, assembly and diagonalization of the Hessian matrix for such large systems can be prohibitively expensive. To overcome this problem, different techniques have been proposed such as the one using a partial Hessian diagonalization (often for the QM region only)[141] or the one computing only modes of interest through algorithms such as vibrational mode-tracking.[142]

A substantially different approach is based on autocorrelation functions (ACFs) extracted directly from a MD simulation.[143,144] It can be shown that the IR spectrum is proportional to the Fourier transform of the dipole-dipole ACF $\langle \mu(t)\mu(0)\rangle$:

$$\mathrm{IR}(\omega) \propto \beta\omega^2 \int_0^\infty e^{i\omega t} \langle \mu(t)\mu(0)\rangle \, \mathrm{d}t \tag{8}$$

similarly, the Raman spectrum is proportional to the Fourier transform of the polarizability ACF.[144] A clear advantage of this method is that it automatically includes anharmonic and finite-temperature effects, while not requiring the identification of all local minima and the corresponding Hessian, which makes this approach particularly appealing for QM/MM calculations.[141,145] On the other hand, assignment of vibrational modes is not straightforward for this approach. A qualitative understanding of vibrational modes can be obtained by computing the velocity autocorrelation function for some atoms or internal coordinates.[146] A more rigorous approach to extract effective normal modes is

based on a variational principle applied to the vibrational frequencies.[143,145]

4.2 Electronic and vibronic spectroscopies

The most common approach employed for QM/MM calculations of electronic absorption spectra is based on the calculation of vertical excitation energies, in many solute/environment configurations, which are sampled by classical MD methods. This "ensemble" method allows considering the effect of the inhomogeneous environment distribution on the band broadening.[147–150]

This approach, however, completely neglects vibronic coupling, which impacts substantially absorption lineshapes even when the vibronic progression itself is hidden under inhomogeneous broadening. In order to overcome this limitation, the vibronic lineshape can be computed at an optimized geometry in implicit solvent, and convoluted with the inhomogeneous distribution function of the excitation energy obtained by MD sampling.[151] In order to include the coupling between the environment and the internal degrees of freedom, this approach should be complemented by a normal-mode calculation at each MD snapshot.[152]

In principle, all the information on the coupling between an electronic excitation $g \rightarrow e$ and the nuclear degrees of freedom of a molecule could be obtained knowing how the excitation energy ΔE changes along the ground-state PES. If the ground-state wavepacket trajectory of the nuclei was known, all the information on vibronic coupling could be determined by the time dependence of ΔE. Based on a ground-state trajectory it is thus possible to compute vibronic properties, such as absorption lineshapes and Resonance Raman scattering, beyond the harmonic approximation, automatically including the effect of the environment. The approach we review here is based on expressing the vibronic coupling in terms of a "spectral density" function (SD), which describes the frequency-dependent linear coupling between the electronic excitation and the nuclear degrees of freedom.[153–156] The electronic lineshapes can be expressed in a time-domain approach within the second-order cumulant expansion formalism.[153]

From a Born-Oppenheimer ground-state trajectory, the SD can be obtained from the Fourier transform of the ACF of the excitation energy fluctuations, i.e., $C_{\mathrm{cl}}(t) = \langle U(t)U(0)\rangle$, where $U(t) = \Delta E(t) - \langle \Delta E\rangle$.

$$J(\omega) = \frac{\beta\omega}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} C_{\mathrm{cl}}(t)\mathrm{d}t \tag{9}$$

Here, $\beta = 1/k_B T$ is the inverse temperature along the Born-Oppenheimer trajectory. The subscript "cl" specifies that the trajectory of the nuclei is classical, and obtained through Newton's equations of motion. The temperature-dependent prefactor in eq. (9) allows reconstructing the spectral density for a quantum trajectory of the nuclei, based on the classical trajectory ACF.[154] In practice, the time-series of the excitation energy is calculated with an excited-state method along the ground-state trajectory.

In order to obtain the absorption lineshape, it is useful to define the *lineshape function* $g_{eg}(t)$, which describes the dephasing between the ground- and excited-state wavepacket dynamics following excitation.[153] In the second-order cumulant expansion, the lineshape function is given by:

$$g(t) = -\int_0^{\infty} \mathrm{d}\omega \frac{J(\omega)}{\omega^2}\left[\coth\left(\frac{\beta\hbar\omega}{2}\right)\left(\cos(\omega t)-1\right) - i\left(\sin(\omega t)-\omega t\right)\right] \tag{10}$$

and the absorption (homogeneous) lineshape can be obtained as:

$$S(\omega-\omega_{eg}) = \Re\int_0^{\infty} \mathrm{d}t\, e^{i(\omega-\omega_{eg})t - g(t)}. \tag{11}$$

where ω_{eg} is the frequency corresponding to the $g \rightarrow e$ vertical excitation. The second-order cumulant expansion is exact only when the ground-state PES is harmonic, and the excitation energy depends linearly on the nuclear coordinates,[154,157] and approximate in the general anharmonic case. Recently, however, it was demonstrated that the second-order cumulant expansion gives accurate results even for anharmonic potentials.[158]

When considering a chromophore in a complex environment such as a protein matrix, the spectral density has contributions from both the internal vibrations of the chromophore and from the motions of the environment. Generally, the former give rise to sharp peaks in the SD, whereas the latter contribute to a low-frequency smooth background.

The environment affects vibronic coupling in several ways. On the one hand it affects directly the chromophore's electronic density, and thus modifies the dependence of ΔE on the nuclear coordinates. On the other hand, it can also affect the nuclear trajectory on which ΔE is computed, by modifying both the equilibrium geometry and the vibrational frequencies. The accuracy of the vibronic calculation is thus directly connected to the embedding method used for the excitation energy calculations and to the one used for the ground-state trajectory. Combining a polarizable QM/MM calculation of excitation energies with a polarizable QM/MM MD offers a route towards accurate prediction of absorption lineshapes in arbitrarily complex environments. We applied this approach to calculating the absorption lineshape for a dye intercalated in DNA[157,159] and for a keto-carotenoid in a protein.[160]

5 Nonadiabatic dynamics

Modeling excited-state processes, such as photoisomerizations and photoinduced electron and/or proton transfers, often requires theory beyond the Born-Oppenheimer approximation. In fact, when different states get closer in energy, the nonadiabatic couplings become non negligible and the BO approximation breaks down. Among the many methods used to perform nonadiabatic molecular dynamics, the most suited for modeling complex chemical and biological systems are those which present the best compromise between accuracy and computational cost. Among them, the most popular ones are the methods in which the nuclear degrees of freedom are classical and the electronic ones are quantum mechanical, namely the mixed quantum classical (MQC) methods. As methods requiring a precomputed potential energy surface are not readily applicable to systems with many degrees of freedom, the most common strategy is to compute the required energy and properties on the fly.[161–163] The most successful and largely used methods belonging to this family

are the trajectory surface hopping (TSH)[164–167] and the methods based on the multiple spawning.[168,169] The mean field Ehrenfest method has been largely used[170,171] as well, but it has a narrower range of applications, since it is only viable in regions of strong non adiabatic coupling.

Nonadiabatic MD techniques have been extended to QM/MM approaches,[13,172] initially through mechanical embedding,[173] and successively through electrostatic embedding. The latter formulation has been successfully used in combination with TSH[174,175] and multiple spawning.[176,177] Different QM methods have been used in QM/MM nonadiabatic simulations, such as semiempirical methods[176,178], TDDFT[179], and CASSCF/CASPT2 methods.[180–185]

In general, the frameworks used to perform nonadiabatic QM/MM simulations combine different codes, specialized in different tasks. One possibility is to interface an existing classical MD code to an electronic structure code, so that the first drives the dynamics and the second performs the QM calculation and takes care of non adiabatic effects. This choice is less general, but presents a reduced amount of data transfer and it is easier to implement; for such reasons it is often used in development implementations. A second option is to use a program that performs the trajectory propagation and calls the appropriate QM and MM codes for the computation of energies and forces. Many production codes fall in this category. COBRAMM[186] has an interface to several QM codes and uses Amber[187] for the MM part; SHARC[188] and Newton-X[189,190] can use a single program to handle the whole QM/MM calculation or they can interface Tinker[104,105] with another QM method of choice.

The extension of nonadiabatic dynamics to polarizable QM/MM approaches is neither straightforward nor unique. The difficulties come from the fact that a polarizable embedding responds differently to different electronic states. This delicate aspect has been already discussed in section 2.3, where we have introduced two different cases which require different formulations of the QM-MM coupling: LR and SS QM methods. For LR QM methods such as TDDFT, the polarizable embedding can address multiple states at the same time, so the extension to nonadiabatic dynamics is straightforward. However, this approach presents two limitations, the first is related to the response of the environment, which lacks the contribution from the relaxed density of the excited state, as explained in section 2.3. The second is related to LR QM methods, which often are not suited for nonadiabatic dynamics: for instance, modelling internal conversion to the ground state requires multireference methods, because the ground state has a multiconfigurational character close to the crossing seam. In these cases, a SS QM method such as CASSCF has to be used but its extension to polarizable QM/MM nonadiabatic dynamics presents some specific theoretical issues which do not apply to standard electrostatic embeddings. The SS polarizable embedding in fact responds to a specific density, and as such, the correct definition of a nonadiabatic dynamics trajectory is not uniquely defined. Considering TSH as the nonadiabatic method, we can formulate three different strategies for representing the response of the polarizable embedding, which are graphically represented in Fig. 4.

The first strategy allows the environment to respond to the average density, in a state-average formalism. This could be a possible solution in regions of strong coupling where different states are close in energy and have similar characteristics, however it would not be reliable in other regions. A second type of coupling would be describing each state separately, with the environment polarized accordingly. But in this case the computational cost would be much higher and the states would no longer be orthogonal. A third strategy, possibly more reliable, polarizes the environment according to the state that is used to propagate the dynamics. This is computationally simple and retrieves vertical transitions from the selected state to the others. However, it would cause discontinuities at the hops because of the sudden change in the description of the environment.

To conclude this section, it is worth mentioning the strategies that can be used to model nonadiabatic dynamics in multichromophoric systems. Within the formalism of the exciton model mentioned above, it is possible to describe the excitations of a multichromophoric system with reduced computational cost. This has allowed the development of an exciton-model nonadiabatic dynamics framework,[191,192] in which however the exciton couplings are approximated by a dipole interaction. Propagating the dynamics of an exciton model would also require to calculate the analytical derivatives of the exciton couplings.[193] A nonadiabatic exciton model with QM/MM description was developed by Menger et al.,[194] using electrostatic embedding QM/MM. Introducing a polarizable embedding in the nonadiabatic exciton model requires overcoming all the difficulties outlined above. In addition, as the polarizable environment introduces an additional term in the exciton couplings, a correct description of the dynamics would also require the analytical derivatives of this term.

6 Concluding remarks

In this perspective we have given an overview of the state of the art of QM/MM approaches which use a polarizable embedding. The focus has been mostly on the induced point dipole formulation and the analysis which has been presented is mainly applicable to biosystems. We have summarized the most recent developments but also underlined the still present limitations. From a computational point of view, the main limitation lies in the inherent complexity and computational cost of these methods. As discussed in section 2.1, the self-consistent treatment of polarization requires one to perform several operations that are not required for electrostatic embedding, the most expensive being solving the polarization equations, which has to be done at each SCF step and, for excited states and response properties, at each step of the solution of the response equations. We believe that the use of a linear scaling strategy is the key to address such a limitation. Relying on the FMM, polarizable QM/MM calculations can be performed at a cost that scales linearly with the size of the MM system and that is overall small if compared to the cost of solving the QM equations. This linear scaling implementation will thus allow the use of polarizable QM/MM methods not only for static descriptions but also effectively extended to molecular dynamics simulations. In particular, the combination of enhanced sampling techniques with polarizable QM/MM approaches can

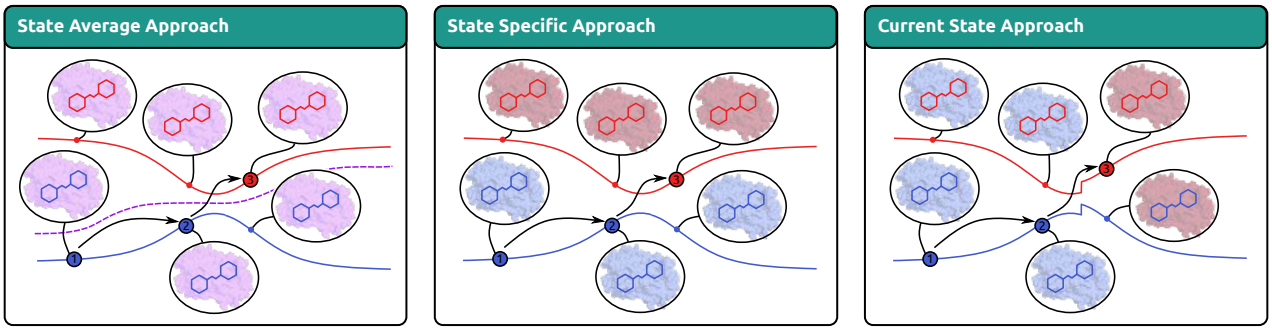

Fig. 4 Schematic representation of the three approaches proposed by the authors to perform non-adiabatic dynamics with PE. A simple SH trajectory on two states (solid red and blue lines) is schematized in three different steps, with an hop occurring between step 2 and 3. The density used for calculating the embedding response is symbolized by the color of the protein in background, while the state calculated for the QM part by the color of the schematized molecule in foreground.

lead to very interesting applications where the PES of a complex system, obtained with high accuracy, can be efficiently explored. This is the most promising strategy in the field of reactive processes in complex biosystems at a new level of accuracy without sacrificing computational efficiency. This extension to dynamics will also have a very important application in the simulation of spectroscopy in complex environments. As a matter of fact, we are confident that polarizable QM/MM might become, in the next decade, the new golden standard for simulating condensed-phase spectroscopy. In the same way, we believe that polarizable embeddings will be largely used to describe excited-state properties and processes. Their extension to non adiabatic dynamics will then come as a natural step for the simulation of complex photophysical processes.

All the applications that we have here presented and discussed, however, require the fulfillment of a fundamental preliminary condition, namely polarizable QM/MM implementations have to be made available in robust, effective and simple-to-use software packages. This is indeed not a simple objective due to the complexity introduced by the mutual polarization between the QM and the MM system. As we have discussed in this perspective, different strategies have been used so far, each with its own advantages and limitations. We hope that in the near future these implementations will become more numerous and higher performing, thus allowing a larger community of users to exploit the advantages of mutual polarization between the QM and the classical subsytems in the simulation of biosystems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

L.C., M.B. and B.M. acknowledge funding by the European Research Council, under the grant ERC-AdG-786714 (LIFETimeS). The authors acknowledge the fundamental contributions given by Carles Curutchet and Jean-Philip Piquemal in the development of the polarizable QM/MM approach here presented and discussed.

Notes and references

1 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
2 H. M. Senn and W. Thiel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 182–187.
3 S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem. Rev.*, 2010, **110**, 949–1017.
4 M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708–2728.
5 A. Warshel, *Angew. Chem. Int. Ed.*, 2014, **53**, 10020–10031.
6 S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes and M. J. Ramos, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2016, **7**, e1281–29.
7 U. N. Morzan, D. J. Alonso de Armiño, N. O. Foglia, F. Ramírez, M. C. González Lebrero, D. A. Scherlis and D. A. Estrin, *Chem. Rev.*, 2018, **118**, 4071–4113.
8 B. Mennucci and S. Corni, *Nat. Rev. Chem.*, 2019, **2**, 1.
9 A. Altun, S. Yokoyama and K. Morokuma, *Photochem. Photobiol.*, 2008, **84**, 845–854.
10 N. H. List, J. M. H. Olsen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20234–20250.
11 E. Boulanger and J. N. Harvey, *Curr. Opin. Struct. Biol.*, 2018, **49**, 72–76.
12 F. Segatta, L. Cupellini, M. Garavelli and B. Mennucci, *Chem. Rev.*, 2019, **119**, 9361–9380.
13 E. Brunk and U. Rothlisberger, *Chem. Rev.*, 2015, **115**, 6217–6263.
14 T. Giovannini, P. Lafiosca and C. Cappelli, *J. Chem. Theory Comput.*, 2017, **13**, 4854–4870.
15 L. V. Slipchenko, M. S. Gordon and K. Ruedenberg, *J. Phys. Chem. A*, 2017, **121**, 9495–9507.
16 T. J. Giese and D. M. York, *J. Phys. Cond. Matter*, 2017, **29**, 1–14.
17 C. Curutchet, L. Cupellini, J. Kongsted, S. Corni, L. Frediani, A. H. Steindal, C. A. Guido, G. Scalmani and B. Mennucci, *J. Chem. Theory Comput.*, 2018, **14**, 1671–1681.
18 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632–672.
19 J. Gao, D. G. Truhlar, Y. Wang, M. J. M. Mazack, P. Löffler, M. R. Provorse and P. Rehak, *Acc. Chem. Res.*, 2014, **47**, 2837–2845.

20 T. A. Halgren and W. Damm, *Curr. Opin. Struct. Biol.*, 2001, **11**, 236–242.

21 A. Warshel, M. Kato and A. V. Pisliakov, *J. Chem. Theory Comput.*, 2007, **3**, 2034–2045.

22 P. Cieplak, F.-Y. Dupradeau, Y. Duan and J. Wang, *J. Phys.: Condens. Matter*, 2009, **21**, 333102–22.

23 Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal and P. Ren, *Annu. Rev. Biophys.*, 2019, **48**, 371–394.

24 E. Boulanger and W. Thiel, *J. Chem. Theory Comput.*, 2012, **8**, 4527–4538.

25 S. Riahi and C. N. Rowley, *J. Comput. Chem.*, 2014, **35**, 2076–2086.

26 C. N. Rowley and B. Roux, *J. Chem. Theory Comput.*, 2012, **8**, 3526–3535.

27 S. K. Sahoo and N. N. Nair, *Front. Chem.*, 2018, **6**, 275.

28 Z. Lu and Y. Zhang, *J. Chem. Theory Comput.*, 2008, **4**, 1237–1248.

29 J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell Jr., *Chem. Rev.*, 2016, **116**, 4983–5013.

30 A. Rappe and W. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.

31 S. W. Rick, S. J. Stuart and B. J. Berne, *J. Chem. Phys.*, 1994, **101**, 6141–6156.

32 F. Lipparini and V. Barone, *J. Chem. Theory Comput.*, 2011, **7**, 3711–3724.

33 P. P. Poier and F. Jensen, *J. Chem. Theory Comput.*, 2019, **15**, 3093–3107.

34 T. Giovannini, A. Puglisi, M. Ambrosetti and C. Cappelli, *J. Chem. Theory Comput.*, 2019, **15**, 2233–2245.

35 M. A. Thompson and G. K. Schenter, *J. Phys. Chem.*, 1995, **99**, 6374–6386.

36 J. Gao, *J. Comput. Chem.*, 1997, **18**, 1061–1071.

37 P. T. Van Duijnen and M. Swart, *J. Phys. Chem. A*, 1998, **102**, 2399–2407.

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

39 D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal and B. Mennucci, *J. Chem. Theory Comput.*, 2016, **12**, 3654–3661.

40 J. M. H. Olsen and J. Kongsted, *Advances in Quantum Chemistry*, Elsevier, 2011, pp. 107–143.

41 J. Dziedzic, Y. Mao, Y. Shao, J. Ponder, T. Head-Gordon, M. Head-Gordon and C.-K. Skylaris, *J. Chem. Phys.*, 2016, **145**, 124106.

42 X. Wu, J.-M. Teuler, F. Cailliez, C. Clavaguéra, D. R. Salahub and A. de la Lande, *J. Chem. Theory Comput.*, 2017, **13**, 3985–4002.

43 J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27–85.

44 P. S. Nerenberg and T. Head-Gordon, *Curr. Opin. Struct. Biol.*, 2018, **49**, 129–138.

45 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

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

47 N. Bernstein, J. R. Kermode and G. Csányi, *Rep. Prog. Phys.*, 2009, **72**, 026501–26.

48 D. Bedrov, J.-P. Piquemal, O. Borodin, A. D. MacKerell, B. Roux and C. Schröder, *Chem. Rev.*, 2019, **119**, 7940–7995.

49 S. Caprasecca, C. Curutchet and B. Mennucci, *J. Chem. Theory Comput.*, 2012, **8**, 4462–4473.

50 S. Caprasecca, S. Jurinovich, L. Lagardère, B. Stamm and F. Lipparini, *J. Chem. Theory Comput.*, 2015, **11**, 694–704.

51 D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2017, **13**, 4025–4033.

52 F. Lipparini, *J. Chem. Theory Comput.*, 2019, **15**, 4312–4317.

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

54 F. Lipparini, G. Scalmani, B. Mennucci, E. Cancès, M. Caricato and M. J. Frisch, *J. Chem. Phys.*, 2010, **133**, 014106.

55 F. Lipparini, L. Lagardère, C. Raynaud, B. Stamm, E. Cancès, B. Mennucci, M. Schnieders, P. Ren, Y. Maday and J. P. Piquemal, *J. Chem. Theory Comput.*, 2015, **11**, 623–634.

56 J. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. Hsieh, H. Lei, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3100.

57 B. T. Thole, *Chem. Phys.*, 1981, **59**, 341.

58 F. Lipparini and B. Mennucci, *J. Chem. Phys.*, 2016, **144**, 160901.

59 F. O. del Valle, M. Aguilar and S. Tolosa, *J. Mol. Struct. Theochem*, 1993, **279**, 223 – 231.

60 F. del Valle and J. Tomasi, *Chem. Phys.*, 1991, **150**, 139 – 150.

61 J. G. Ángyán, *Chem. Phys. Lett.*, 1995, **241**, 51 – 56.

62 K. Sneskov, T. Schwabe, J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2011, **134**, 104108.

63 D. HrÅąak, J. M. H. Olsen and J. Kongsted, *J. Chem. Theory Comput.*, 2018, **14**, 1351–1360.

64 M. Caricato, B. Mennucci, G. Scalmani, G. W. Trucks and M. J. Frisch, *J. Chem. Phys.*, 2010, **132**, 084102.

65 R. Cammi, *J. Chem. Phys.*, 2009, **131**, 164104.

66 F. Lipparini, G. Scalmani and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2009, **11**, 11617–11623.

67 M. Caricato, *J. Chem. Phys.*, 2011, **135**, 074113.

68 M. Caricato, F. Lipparini, G. Scalmani, C. Cappelli and V. Barone, *J. Chem. Theory Comput.*, 2013, **9**, 3035–3042.

69 S. Ren, F. Lipparini, B. Mennucci and M. Caricato, *J. Chem. Theory Comput.*, 2019, **15**, 4485–4496.

70 I. Carnimeo, C. Cappelli and V. Barone, *J. Comput. Chem.*, 2015, **36**, 2271–2290.

71 J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A.

DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

72 W. Wang and R. D. Skeel, *J. Chem. Phys.*, 2005, **123**, 164107.

73 L. Greengard and V. Rokhlin, *J. Comput. Phys.*, 1987, **73**, 325–348.

74 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *GaussianËIJ16 Revision A.03*, 2016, Gaussian Inc. Wallingford CT.

75 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian Development Version Revision J.06+*, 2020, Gaussian Inc. Wallingford CT.

76 M. Scheurer, P. Reinholdt, E. R. Kjellgren, J. M. Haugaard Olsen, A. Dreuw and J. Kongsted, *J. Chem. Theory Comput.*, 2019, **15**, 6154–6163.

77 E. G. Kratz, A. R. Walker, L. LagardÃĺre, F. Lipparini, J.-P. Piquemal and G. AndrÃľs Cisneros, *J. Comput. Chem.*, 2016, **37**, 1019–1029.

78 S. Metz, J. KÃďstner, A. A. Sokol, T. W. Keal and P. Sherwood, *WIRES Computat. Mol. Sci.*, 2013, **4**, 101–110.

79 M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, 1998, **108**, 4439–4449.

80 A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev and M. S. Gordon, *J. Phys. Chem. Lett.*, 2011, **2**, 2184–2192.

81 S. Corni, R. Cammi, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 2005, **123**, 134512.

82 B. Lunkenheimer and A. Köhn, *J. Chem. Theory Comput.*, 2013, **9**, 977–994.

83 M. Caricato, B. Mennucci, J. Tomasi, F. Ingrosso, R. Cammi, S. Corni and G. Scalmani, *J. Chem. Phys.*, 2006, **124**, 124520.

84 H. Schröder and T. Schwabe, *J. Chem. Theory Comput.*, 2018, **14**, 833–842.

85 A. V. Marenich, C. J. Cramer, D. G. Truhlar, C. A. Guido, B. Mennucci, G. Scalmani and M. J. Frisch, *Chem. Sci.*, 2011, **2**, 2143–2161.

86 C. A. Guido, D. Jacquemin, C. Adamo and B. Mennucci, *J. Chem. Theory Comput.*, 2015, **11**, 5782–5790.

87 Lin and J. Gao, *J. Chem. Theory Comput.*, 2007, **3**, 1484–1493.

88 P. Arora, L. V. Slipchenko, S. P. Webb, A. DeFusco and M. S. Gordon, *J. Phys. Chem. A*, 2010, **114**, 6742–6750.

89 Q. Li, B. Mennucci, M. A. Robb, L. Blancafort and C. Curutchet, *J. Chem. Theory Comput.*, 2015, **11**, 1674–1682.

90 R. Guareschi, O. Valsson, C. Curutchet, B. Mennucci and C. Filippi, *J. Phys. Chem. Lett.*, 2016, **7**, 4547–4553.

91 R. Nifosì, B. Mennucci and C. Filippi, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18988–18998.

92 Z.-Q. You and C.-P. Hsu, *Int. J. Quantum Chem.*, 2014, **114**, 102–115.

93 C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294–343.

94 C. König and J. Neugebauer, *ChemPhysChem*, 2011, **13**, 386–425.

95 T. Renger and F. Müh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3348–24.

96 L. Cupellini, M. Bondanza, M. Nottoli and B. Mennucci, *BBA - Bioen.*, 2020, **1861**, 148049.

97 C.-P. Hsu, G. R. Fleming, M. Head-Gordon and T. Head-Gordon, *J. Chem. Phys.*, 2001, **114**, 3065.

98 M. F. Iozzi, B. Mennucci, J. Tomasi and R. Cammi, *J. Chem. Phys.*, 2004, **120**, 7029–12.

99 G. D. Scholes, C. Curutchet, B. Mennucci, R. Cammi and J. Tomasi, *J. Phyis. Chem. B*, 2007, **111**, 6978–6982.

100 C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hossein-Nejad, G. D. Scholes and B. Mennucci, *J. Am. Chem. Soc.*, 2011, **133**, 3078–3084.

101 M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan and G. Mathias, *J. Chem. Phys.*, 2013, **138**, 244103–14.

102 V. Vitale, J. Dziedzic, A. Albaugh, A. M. N. Niklasson, T. Head-Gordon and C.-K. Skylaris, *J. Chem. Phys.*, 2017, **146**, 124115.

103 D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci and J.-P. Piquemal, *Chem. Sci.*, 2019, **10**, 7200–7211.

104 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem.*

Theory Comput., 2018, **14**, 5273–5289.

105 L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder and J.-P. Piquemal, *Chem. Sci.*, 2018, **9**, 956–972.

106 S. Caprasecca, S. Jurinovich, L. Viani, C. Curutchet and B. Mennucci, *J. Chem. Theory Comput.*, 2014, **10**, 1588–1598.

107 M. F. S. J. Menger, S. Caprasecca and B. Mennucci, *J. Chem. Theory Comput.*, 2017, **13**, 3778–3786.

108 E.-S. RiihimÃďki, J. M. MartÃ■nez and L. Kloo, *J. Mol. Struct.: Theochem*, 2006, **760**, 91–98.

109 J. W. Essex and W. L. Jorgensen, *J. Comput. Chem.*, 1995, **16**, 951–972.

110 N. Rega, G. Brancato and V. Barone, *Chem. Phys. Lett.*, 2006, **422**, 367–371.

111 G. Brancato, V. Barone and N. Rega, *Theor. Chem. Acc.*, 2007, **117**, 1001–1015.

112 G. Brancato, N. Rega and V. Barone, *J. Chem. Phys.*, 2008, **128**, 144501.

113 A. H. Steindal, K. Ruud, L. Frediani, K. Aidas and J. Kongsted, *J. Phys. Chem. B*, 2011, **115**, 3027–3037.

114 D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. M. Deneroff, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson and K. J. Bowers, *Comm. Acm*, 2008, **51**, 91.

115 D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan and B. Towles, Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, New York, NY, USA, 2009.

116 C. Abrams and G. Bussi, *Entropy*, 2013, **16**, 163–199.

117 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

118 D. Hamelberg, C. A. F. de Oliveira and J. A. McCammon, *J. Chem. Phys.*, 2007, **127**, 155102.

119 Y. Miao and J. A. McCammon, *Mol. Simul.*, 2016, **42**, 1046–1055.

120 Y. Miao, V. A. Feher and J. A. McCammon, *J. Chem. Theory Comput.*, 2015, **11**, 3584–3595.

121 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12562–12566.

122 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

123 G. Bussi and A. Laio, *Nat. Rev. Phys.*, 2020, **2**, 200–212.

124 L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.*, 2006, **426**, 168–175.

125 R. C. Bernardi, M. C. Melo and K. Schulten, *Biochim. Biophys. Acta - Gen. Subj.*, 2015, **1850**, 872–877.

126 L. Ridder, I. M. C. M. Rietjens, J. Vervoort and A. J. Mulholland, *J. Am. Chem. Soc.*, 2002, **124**, 9926–9936.

127 A. Crespo, M. A. Martí, D. A. Estrin and A. E. Roitberg, *J. Am. Chem. Soc.*, 2005, **127**, 6940–6941.

128 H. M. Senn, S. Thiel and W. Thiel, *J. Chem. Theory Comput.*, 2005, **1**, 494–505.

129 S. R. Billeter, C. F. W. Hanser, T. Z. Mordasini, M. Scholten, W. Thiel and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2001, **3**, 688–695.

130 J. Chandrasekhar, S. Shariffskul and W. L. Jorgensen, *J. Phys. Chem. B*, 2002, **106**, 8078–8085.

131 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.

132 J. E. Basner and S. D. Schwartz, *J. Am. Chem. Soc.*, 2005, **127**, 13822–13831.

133 D. G. Fedorov, Y. Sugita and C. H. Choi, *J. Phys. Chem. B*, 2013, **117**, 7996–8002.

134 M. Schwörer, C. Wichmann, E. Gawehn and G. Mathias, *J. Chem. Theory Comput.*, 2016, **12**, 992–999.

135 F. CÃľlerse, L. LagardÃĺre, E. Derat and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2019, **15**, 3694–3709.

136 The PLUMED consortium, *Nat. Methods*, 2019, **16**, 670–673.

137 B. Mennucci, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6583–12.

138 F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri and V. Barone, *J. Chem. Theory Comput.*, 2012, **8**, 4270–4278.

139 M. Schwörer, C. Wichmann and P. Tavan, *J. Chem. Phys.*, 2016, **144**, 114504.

140 Q. Cui and M. Karplus, *J. Chem. Phys.*, 2000, **112**, 1133–1149.

141 M. Nonella, G. Mathias and P. Tavan, *J. Phys. Chem. A*, 2003, **107**, 8638–8647.

142 C. Herrmann, J. Neugebauer and M. Reiher, *J. Comput. Chem,*, 2008, **29**, 2460–2470.

143 M. Martinez, M.-P. Gaigeot, D. Borgis and R. Vuilleumier, *J. Chem. Phys.*, 2006, **125**, 144106.

144 M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–22.

145 D. Bovi, A. Mezzetti, R. Vuilleumier, M.-P. Gaigeot, B. Chazallon, R. Spezia and L. Guidoni, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20954.

146 M.-P. Gaigeot and M. Sprik, *J. Phys. Chem. B*, 2003, **107**, 10344–10358.

147 C. M. Isborn, A. W. GÃűtz, M. A. Clark, R. C. Walker and T. J. Martínez, *J. Chem. Theory Comput.*, 2012, **8**, 5092–5106.

148 T. J. Zuehlsdorff and C. M. Isborn, *Int. J. Quantum Chem.*, 2019, **119**, e25719.

149 K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin and A. I. Krylov, *Acc. Chem. Res.*, 2012, **45**, 265–275.

150 B. Mennucci, *Int. J. Quantum Chem.*, 2015, **115**, 1202–1208.

151 J. Cerezo, F. J. Avila Ferrer, G. Prampolini and F. Santoro, *J. Chem. Theory Comput.*, 2015, **11**, 5810–5825.

152 J. Cerezo, D. Aranda, F. J. Avila Ferrer, G. Prampolini and F. Santoro, *J. Chem. Theory Comput.*, 2020, **16**, 1215–1231.

153 S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*, Oxford University Press, New York, 1995.

154 S. Valleau, A. Eisfeld and A. Aspuru-Guzik, *J. Chem. Phys.*, 2012, **137**, 224103.

155 S. A. Egorov, E. Rabani and B. J. Berne, *J. Chem. Phys.*, 1998, **108**, 1407–1422.

156 S. Karsten, S. D. Ivanov, S. I. Bokarev and O. Kühn, *J. Chem. Phys.*, 2018, **149**, 194103–14.

157 D. Loco and L. Cupellini, *Int. J. Quantum Chem.*, 2019, **119**, e25726.

158 T. J. Zuehlsdorff, A. Montoya-Castillo, J. A. Napoli, T. E. Markland and C. M. Isborn, *J. Chem. Phys.*, 2019, **151**, 074111.

159 D. Loco, S. Jurinovich, L. Cupellini, M. F. S. J. Menger and B. Mennucci, *Photochem. Photobiol. Sci.*, 2018, **17**, 552–560.

160 M. Bondanza, L. Cupellini, F. Lipparini and B. Mennucci, *Chem*, 2020, **6**, 187–203.

161 M. Persico and G. Granucci, *Theor. Chem. Acc.*, 2014, **133**, 1526–28.

162 R. Crespo-Otero and M. Barbatti, *Chem. Rev.*, 2018, **118**, 7026–7068.

163 T. R. Nelson, A. J. White, J. A. Bjorgaard, A. E. Sifain, Y. Zhang, B. Nebgen, S. Fernandez-Alberti, D. Mozyrsky, A. E. Roitberg and S. Tretiak, *Chem. Rev.*, 2020, **120**, 2215–2287.

164 J. C. Tully, *J. Chem. Phys.*, 1990, **93**, 1061–1071.

165 G. Granucci and M. Persico, *J. Chem. Phys.*, 2007, **126**, 134114.

166 M. Barbatti, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 620–633.

167 J. E. Subotnik, A. Jain, B. Landry, A. Petit, W. Ouyang and N. Bellonzi, *Annu. Rev. Phys. Chem.*, 2016, **67**, 387–417.

168 M. Ben-Nun, J. Quenneville and T. J. MartÃnez, *J. Phys. Chem. A*, 2000, **104**, 5161–5175.

169 B. F. E. Curchod and T. J. MartÃnez, *Chem. Rev.*, 2018, **118**, 3305–3336.

170 J. C. Tully, *Faraday Discuss.*, 1998, **110**, 407–419.

171 M. Vacher, M. J. Bearpark and M. A. Robb, *Theor. Chem. Acc.*, 2016, **135**, 187.

172 B. F. E. Curchod, P. Campomanes, A. Laktionov, M. Neri, T. J. Penfold, S. Vanni, I. Tavernelli and U. Rothlisberger, *Chimia Int. J. Chem.*, 2011, **65**, 330–333.

173 M. Garavelli, B. Frabboni, M. Fato, P. Celani, F. Bernardi, M. A. Robb and M. Olivucci, *J. Am. Chem. Soc.*, 1999, **121**, 1537–1545.

174 A. Toniolo, G. Granucci and T. J. MartÃnez, *J. Phys. Chem. A*, 2003, **107**, 3822–3830.

175 M. Ruckenbauer, M. Barbatti, T. MÃijller and H. Lischka, *J. Phys. Chem. A*, 2010, **114**, 6757–6765.

176 A. Toniolo, S. Olsen, L. Manohar and T. J. MartÃnez, *Faraday Discuss.*, 2004, **127**, 149–163.

177 A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko and T. J. MartÃnez, *J. Phys. Chem. B*, 2009, **113**, 3280–3291.

178 G. Cui and W. Thiel, *J. Chem. Phys.*, 2014, **141**, 124101.

179 I. Tavernelli, B. F. Curchod and U. Rothlisberger, *Chem. Phys.*, 2011, **391**, 101–109.

180 G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark and M. A. Robb, *J. Am. Chem. Soc.*, 2004, **126**, 4228–4233.

181 G. Groenhof, L. V. SchÃď fer, M. Boggio-Pasqua, M. Goette, H. GrubmÃijller and M. A. Robb, *J. Am. Chem. Soc.*, 2007, **129**, 6812–6819.

182 M. BÃűckmann, N. L. Doltsinis and D. Marx, *J. Phys. Chem. A*, 2010, **114**, 745–754.

183 D. Polli, P. AltoÃĺ, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R. A. Mathies, M. Garavelli and G. Cerullo, *Nat.*, 2010, **467**, 440–443.

184 I. Conti, P. AltoÃĺ, M. Stenta, M. Garavelli and G. Orlandi, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5016.

185 J. W. Park and T. Shiozaki, *J. Chem. Theory Comput.*, 2017, **13**, 3676–3683.

186 P. AltoÃĺ, M. Stenta, A. Bottoni and M. Garavelli, *Theor. Chem. Acc.*, 2007, **118**, 219–240.

187 W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell and P. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

188 S. Mai, M. Richter, M. Heindl, M. F. S. J. Menger, A. Atkins, M. Ruckenbauer, F. Plasser, L. M. Ibele, S. Kropf, M. Oppel, P. Marquetand and L. González, *SHARC2.1: Surface Hopping Including Arbitrary Couplings âĂŞ Program Package for Non-Adiabatic Dynamics*, 2019.

189 M. Barbatti, G. Granucci, M. Persico, M. Ruckenbauer, M. Vazdar, M. Eckert-Maksić and H. Lischka, *J. Photochem. Photobiol. A: Chem.*, 2007, **190**, 228–240.

190 M. Barbatti, M. Ruckenbauer, F. Plasser, J. Pittner, G. Granucci, M. Persico and H. Lischka, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **4**, 26–33.

191 A. Sisto, D. R. Glowacki and T. J. Martinez, *Acc. Chem. Res.*, 2014, **47**, 2857–2866.

192 A. Sisto, C. Stross, M. W. van der Kamp, M. O'Connor, S. McIntosh-Smith, G. T. Johnson, E. G. Hohenstein, F. R. Manby, D. R. Glowacki and T. J. Martinez, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14924–14936.

193 A. F. Morrison and J. M. Herbert, *J. Chem. Phys.*, 2017, **146**, 224110.

194 M. F. S. J. Menger, F. Plasser, B. Mennucci and L. González, *J. Chem. Theory Comput.*, 2018, **14**, 6139–6148.