Theor Chem Acc (2007) 118:219–240
DOI 10.1007/s00214-007-0275-9

REGULAR ARTICLE

A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction

Piero Altoè · Marco Stenta · Andrea Bottoni · Marco Garavelli

Received: 29 November 2006 / Accepted: 22 February 2007 / Published online: 28 March 2007

Abstract In the last decade combined quantum mechanic/molecular mechanic (QM/MM) methods have been applied to a large variety of chemical problems. This paper describes a new QM/MM implementation that acts as a flexible computational environment. Specifically, geometry optimizations, frequency calculations and molecular dynamics can be performed on the investigated system that can be split up to three different layers corresponding to different levels of accuracy. Here we report, together with a detailed description of the method and its implementation, some test examples on very different chemical problems, which span the wide and diversified area of chemistry (from ground to excited states topics) and show the flexibility, general applicability and accuracy of the presented hybrid approach. Biochemical, photobiological and supra/super-molecular applications are presented for this purpose: (a) the optimized geometry of a rotaxane is compared with its X-ray structure; (b) the computed absorption spectra of the green fluorescent protein and rhodopsin chromophores in different environments (namely solvent and protein) are compared to the corresponding experimental values and the role of the counter ion and ion pairs in tuning the geometrical and optical properties of charged organic chromophores in polar solvents is explored and discussed; (c) problems and open questions related to the model set-up of a protein are investigated in the framework of the TcPRAC-protein racemase; (d) similarities and differences between the QM and QM/MM reaction path for the HIV1-protease enzymatic mechanism are shown and discussed; (e) the delicate anomeric equilibrium of α- and β-*D*-glucopyranose in water is investigated via QM/MM optimizations and molecular dynamics to show the reliability of the actual implementation in the simulation of solvation effects and delicate balances. Finally, it will be shown that the current implementation (called COBRAMM: Computations at Bologna Relating Ab-initio and Molecular Mechanics Methods) is more than a simple QM/MM method, but a more general hybrid approach with a modular structure that is able to integrate some specialized programs, which may increase the flexibility/efficiency of QM, MM and QM/MM calculations.

Keywords QM/MM · Reaction mechanisms · Solvation · Spectroscopy · Super/supra-molecules · Enzymatic processes

1 Introduction

The study of large molecular systems takes advantage of the use of computational methods based on molecular mechanics (MM) [1,2]. These methods simulate atoms almost as rigid charged spheres; the interactions between atoms are modeled on the basis of chemical connectivity, using simple harmonic potentials (or sometimes more complex functions) to describe bonds, angle bending and torsions. Non-bonded electrostatic and Van der Waals interactions are accounted for on the basis of the charge (or dipole) assigned to the atoms using a simple Coulombic potential and by means of a Lennard–Jones potential (or similar), respectively. The

Contribution to the Fernando Bernardi Memorial Issue.

P. Altoè · A. Bottoni · M. Garavelli (✉)
Dipartimento di Chimica 'G. Ciamician', Universita' di Bologna, Via F. Selmi, 2, 40126 Bologna, Italy
e-mail: marco.garavelli@unibo.it

P. Altoè
INSTM, UdR Bologna, Bologna, Italy

M. Stenta
Dipartimento di Chimica"A. Mangini", Universita' di Bologna, Via San Giacomo, 11, 40126 Bologna, Italy

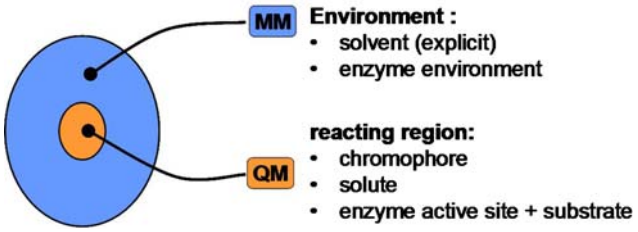


Fig. 1 Schematic representation of the hybrid QM/MM approach

analytic functional form of the equations used to compute energies and forces makes the MM calculation fast even for large molecules. Anyway, the drawback of these methods is their inability in describing processes involving a change of the "nature" of atoms. Thus, chemical reactions (where bonds are broken and formed, and where connectivity and atom-types are not preserved) cannot be described by MM methods. On the other hand, computational methods based on quantum mechanics (QM) [2] have proved to be successful in dealing with such problems because they explicitly treat the electrons (and their couplings) by means of the calculation of the associated wave function.

Unfortunately, the QM methods are much more expensive than MM ones in terms of computational cost and cannot be used to study very large systems. Thus, a problem arises when studying the chemical reactivity of large molecular systems; in response to this need the last three decades were spent on theoretical studies for the development of new computational methods. A promising technique is the partitioning of the whole system (called *real* in the following discussion) in two regions (see Fig. 1): a small part, containing the atoms involved in the chemical process, is described at QM level, while the remaining atoms are treated at MM level, in order to speed-up the calculation and simulate (although at a lower level) the influence of the environment on the reactive core. This hybrid approach is usually called "QM/MM" [3–7]. Anyway, this method has to overcome a difficult technical problem, which is often a source of significant errors and approximations: the correct description of the boundary region. Consequently, great care is addressed to this problem when developing QM/MM methods [6–8], because it strongly influences the capability of reproducing the effects of the surrounding regions on the investigated process.

The wide literature [3–11] on hybrid methods gave us a complete landscape on past attempts to develop a solid and general QM/MM approach. We took advantage of this experience for developing a new versatile QM/MM algorithm and in this paper we describe the details of our implementation, focusing into its advantages as well as its limitations. More generally, our algorithm (called COBRAMM hereafter: Computations at Bologna Relating Ab-initio and Molecular Mechanics Methods) is a hybrid approach that acts as an interface between different programs; these programs execute the single phases of the whole calculation and are allowed to communicate each other by the interface, whilst the partial data are gathered and manipulated to give the final result. Modularity is the main feature of this approach, allowing the user to tailor the computational level by selecting (and combining together) specific programs according to the specific requirements of the investigation. The final purpose being to perform the computational task with the more appropriate (and efficient) computational tools available. A similar approach was used by P. Sherwood and co-workers for ChemShell [12].

To confirm the effectiveness of this implementation, comparative tests are mandatory. The first is a comparison of computed and experimental geometrical parameters for a complex molecular architecture such as that of a rotaxane, whose structure and functionality depends on weak (hydrogen bonds) interactions. Then, it will be shown that the optical properties of two well known and widely studied biological chromophores (i.e., the rhodopsin and green fluorescent protein chromophores) are nicely reproduced both in solution and in the protein. The flexibility and potentiality of our code (namely its capability to perform optimizations as well as molecular dynamics at both QM, MM and QM/MM level) will be exploited for the investigation of the delicate equilibrium in water of the two (α and β) anomers of *D*-glucopyranose. Moreover, we will describe in detail the set-up phase for the study of a catalyzed enzymatic reaction: the enzyme TcPRAC (*Tripanosoma Cruzi* proline racemase) will be taken as a case study to discuss some general aspects inherent to the set-up phase of complex molecular system. Finally, the QM energy profile of the reaction mechanism of HIV1-protease will be compared to the QM/MM one, analyzing and discussing the differences obtained and the reliability of the results; we will show how inclusion of the enzymatic environment is essential for the correct description of the enzymatic process.

It is worth noting that we neither want to present novel chemical applications, nor a detailed review or an exhaustive/conclusive analysis of each problem. Herein, we just want to provide the reader with a bunch of diversified and already well-known chemical issues (spanning from photochemistry to biology, from structural and supramolecular chemistry to condensed phase chemistry, etc.) to show how versatile COBRAMM is and how fruitfully it can be applied, pointing to its flexibility, potentiality and reliability. Under this respect, the applications presented (Sect. 3) pursue a documentary more than an innovative approach.

2 Methods

The computational approach for modeling chemical processes usually involves Geometry Optimizations (Opt) and Molecular Dynamics (MD). Both techniques mainly involve

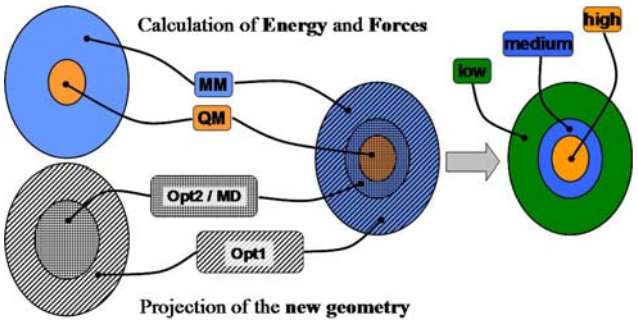


Fig. 2 Graphical representation of the tree-layer scheme adopted in our code

three iterative steps: (1) the calculation of energy E, (2) the calculation of forces **F** (i.e., the first derivatives of the energy with respect to the geometric variables) and (3) the construction of a new geometry on the basis of E and **F** (and using velocities v in the case of MD). This procedure is reiterated until the chosen convergence criteria are fulfilled (in the case of geometry optimizations) or the total simulation time reaches the defined value (in the case of MD).

2.1 The QM/MM partitioning scheme

In order to construct a versatile QM/MM scheme we adopted a modular approach, combining different available codes to exploit their recognized ability in dealing with a specific aspect of the whole problem. Such an approach allows a great flexibility in the partitioning of the system, because the computation of E and **F** on one hand, and the projection of the new geometry on the other, are performed separately at different points of the whole QM/MM calculation (see Fig. 2). More specifically, E and **F** are calculated according to a two-region partitioning of the system, where a QM and an MM calculation are independently performed on the two aforementioned sub-regions. This information is then processed (see below for the details) to obtain the total QM/MM energy and the complete set of forces. These data are then used to generate the new geometry and during this step the system is also partitioned into two independent regions, which do not necessarily coincide with the first ones: the external/larger one (**Opt1**) is handled by means of a fast and rough algorithm, like the Steepest Descents, while the smaller region (**Opt2/MD**) (including the **QM** sub-region) is treated with a more sophisticated optimization algorithm, like BFGS [13], or with a molecular dynamics code based on the Beeman or the Velocity-Verlet algorithms [1].

In our approach the **QM** region is a subset of the **Opt2/MD** region. When **QM** and **Opt2/MD** do coincide, then the system is divided in to two layers called, respectively, *high*, treated at QM level and optimized with the accurate algorithm, and *low*, treated at the MM level and optimized independently using the fast algorithm. When **QM** and **Opt2/MD** do not coincide (i.e., **Opt2/MD** is larger than **QM**), then some MM atoms are involved, together with the QM atoms, in the accurate geometry optimization or MD (i.e., these MM and QM atoms are coupled together). This buffer (intermediate) region of MM atoms, called *medium*, represents a significant improvement and makes the difference with respect to the three layers ONIOM (MO1:MO2:MM) approach from Morokuma [10,11] where the system is divided in to three layers: the inner one (MO1, which is treated at high QM level), the outer (MM, which is treated at MM level) and the intermediate layer (MO2, which is treated at a low QM level, often semiempirical). This scheme is intended to improve the simulation of the electrostatic interaction between QM and the MM regions by interposing a buffer layer treated at a low QM level. In our experience we found that a simpler scheme, like our one, is efficient if an appropriate MM force field is used and if an electrostatic embedding [4,7] scheme (see below) is adopted. In our approach the intermediate region has a different function with respect to Morokuma's approach and is used to improve the reliability and efficiency in the optimization (or molecular dynamic) phase. The capability of handling coupled together with the *high-medium* region (**Opt2/MD**) allows the explicit treatment of large molecular motions around the reactive region without increasing the computational cost, because E and **F** of non-reacting (MM) atoms included in the **Opt2/MD** region are computed at the MM level.

As mentioned above, our scheme uses the electrostatic embedding [4,7] approach to account for the electrostatic influence of the surrounding MM region on the QM layer. Analogously, the effect of charge changes occurring during the chemical process in the *high* (QM) layer is accounted for by using in the MM calculations the QM atomic point charges coming from QM computations. Furthermore, our approach also allows a full independent optimization of the *low* layer (using the fast algorithm) per each optimization step of the *high-medium* layer; this feature nearly resembles the so-called "micro-iteration" [14] technique of ONIOM that, for example, can be useful in the study of a QM solute in a MM solvent. Some authors [15] pointed out the need of taking into account, when studying complex systems like enzymes, the contribution from the protein reorganization energy. We developed the aforementioned optimization scheme to deal with "the linear response of the protein to the movement along the reaction coordinate", with the aim of improving the description of this phenomenon.

The algorithm is intended to deal with all the possible combinations of layers in order to customize calculations according to the specific problem under investigation. The full set of possible combinations actually supported is reported in Table 1; the calculation level is indicated using up to three figures, which are the first figures of the name of the layers used. This notation will be used all through the paper.

Table 1 Calculation types

Calculation type	Layer used		
	high	*medium*	*low*
HML	×	×	×
HM	×	×	
HL	×		×
ML		×	×
H	×		
M		×	

For example, HML refers to a three layers system: this can be useful, for instance, to study the reaction mechanism of an enzymatic process when including a small part of the active site and the substrate in the *high* layer, assigning the remaining part of the active site to the *medium* layer and leaving all the other residues of the enzyme in the *low* layer. An HM calculation can be employed for the accurate optimization, or for the molecular dynamics, of the whole molecular system and allows a great saving in computational resources by treating at the QM level (*high* layer) only the reactive sub-region, while the remaining part of the system is treated at the MM level (*medium* layer). As mentioned above, the HL scheme can give good results for the study of a solute (*high* layer) in a solvent (*low* layer) composed by a large number of explicit molecules. The ML calculation is fully performed at the MM level; it can be used, for example, to obtain a preliminary guess for a critical point on the Potential Energy Surface (PES) associated to the reaction under study. In the H (full QM) calculation the energy and forces evaluation from a code is combined with the optimization (or MD) algorithm of a different code to study, for example, photoinduced ultra-fast processes with QM-MD simulations. The M (fully MM) calculation can be used to get a guess geometry for an H calculation. ML and M calculations can be used only for minima or conformational transition states (with no change in the chemical connectivity); they also give the opportunity to use the force field from an MM code and the optimization or MD algorithm from a different one.

2.2 The boundary region

As stated in "Introduction", handling the boundary between the QM and MM regions need extreme caution, because wrong assumptions can easily lead to unphysical results. In some cases the boundary does not go through a covalent bond: this is the case of a solute (QM level) immersed in a large number of explicit solvent molecules (MM level); this case is very easy to handle and do not need special assumptions. However, in many cases, one cannot avoid passing the QM-MM boundary crosswise one (or more) covalent bond(s), as it is for enzymatic systems. Two strategies have been developed to overcome this problem: (a) the *atom-link* approach [7,16–18] and (b) the modified orbital methods [5,19–21]. We will focus on the former approach because we adopted it to handle boundary regions in our QM/MM method.

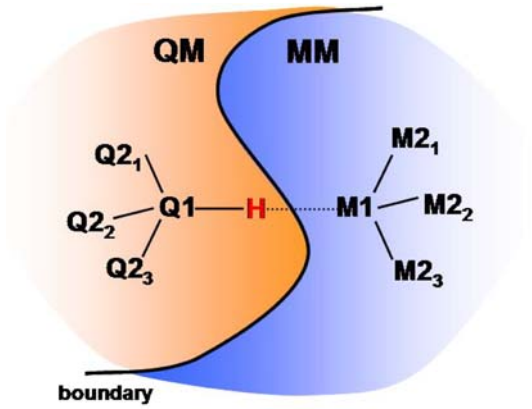


Fig. 3 Boundary region crossing a covalent bond

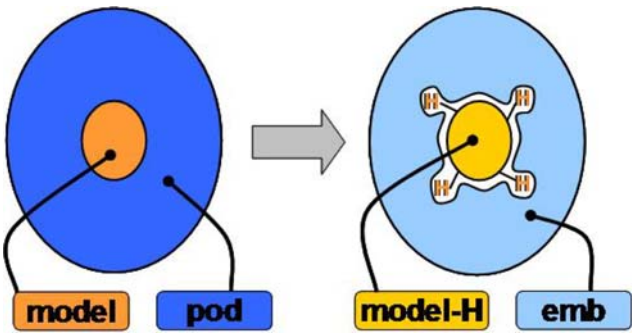


Fig. 4 The atom-link approach implies a change in the original charges (*pod*) of the MM region to be introduced in the QM calculation, according to the electrostatic embedding scheme. The procedure to obtain the new set of charges (*emb*) is described in the text

When the QM-MM boundary cuts a covalent bond, then the valence of the frontier QM atom (Q1 in Fig. 3) remains unfilled; we call this dangling QM structure as *model* and we can observe that it coincides with the *high* layer, as mentioned above. To perform the QM calculation we need to saturate the dangling bond with a "frontier atom". This link atom is usually a hydrogen atom, in which case from *model* we obtain *model-H* (Fig. 4). The *atom-link* approach has, like each boundary treatment, strong advantages as well as weaknesses [7]; is widely used and proved to give good results, is straightforward to understand and it is easy to implement. However, the introduction of artificial atoms that are not present in the original system, can give problems in obtaining the total QM/MM energy value and in geometry optimizations.

2.2.1 The atom-link position

Obtaining the correct QM/MM energy can be difficult because *model-H*, used for QM calculation, does not coincide with the *high* layer, due to the presence of additional hydrogen link atoms and because the MM atomic point charge at the boundary has to be modified [7] (see further sections). We adopted a subtractive approach, similar to the ONIOM scheme [10,11], but not identical to it, because of a different partitioning scheme (see Sect. 3).

To improve the geometry optimization procedure and avoid the artifacts arising from the presence of foreign H atoms, we tried to eliminate their extra degrees of freedom by making their position dependent on the position of the Q1 (QM) and M1 (MM) atoms. Thus, during a geometry optimization or a molecular dynamics the position of the hydrogen link atom is reassigned at each step and is chosen to be on the line connecting Q1 an M1 atoms, removing two degrees of freedom (see Fig. 3). The last degree of freedom can be eliminated in several ways, but the simplest way is to keep fixed the Q1-*H* distance. Herein, we impose a fixed distance of 1.09 Å, which is a standard value for a C(sp^3)–H bond; in principle, to reduce the error associated with this technique, this choice implies the boundary bond type to be only C(sp^3)–C(sp^3). We are now developing new recipes for selecting different boundary bonds [7]. It is worth noting that the bond itself between Q1 and M1 is calculated molecular mechanically.

2.2.2 Handling QM/MM non-bonding cross-terms

In the presented algorithm, all the energy components of the *high* layer (QM region) are given by a QM calculation on *model-H*; the bonded (stretching, bending, torsions) and non-bonded (electrostatic and Van der Waals) interactions within the *medium* and *low* layers (MM region) are handled at the MM level. Managing the cross terms is more difficult and can be done with different approaches. In our scheme, all the bonding and the Van der Waals terms are treated at the MM level, while electrostatic interactions between the QM and the MM region are computed at the high QM level. For this purpose, an electrostatic embedding [7] scheme was adopted: QM computation are performed on *model-H* surrounded by the atomic point charges of the MM layers (i.e., the *emb* charges). In our notation *pod* and *emb* are two subsets of atomic point charges of the MM regions. The *pod* set contains the unmodified atomic point charges from the MM force field, while the *emb* set of charges do differ form the *pod* one (see Fig. 4) only at the boundary region. Indeed, it has been shown that *pod* charges, when coupled to the *atom-link* method, lead to an unrealistic hyperpolarization of the wave function at the Q1-*H* bond. This unphysical effect is mainly due to the presence of the atomic point charge on the M1 atom near the hydrogen atom link (see Fig. 3). To avoid this problem, the charge on M1 is redistributed (i.e., it is spread) on the neighbouring MM atoms (M2) while it is set to zero on M1, so that the total charge of the system is preserved. It is worth to specify that the charge is not equally redistributed on the M2 atoms, but the added fractional charge is proportional to the module of the original MM charge hosted on that atom. We realized that this charge redistribution scheme gives better results if compared to other strategies, like the zeroing of all the charges on the M2 atoms. Anyway, based on recent publications [4,7], we are now trying to improve this redistribution scheme in order to preserve the polarity of the original Q1–M1 bond as well as the one of the M1–M2 bonds; this will be particularly important for the study of systems in which the frontier does cut bonds other than the standard (substantially non-polar) C(sp^3)–C(sp^3) ones.

2.3 QM/MM formalism

The QM/MM total energy follows a subtractive scheme [4] and can be written as:

$$E^{\mathrm{tot}} = E_{\mathrm{QM}}^{\mathrm{model-H}} + E_{\mathrm{MM}}^{\mathrm{real}} + E_{\mathrm{QM}}^{\mathrm{el.model-H/emb}} - E_{\mathrm{MM}}^{\mathrm{el.model/pod}} - E_{\mathrm{MM}}^{\mathrm{model-H}} \quad (1)$$

where the first and third terms are calculated at the QM level. $E_{QM}^{model-H}$ is the QM energy of *model-H* in vacuo, while $E_{QM}^{\mathrm{el.model-H/emb}}$ is the electrostatic interaction between *model-H* and the charge embedding (emb). In this way the wave function is perturbed by the surrounding charges, ensuring the polarization of the QM part by the MM environment. This method is usually referred to as electrostatic embedding [7] and, practically speaking, the two $\mathrm{E}_{\mathrm{QM}}^{\mathrm{model-H}}$ and $\mathrm{E}_{\mathrm{QM}}^{\mathrm{el.model-H/emb}}$ terms are computed together via a single QM computation on *model-H* in the bath of emb charges.

The second, fourth and fifth terms of Eq. (1) ($E_{\mathrm{MM}}^{\mathrm{real}}$, $E_{QM}^{\mathrm{el.model/pod}}$, $E_{MM}^{\mathrm{model-H}}$) are MM energies. In particular, they can be divided into the following single contributions:

$$E_{\mathrm{MM}} = E_{\mathrm{MM}}^{\mathrm{real}} - E_{\mathrm{MM}}^{\mathrm{el.model/pod}} - E_{\mathrm{MM}}^{\mathrm{model-H}} = E_{\mathrm{MM}}^{\mathrm{pod}} + E_{\mathrm{MM}}^{\mathrm{VdW_{model/pod}}} + E_{MM}^{\mathrm{bond,bend.,tors.model/pod}} \quad (2)$$

where the MM contribution to the total energy has been split into three terms: the first ($\mathrm{E}_{\mathrm{MM}}^{\mathrm{pod}}$) involves only MM atoms, the second ($E_{MM}^{VdW_{model/pod}}$) and the third ($E_{MM}^{\mathrm{bond,bend.,tors.model/pod}}$) are cross contributions between the QM and MM atoms (Van der Waals and frontier covalent bond terms, respectively).

Two approximations are implied in this description:

- The *atom-link* scheme (i.e., the atom added to saturate the QM part) is a hydrogen link atom (see Sect. 2.1).
- The charges on M1 atoms (i.e., the frontier MM atoms covalently bonded to the QM region) are redistributed to its nearest-neighbors (M2, see Sect. 2.2).

Analytical derivatives of E^{tot} lead to the forces used for the QM part, while the MM region is optimized on the basis of a pure MM calculation. During this later (classical MM) step, the electrostatic interaction between the QM and MM part is computed using point charges derived from the QM wave function, according to the CHELPG [22] or ESP [23,24] schemes. The *atom-link* charge is redistributed over all the QM atoms in order to preserve the total charge (see Appendix 2 for technical details).

2.4 Implementation

A series of standard commercial packages such as MOLCAS [25], Gaussian03 [26], Turbomole [27], ORCA [28,29] (for QM calculations), Amber [30] and Tinker [31] (for the MM part) may be plugged to the developed software. In particular, an efficient strategy resulted when using Gaussian and Tinker algorithms to perform, respectively, geometry optimizations or molecular dynamics on the *high-medium* region, while Amber for MM calculations (i.e., optimization of the *low* layer and evaluation of MM energy and forces), since it allows the treatment of systems whose parameters are not included in many force fields: indeed, the standard Amber force field (*ff03*) [32] is fully compatible with the Generalized Amber Force Field (GAFF) [33] developed to describe almost every organic molecule.

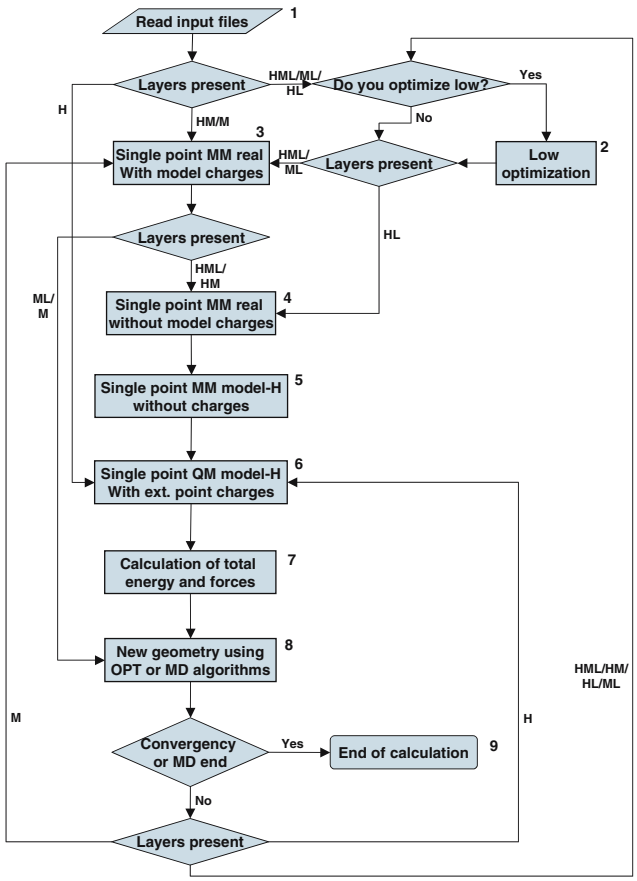
The flowchart of the QM/MM interface is presented in Scheme 1: in particular, points 2, 3, 4 and 5 do need external MM programs to calculate energies and forces, while point 6 uses a QM software to get the electrostatically embedded wave function and its energy and first derivatives. Point 7 merges all these results to give the total energy and gradient. Finally, in the nineth section, an external program for geometry optimization or molecular dynamics is used to generate the new geometry to restart the cycle.

The calculation terminates when the convergence criteria are satisfied or when the time for a MD run is over.

3 Results and discussion

3.1 Optimization of a complex molecular structure: a fumaramide based rotaxane

The supra/super molecules are often characterized by a very high flexibility; in fact, the weak interaction between



Scheme 1 QM/MM implementation flowchart

subunits leaves the system free to move in a quite flat potential energy surface. The investigation of this class of molecules can be very hard in particular using a QM/MM approach where the forces (which drive the research of the minima) are calculated using different methods in different regions of the molecule.

This section has no purpose other than showing the accuracy of the current QM/MM implementation in the description of a structural problem for a delicate supermolecular system, i.e., a rotaxane based on a fumaramidic group [34]. Here, the thread and the ring interact by hydrogen bonds (which is where the QM-MM boundary is passing through) and π stacking (see Fig. 5).

Only the central dimethyl-fumaride has been treated at the QM level (this is the photo-reactive part whose activity must be accounted for by QM methods) using HF/6-31G*, while the rest of the thread and the ring are computed using a classical (MM) force field (GAFF [33]), see Fig. 5a. The main part of the interactions between the thread and the ring comes from QM and MM cross terms.

Fully unconstrained optimizations have been done according to a two layers *high* and *medium* (HM) partitioning scheme.

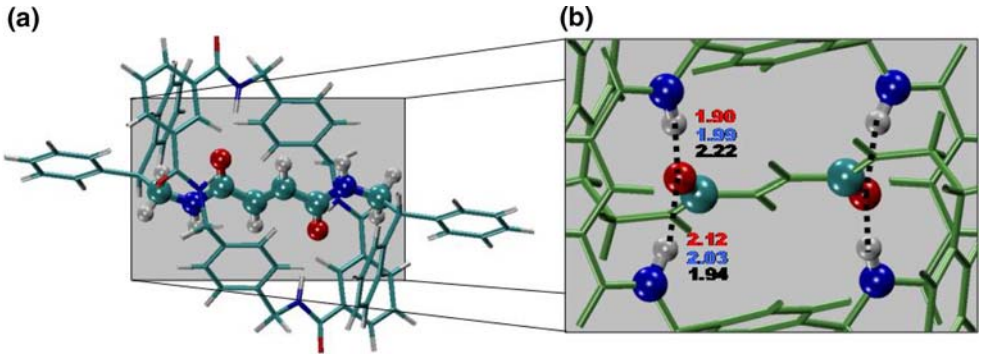


Fig. 5 Rotaxane ground state structure. **a** The dimethyl-fumaramide (*ball and stick*) is treated at the QM level while the rest (tube) is calculated at the MM level. **b** Hydrogen bonds (*dashed lines*) linking the ring to the thread: QM/MM (*red*), mm3 (*blue*) [7] and experimental X-ray (*black*) [35] values

The results are in good agreement with the experimental X-ray [34] structure and previously published MM3 [35] values [34] (Fig. 5b). Specifically, we have selected hydrogen bond distances as a measure of the accuracy; the errors in the hydrogen bond distances between the QM and MM regions are $\leq$0.3 Å.

It can be concluded that this QM/MM approach may be a reliable method for structure determination also in cases where the molecular structure does depend by a subtle balance of factors, and small errors in the forces (such as the ones at the frontier region) may lead to large geometrical variations, like it is the case for this rotaxane.

3.2 The anomeric equilibrium of D-glucopyranose: a delicate example of solvation effects

The simulation of solvation is a crucial problem in general, and in biological systems in particular, since here polar protic solvents (i.e. water) are usually involved, which strongly interact with the solute. Two different computational ways can be used for simulating solvation: (1) the implicit solvent and (2) the explicit solvent model. Polarizable continuum solvent (PCM) [36] for QM calculations and Generalized Born (GB) method [37] for MM calculations are examples of the former, while the latter implies that solvent molecules are explicitly taken into account to obtain the energy of the system. Obviously, the computation of explicit solvent is more time consuming with respect to the implicit approach and can be performed only by considering the solvent molecules at a low (such as MM) level of theory. Thus, a simple way to do this is to use QM/MM computations where the solvent bulk and the solute are treated at the MM and QM levels, respectively.

In this section, the code will be applied to a well known and delicate equilibrium problem: the relative stability of the two anomers of the pyranoid form of D-glucose (i.e. α- and β-D-glucopyranose, see Fig. 6) in water. Rather than focusing on the result itself (which will be shown to go in the right direction, anyway), here we will stress on the flexibility of the implemented algorithm, showing how geometry optimizations and molecular dynamics computations by pure MM or QM/MM techniques may be successfully applied to the aforementioned study.

Many studies [38–41] in the past have focused on this topic revealing all the difficulties in simulating carbohydrates because of the high conformational freedom of their hydroxyl and hydroxymethyl groups; also, the interaction with polar protic solvents, such as water, plays a key role in determining the relative stability of the different isomers, but this is often a subtle and difficult term to estimate. Because of the high computational cost of a complete exploration of the conformational space of this system, we used only a small number of conformers chosen among the most stable ones. We found two minima, on the basis of geometry optimizations at the QM (DFT/B3LYP/6-31G*) [42,43] level, simulating solvent by means of the SCRF-CPCM [44] method. We used these geometries as a guess to create the model system used in the further studies. In all the model systems investigated here the solute molecule (α- or β-D-glucopyranose) is surrounded by a 20 Å radius drop of water (1,490 water molecules): a shell (**shell1**) of 10 Å around the solute is left free to move during the optimization or MD runs, while the water molecules in the outer 10 Å shell (**shell2**) are fixed to their initial geometries (see Fig. 7). Starting solvent geometries are obtained adding the water molecules to the solute by mean of the "leap" module of the AMBER8 program, which uses a standard equilibrated TIP3P [1,2] water library. It is worth noting here that **shell2** prevents molecules of **shell1** to spread out (which would be unphysical), thus acting as a physical barrier during all the calculations.

3.2.1 The L level optimization

We performed full MM minimization using the standard *ff03* [32] force field and a flexible TIP3P model for the water

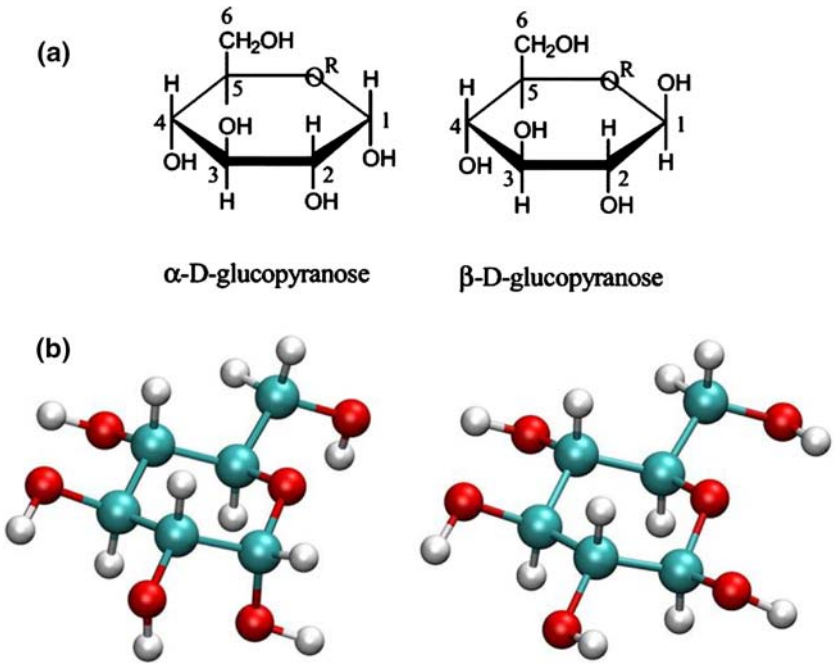


Fig. 6 **a** Haworth and **b** ball and stick representations of the α- and β-D-glucopyranose

molecules and the Glycam [45] parameters for the solute obtaining a $\Delta E = (E_\beta - E_\alpha)$ value of about 21 kcal/mol in terms of internal energy (with α more stable than β), which is in complete disagreement with the experimental value of about −0.4 kcal/mol in terms of Gibbs free energy (the experimental ratio between α and β is 36:64). This result shows that a simple MM optimization cannot reproduce the relative stability of the two anomers and that the force field is not accurate enough to account for solute/solvent interactions.

3.2.2 The HL level optimization

We performed QM/MM calculations with a two layer (*high* and *low*) approach (HL): *high* is constituted by one α- or β-D-glucopyranose molecule treated at the QM (DFT/B3LYP/6-31G*) [42,43] level; *low* contains all the MM solvent molecules (*ff03* force field and flexible TIP3P water model) [32]. After geometry optimization of the solute (*high* layer) and **shell1** (which, together with **shell2**, does represent the *low* layer) we computed the energy difference between the two anomers. The calculations were repeated with different initial configurations of the mobile water shell and we found the two anomers to be almost isoenergetic, with oscillations of the $\Delta E = (E_\beta - E_\alpha)$, in terms of internal energy, between 1.5 and −1.5 kcal/mol. Considering the approximations involved in our approach, this is a quite good result for such a straightforward calculation because it is near to the experimental evidence. Anyway, the relative energy of the two conformers is strongly affected by the conformation assumed by their free –OH groups, which is in turn affected very much by the surrounding solvent molecules. From these simple considerations, it results that optimizations of local minima can only deliver snapshots of the real behavior and that a dynamical treatment could be necessary.

3.2.3 MM molecular dynamics: the ML level

To dynamically explore the interaction between solute and solvent molecules we performed short MD simulations on a full MM system. The *ff03* [32] force field and a flexible TIP3P model was used the for the water molecules and the Glycam [45] parameters for the solute. The outer 10 Å shell (**shell2**) water molecules were assigned the *low* layer (which was kept frozen throughout all the computations), while the inner ones (**shell1**) plus the solute constituted the *medium* layer. The NVT molecular dynamics run of the *medium* layer was performed using the "Beeman" algorithm of the Tinker program with a time step of 1 fs for a total time of 3 ps at the temperature of 298 K. It is worth noting that the implemented code does exploit the AMBER8 package for energy and first derivatives MM computations, while the Tinker program is used for MD: this grants a great flexibility to the procedure since many common MD algorithms do fail in handling a largely frozen system (e.g., our attempt to run a few ps MD run with the standard "sander" program—the molecular dynamics module of Amber8—with the **shell2** kept frozen resulted in the blowing-up of the system). An ML calculation with our code gives the possibility to study large systems keeping frozen all (or part of) the *low* layer. This is possible because the MD run (or the optimization) is performed on the *medium* layer in the presence of the geometry of *low*, which acts as a barrier (with its associated force field) around the

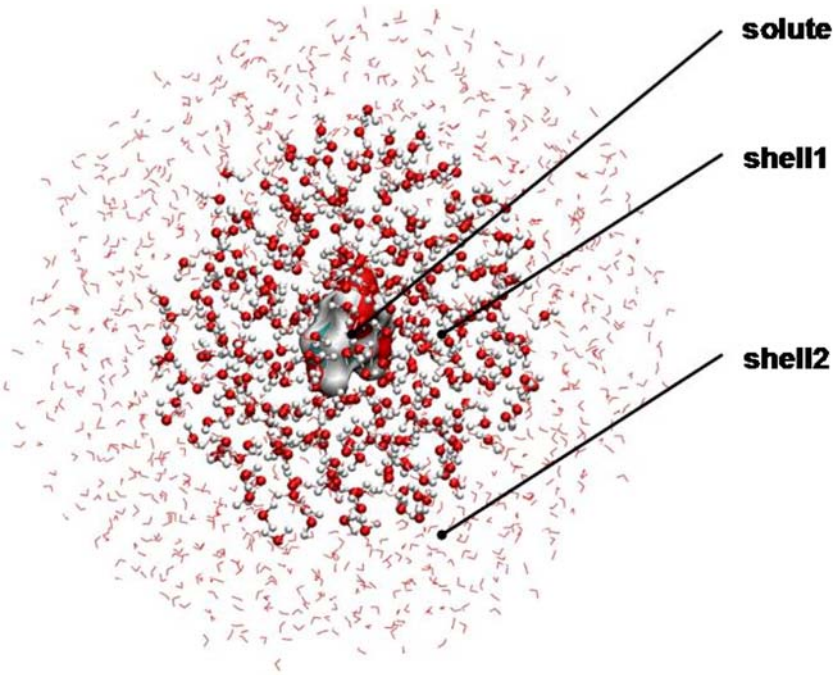


Fig. 7 Schematic representation of the model system used to study the (α- and β-) *D*-glucopyranose (solute) in solution. The solvent molecules (water) are partitioned between shell1 (free to move) and shell2 (kept frozen to the initial geometry)

smaller *medium* layer. The same considerations are valid if the *high* layer is also present.

The equilibration phase covers the first 1,500 steps (1.5 ps) for both anomers (see Fig. 8). From the obtained trajectories it is possible to observe the high mobility of all hydroxyl groups, while the methoxylic group conserves their initial conformation. Averaging the total energy (potential MM energy plus kinetic energy) over the last 2.5 ps we obtain a value for ΔE of 16.34 kcal/mol (see Fig. 8a). This value is very similar to the result obtained from geometry optimization at the MM level (Sect. 2.1) and is still in contrast with experimental evidence calling (unrealistically) for a more stable α anomer.

3.2.4 QM/MM molecular dynamics: the HML level

A three layers approach was used for QM/MM molecular dynamics computations: the solute (D-glucopyranose) was assigned the *high* layer and the QM (DFT/B3LYP/DZVP) level; the **shell1** and **shell2** water molecules were assigned the *medium* (free to move) and the *low* (frozen) layer, respectively, and the MM level (*ff03*) [32] was used. Starting from the pure MM MD results (geometries, velocities, accelerations) described in the previous section, we performed a MD (NVT) run of 3 ps at a temperature of 300 K with a time step of 1 fs using the same modular (AMBER//TINKER) approach as described above (2.3). The system equilibrates in 1.5 ps and the production takes 1.5 ps. By averaging the total energy (potential QM/MM energy plus kinetic energy) over the last 1.5 ps, a value of 6.47 kcal/mol for ΔE was obtained (see Fig. 8b). These results, although not yet in agreement with the experiments, are much better than the MM ones. In the present case we believe that a single and short (3 ps) QM/MM–MD run is inadequate to simulate the overall physical behavior of solvated D-glucopyranose, because the conformational space is not fully explored. The results can be improved by running multiple trajectories, each one starting from a different conformational minimum. However, a better reproduction of the experimental data has already been accomplished in the past [38] and is beyond the purpose of our short discussion. Herein, we wanted only to show an application of this flexible approach to a well-known solvation problem.

3.3 Rhodopsin and GFP chromophores: solvent and protein

The optical properties of an organic chromophore are among the ones that can be highly influenced by the environment (such as the solvent or the protein). This effect is particularly pronounced for charged molecules where, usually, the bright excited state has (at least partially) a charge transfer character. Thus, for example, while the solvent may play a role in stabilizing (by solvation) the ground state of the charged chromophore, it may play in the opposite direction for the excited state. This behavior leads generally to a blue shift of the absorption maximum going from *vacuo* to solvent. For the protein a similar effect may be envisaged.

Two biologically important chromophores belong to this class of compounds: the anionic 4′-hydroxybenzyldiene-2,3-dimethyl-imidazolinone (HBDI, see Fig. 9a) and the proton-

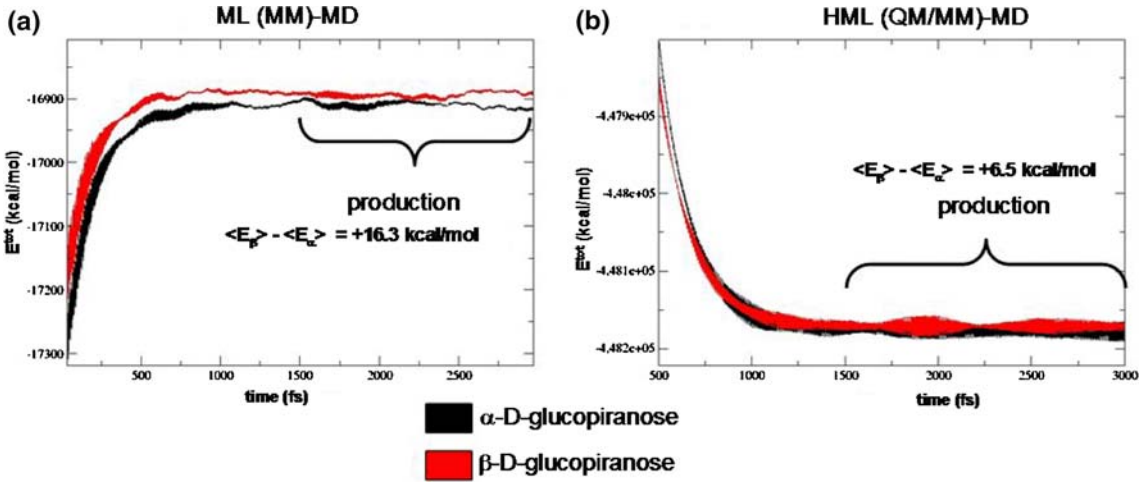


Fig. 8 Molecular dynamics results from **a** the MM system (ML calculation type) and **b** the QM-MM system (HML calculation type). The behavior of the α- (*red line*) and β- (*black line*) anomers is reported, highlighting the production part of each run. The difference between the averaged energies of the two anomers is also reported

ated Schiff base of retinal (RPSB see Fig. 9b), which are the chromophores of the green fluorescent protein (GFP) and rhodopsin proteins, respectively. This section focuses on the description of the optical properties of these systems in different environments by means of QM/MM computations. An advantage is that both experimental and computed spectra are available in vacuo, solvents, as well as protein [46–50]. Thus, it is possible to compare (and validate) our QM/MM results against the ones already in the literature.

3.3.1 Structure and optical properties for solvated HBDI and GFP: flexible QM/MM optimizations

Thanks to its highly fluorescent chromophore and the fact that this is directly formed from a tri-peptide unit (Ser65-Thr66-Gly67) [48] during the protein folding without cofactors, GFP is widely used in biochemistry as a fluorescent marker [48]. The anion is considered to be the active form of the chromophore and, while in protein it strongly fluoresces, in solution the fluorescence is quenched. Therefore, the study of this chromophore in different environments is crucial to understand its photophysical/photochemical properties.

3.3.1.1 Solvation by HL and HML QM/MM computations

HBDI (Fig. 9a) has been chosen as a model for the GFP chromophore because it preserves all the conjugate system as well as the structure of the original protein-embedded chromophore and it has been widely investigated in water solution. QM/MM computations have been used to simulate solvation with the chromophore (anionic HBDI) embedded in a 16 Å TIP3P water box, using an MM Na^+ as the counter ion (as

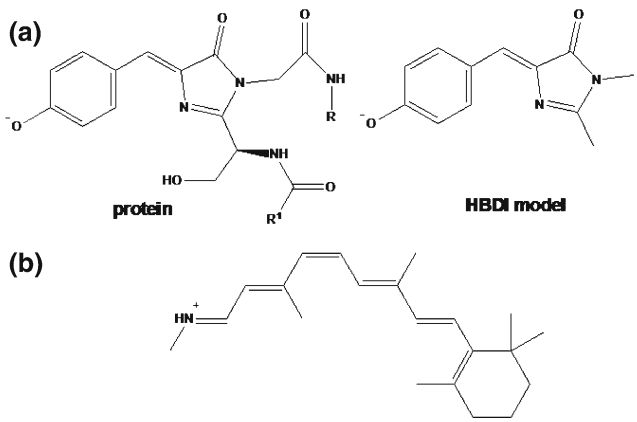


Fig. 9 Chemical structure of the **a** GFP chromophore and its isolated model (HBDI) and **b** the protonated Schiff base of 11-it cis retinal

it is experimentally) [46]. The CASSCF(6-31G*) level has been extensively used for the chromophore during QM/MM optimizations and then the energy has been refined to account for correlation effects using single point CASPT2 computations. This CASPT2//CASSCF approach, which defines the *high* (QM) layer, has been widely used for the photochemistry of organic chromophores (proving to give reliable results of experimental accuracy) [51] and represents the state of the art to calculate excited state properties of organic systems. In order to obtain a good compromise between speed and accuracy, the full active π-space (16 electrons in 14 orbitals) has been reduced (according to our previous paper [52]) to 12 electrons in 11 orbitals, where the orbital localized on the amidic group has been removed together with the highest and lowest energy π-orbitals of the benzene ring.

Table 2 Absorption maxima for the anionic GFP chromophore

Structure	Absorption maximum (nm)	
	QM or QM/MM	Experimental
vacuo	465[a], 493[b],	479[d]
Solvent	434[c]	425[e] – 432[f]
(bound ionic pair)		
Solvent	499 (HL),	
(loose ionic pair)	460 (HML1), 441 (HML2)	
Protein	504, 468[c]	495[g]

(a) ref. [55]; (b) ref. [52]; (c) ref. [51]; (d) ref. [47]; (e) ref. [93]; (f) ref. [46]; (g) ref. [58]

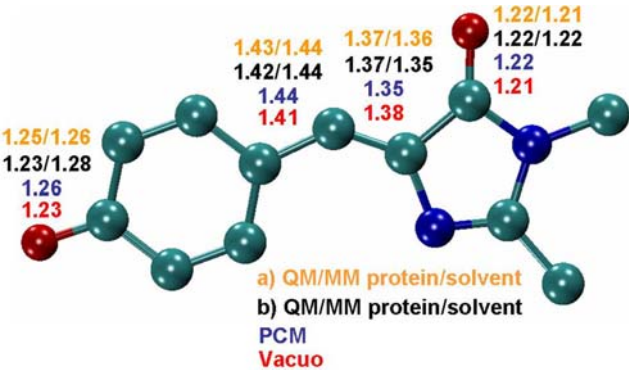


Fig. 10 Bond distance values calculated for the GFP chromophore in vacuo, solvent (PCM and QM/MM) and protein: **a** QM/MM values from ref. [51] and **b** our QM/MM results

To generate a good starting point we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry in vacuo with point charges obtained using the AM1-BCC procedure [53,54], the solvent and counter ion were kept free to move, so that during the 1 ns MD the counter ion is able to diffuse in the water bulk. The lowest potential energy geometry was selected along the trajectory as a representative point of the sampling and as starting guess for QM/MM computations.

Three different QM/MM optimization types were performed on the system according to different and increasing levels of mobility (i.e., the parts comprising the atoms that are free to move and are optimized can be changed according to the partitioning scheme used): (1) HL optimizations, where only the *high* (QM) layer (i.e., the chromophore) is optimized, while all the other (MM) waters (i.e., the *low* layer) are kept frozen; (2) HML optimizations, where the *high* (QM) layer (i.e., the chromophore) and a *medium* (MM) layer (i.e., the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimized together, with all the other (MM) waters (i.e., the *low* layer) kept frozen; (3) HML optimizations, where the *high* (QM) layer (i.e., the chromophore) and a *medium* (MM) layer (i.e., the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimized together, while all the other (MM) waters (i.e., the *low* layer) are free to move (independently with respect to the first two ones) up to a 10 Å radius around the chromophore (and the others are kept frozen to their initial positions). These three different computational levels (which are identified in Table 2 as HL, HML1 and HML2, respectively) highlight the high flexibility of the optimization procedure implemented in the current QM/MM code and, very remarkably, are shown to produce results of increasing accuracy (see Table 2).

It is worth noting that the counter ion (Na^+) is 7 Å away from the chromophore in all the QM/MM optimized structures, revealing a weakly bound ionic pair. Although this is only one of the many possible stable solute/solvent configurations, it is remarkable that the computed absorption energy (499, 460 and 441 nm for (1), (2) and (3), respectively) is progressively getting very close to the experimental value recorded for the chromophore in water (425 nm) and does perfectly account for the blue shift observed on going from vacuo (where, again, the computed 465 nm absorption maximum [51,52] does nicely reproduce the experimental value of 479 nm [46]) to water (see Table 2). Figure 10 reports the relevant bond distances for HBDI in vacuo, explicit solvent and PCM [51,52,55]. It is apparent that the solvent also has a key role in tuning the chromophore molecular structure, since it shifts the geometry to the individual resonance structure carrying the negative charge on the phenolic oxygen, while a resonance hybrid is more likely to represent the chromophore in vacuo. Furthermore, it results that implicit (PCM) as well as explicit (QM/MM) techniques do yield similar results [51,52].

3.3.1.2 GFP by HML QM/MM computations Finally, QM/MM computations have been used to investigate the geometry and absorption maximum of the chromophore in the protein (GFP). The crystallographic structure (code 1GLF) [56] available in the protein data bank archive has been selected. After protonation, using the **H++** [57] procedure, hydrogen atoms have been relaxed in order to get a reliable starting structure. A three-layer (HML) partitioning scheme has been adopted: the *high* (QM) layer (i.e., the electrostatic embedded chromophore, see Fig. 11) and the *medium* layer (see tube representation in Fig. 11) constitute the mobile part (following the receipt of Sinicropi et al. [51]), while all the remaining (MM) protein is kept frozen and represents the *low* layer. The same QM level as for HBDI in solution has been adopted for the chromophore, while protein and water molecules are computed with the Amber force field. Very remarkably, the computed 504 nm absorption maximum for the I anionic state of wild type GFP does nicely reproduce the experimental value of 495 nm [58]

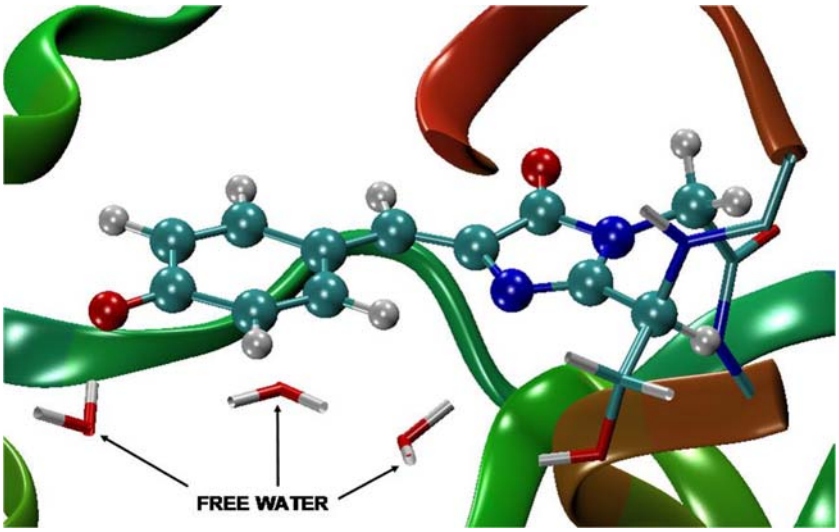


Fig. 11 Model system used for the QM/MM study of the anionic GFP. *Ball* and *stick* is the QM part, the *tube* is the movable medium layer, while the rest is kept frozen

(see Table 2). The same applies also for the chromophore structure (see Fig. 10).

3.3.2 11-cis retinal PSB in solution

A similar QM/MM investigation as before has been performed for solvated RPSB. Specifically, in the present study our effort is focused to simulate the effect of methanol in the spectral tuning of the 11-*cis*-retinal chromophore. Thus, for this purpose, QM/MM computations have been used with the chromophore (RPSB) embedded in a cubic 45 Å methanol box, using an MM Cl^- as the counter ion (as it is experimentally) [50]. Similar to the examples above and previous computational works [59,60], a CASPT2//CASSCF(6-31G*) approach (with electrostatic embedding) has been adopted for the chromophore (using the full π-space of 12 π-electrons in 12 π-orbitals), which defines the *high* (QM) layer, during QM/MM optimizations.

To generate a good starting point for QM/MM optimizations we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry in vacuo [61] with point charges obtained using AM1-BCC procedure [53,54], the solvent and counter ion were kept free to move, so that the counter ion is able to diffuse in the methanol solvent. After a MD run of 1 ns, we see (at least for this specific trajectory) that most of the time the chloride counter ion resides near the retinal nitrogen (i.e., close to the cationic head, in a tight ionic pair arrangement). Thus, a low energy snapshot with the counter ion close to nitrogen has been selected as a representative point of the sampling and as starting guess for QM/MM computations.

A two layer (HL) QM/MM optimization has been performed: the *high* (QM) layer is given by the chromophore (as stated above), while all solvent molecules and the counter ion constitute the *low* (MM) layer (thus excluding a *medium* buffer), which has been left free to move independently up to a distance of 10 Å from the chromophore (while the rest is kept frozen at its initial geometry).

Table 3 Absorption maxima for RPSB

Structure	Absorption maximum (nm)	
	QM or QM/MM	Experimental
Vacuo	545[a]	610[c]
Solvent (bound ionic pair)	453, 429[b]	440[d]
Solvent(loose ionic pair)	442	

(a) ref. [61]; (b) ref. [62]; (c) ref. [49]; (d) ref. [50]

Remarkably, the resulting absorption maximum (453 nm) is in good agreement with the experimental value of 440 nm [50] and with previous QM/MM computations [62] (see Table 3), but in this case the optimized geometry does not undergo significant variations on going from vacuo to solution (see Fig. 12): while in HBDI geometrical modifications come from a different weight of the two resonance structures, in the retinal chromophore the structure does not change a lot, because the geometry is described only by one resonance structure (i.e., the one with the positive charge on nitrogen). This means that the change of the absorption spectrum on going from vacuo to solvent (see Table 3) is addressed only to the perturbation of the wave function by the electrostatic effects of the external point charges and that electrostatic embedding is enough to get a correct description of the chemico-physical properties of the chromophore in solution.

To explore the effects of the counter ion (Cl^-) position on the absorption energy of the chromophore, we have also investigated the optical properties of a loose ionic pair: thus, a low energy snapshot with this arrangement has been selected along the previous MD run as a starting point for new QM/MM optimizations (which have been performed at the

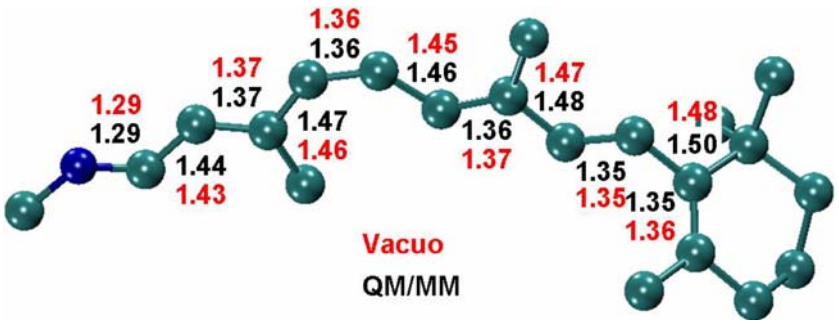
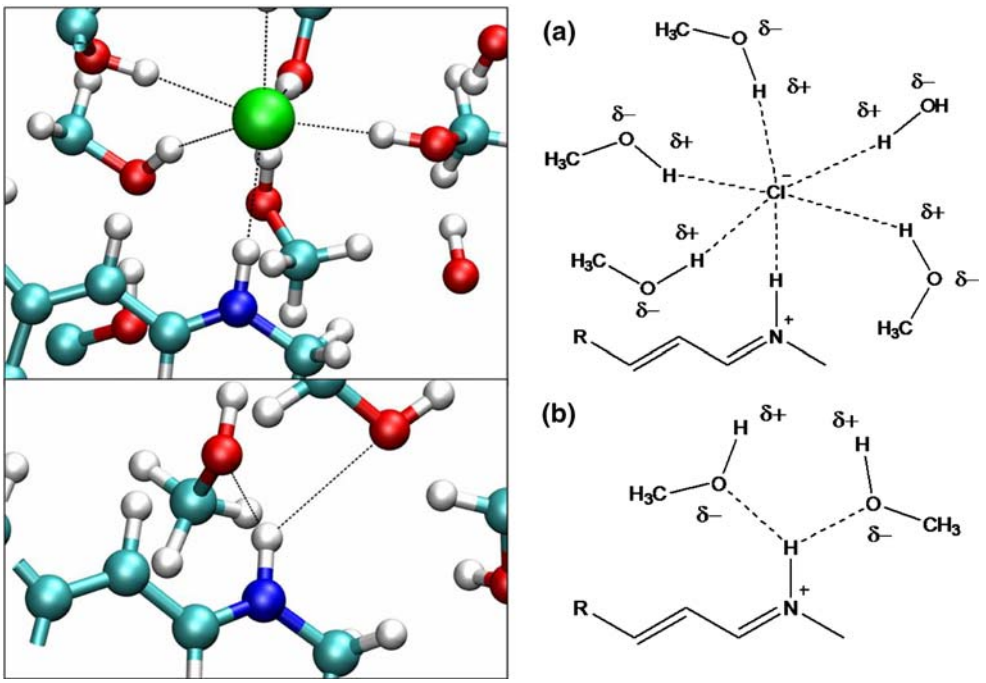


Fig. 12 Bond distances computed for RPSB in vacuo (*red*) and in protein (*black*)



Scheme 2 Schematic representation of interation between solvent, PSB and counter ion: **a** tight bound ionic pair, **b** dissociated couple

same HL level as before). Very interestingly, the absorption maximum computed for this QM/MM optimized loose ionic pair is substantially unchanged (442 nm) (see Table 3) and is in agreement with the experimental value.

3.3.3 Solvated ion pairs: structure and properties

Interestingly, our results for the optimized geometry and the absorption maximum of solvated HBDI (Sect. 3.1) match perfectly (see Fig. 10; Table 2) with the ones reported in previous QM/MM computations [51] where an analogous CASPT2//CASSCF approach was employed (see Table 2). Anyway, it is worth noting that there the counter ion was forced to stay at the vicinity of HBDI (i.e., a tight bound ionic pair arrangement was selected), instead of being far away as it is in our case. The same results appear for solvated RPSB, where the close and the loose ion pairs do lead to very similar optical and structural properties (Sect. 3.2). This leads to the remarkable conclusion that the solvent in the loose ionic pair interacts with the solute as if it were the real counter ion, i.e., solvation shells form a *virtual counter ion* that has the same (electrostatic) effect as the real one in a tight bound ionic pair. This effect is due to the reorientation (i.e. polarization) of the polar solvent close to the chromophore (see Scheme 2): the polarized permanent dipoles of the solvent act similarly to the counter ion, stabilizing the ground state with respect to the charge transfer excited state.

This discussion implies that tight bound ionic pairs may be seen as qualitatively good models for solvated charged chromophores in general, as anticipated in our previous works

[59,63]. This study represents a further validation of the aforementioned statement and is in line with previous suggestions by Sakurai and co-workers [64,65].

3.4 Comparison between QM and QM/MM results for an enzymatic reaction mechanism

In this section QM and QM/MM results will be compared. The system under study is the homodimeric protease of the HIV1 virus (HIV1-Pr). This enzyme catalyzes the proteolysis of a long peptide [66–68] synthesized from viral reverse-transcripted DNA and is essential for viral replication. In the last two decades many studies [69–77] have been addressed to understand the catalytic process of this enzyme in order to speedup the rational design of new inhibitors. The enzymatic reaction mechanism involves the Asp25 and Asp25′ residues (each one belonging to one of the two identical subunits of the dimeric enzyme); their side chains are symmetrically disposed in the active site, sharing a proton and a negative charge; they strictly bind a water molecule, forming a low barrier hydrogen bond system (see Fig. 14). Experimental [78–80] and theoretical [66,81] results suggest that the hydrolysis of the amide bond takes place after the formal addition of the water molecule, leading to a metastable gem-diolic intermediate. The aspartic dyad is believed to play the key role in the catalysis by activating both the amidic substrate and the water molecule via a double proton transfer (i.e., carbonylic group protonation and water molecule deprotonation, respectively).

3.4.1 QM reaction profile

To explore the details of the reaction mechanism we performed a pure QM investigation of the potential energy surface of a small model system by means of geometry optimizations. We included in this model the catalytic dyad, some nearby fragments, the nucleophilic water molecule, the Phe-Pro dipeptide (as a model of the substrate) and a structural water molecule bounded to the substrate (Fig. 13a). This model system (MOD_QM) was constructed on the basis of the crystallographic pdb data file (1G6L) [82] taking the Cartesian coordinates of the interesting atoms and discarding the rest of the enzyme. Missing hydrogen atoms were added to saturate the valence of dangling bonds. Our previous experience [83,84] on QM studies of enzymatic systems encouraged us to partially consider the (steric) effect of the neglected part of the system by keeping frozen the border atoms of MOD_QM to their crystallographic position (Fig. 13a). This has proven to avoid unrealistic distortions of the model system due to additional degrees of freedom. The MOD_QM model system was investigated at the DFT/B3LYP level and, to reduce the computational cost, we assigned three different basis-sets to the atoms (see Fig. 13b), on the basis of their relevance in the reactive process. We assigned the most accurate basis-set (DZVP) to the atoms of the Asp dyad, to the nucleophilic water molecule and to the peptidic bond of the substrate, because they are involved in the bond breaking/forming processes. The outer atoms were treated with a low level basis set (STO-3G), while for the other atoms we used an intermediate level (3-21G*).

Figure 14a shows a schematic representation of the reaction profile computed for MOD_QM and is briefly discussed here. The reaction path starts with the so-called Michaelis complex (M1a) and after a small rearrangement of the nucleophilic water (TS1a) the rate limiting step takes place between the two minima M2a and M3a (Fig. 14a). The transition state TS2a corresponds to a concerted double proton transfer and the nucleophilic attack of the water molecule to the carbonyl of the amidic group (see Fig. 14c). Thus, water is activated (enhancing its nucleophilicity) by Asp25′ deprotonation, while a concurrent protonation of the amidic carbonyl oxygen by Asp25 increases electrophilicity of the amidic carbon atom. The M3a minimum corresponds to a diolic intermediate with both hydroxylic groups that are H-bonded to the same unprotonated carboxylic group of the Asp dyad. The other steps from M3a to M4a correspond to structural rearrangements and proton displacements needed to complete the amide bond hydrolysis. The overall reaction path is apparently in good agreement with the observed behavior of the enzyme (the reaction is spontaneous at room temperature), being the simulated process exo-ergonic (about 8 kcal/mol). Moreover, the taller energy barrier corresponds to the experimental value [79,80] of the rate limiting step (i.e., formation of the diolic intermediate).

3.4.2 QM/MM reaction profile

A QM/MM model system (MOD_QMMM) is then used both to test our QM/MM code and to see the effects of the environment on the previous results. The system is constructed on the basis of the crystallographic coordinates used for the QM model (pdb code:1G6L); the protonation state is assigned by means of the **H++** program [57], except for the Asp residues of the dyad [76,77]. The substrate is simulated by an epta-peptide, in which the hydrolyzed bond is the Phe-Pro amidic bond. The system was fully minimized at the MM level using the Amber8 program and the *ff03* [32] force field. Solvation was also simulated by means of the Born generalized method [37]. This represents the starting structure for QM/MM computations, which were performed according to a three layer (HML) partitioning (see Fig. 15): the *high* layer was chosen to coincide with MOD_QM (also the basis-set is the same); few atoms (backbone only) of the residues near the *high* layer were assigned to the *medium* one; all the other atoms of the system were assigned to the *low* level.

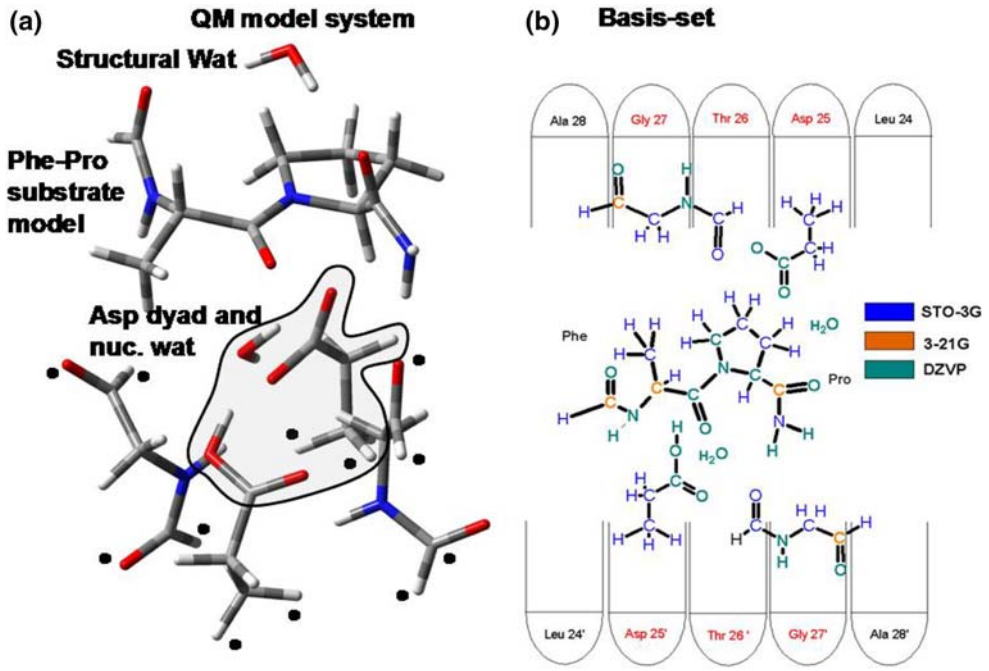


Fig. 13 **a** Model system used for the QM study; the • symbol is used to highlight the atoms kept frozen to their initial positions. **b** The basis set adopted for the atoms of the QM system is shown using different colors. The *model-H* region used in QM/MM computations do coincide with the QM model system and the basis set adopted is the same

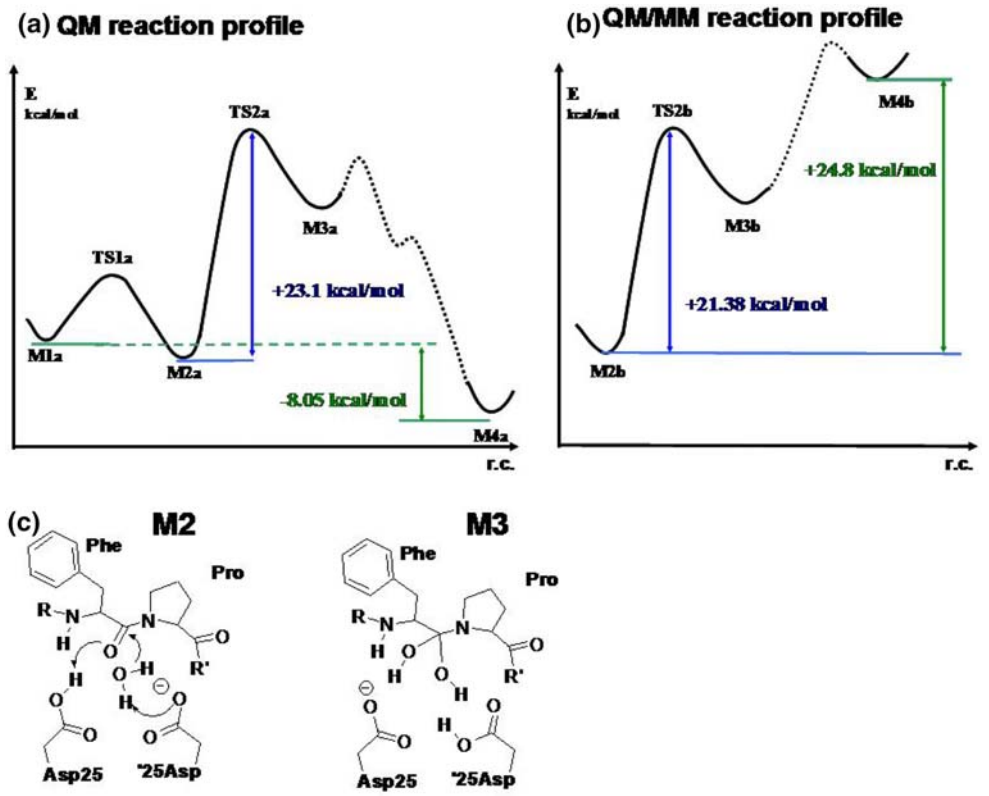


Fig. 14 **a** Pure QM and **b** QM/MM reaction profiles. **c** Schematic representation of the M2 and M3 species; the two Asp residues of the aspartic diad (Asp25 and Asp25′) belong to two identical chains

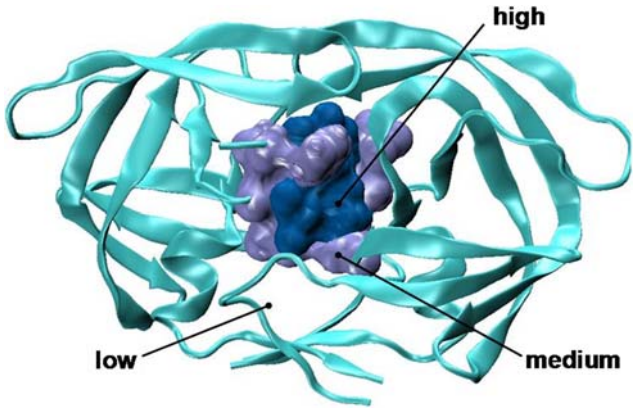


Fig. 15 Model system used to study the QM/MM reaction profile of the enzyme HIV1-Pr

The residues of the *low* layer were kept frozen to their initial position because they are largely exposed to the surface and in our simulation solvent molecules are not included.

The QM/MM PES showed interesting deviations from the behavior observed with the pure QM model. The minimization of the Michaelis complex gave a structure similar to the reacting complex M2 of MOD_QM. The first reaction step (21.38 kcal/mol of activation energy) leads directly to the diolic intermediate, which is less stable than the reactant. This value is very similar to the one obtained for the QM model and also the transition state is very similar. Before mapping the whole reaction path we searched for the product, finding the process to be endo-ergonic by about 25 kcal/mol (Fig. 14b).

These results may be accounted for if we consider the structure of the system and the difference between reactant and products. It is well known that the structure of the free enzyme with respect to that of the complex may be quite different. As a matter of fact, when the substrate binds to the open active site it causes a large conformational change on the mobile regions of the enzyme (called "flaps") [85], leading to a closed form. Here, the peptide chain is fixed to the active site by an hydrogen bond network, but when the peptide bond is broken, the lid of the box opens again and the two proteic segments are released to the aqueous media: the lower affinity of the closed form of the enzyme for the hydrolyzed product causes an opposite conformational change on the flaps to give the free enzyme in its open form and the product.

In the QM model the destabilizing interactions between the enzyme and the products are not taken into account properly, since most of the enzyme (like the "flaps") are neglected along the whole reaction profile. Thus, the computed energies only reflect the enthalpic components of the hydrolytic process itself, because all the successive phases (e.g., enzyme rearrangement, etc.) cannot be explored with such a small system. On the other hand, in the QM/MM model most of the enzyme (like the "flaps") is kept frozen (along the whole reaction profile) at the conformation of the starting complex (i.e., the closed form) and the computed energy for the M4 minimum (i.e., the products) has no physical sense, because such a structure is not reliable for the real biological system which would naturally relax by opening the active site via rotation of the "flaps". This dynamical process cannot be accounted for in our QM/MM model and computations correctly predict a high energy value for such a stressed geometry. However, the first part of the path if very similar for both the QM and QM/MM model systems (e.g., the barrier to the formation of the diolic intermediate is almost the same). Thus, we hypothesized that in the early region of the reaction coordinate the destabilizing factors are too low to play a key role, because the substrate backbone is substantially unchanged with respect to the products and the hydrogen bond network (responsible for substrate recognition) is still effective. As the system moves along the reaction coordinate, the breakable peptide bond becomes weaker until the substrate is broken in to two fragments, which immediately go apart. We believe that the new hydrogen bonds configuration can destabilize the complex between enzyme and products, thus preparing the opening of the flaps and the release of the hydrolyzed product [85,86]. The endo-ergonic nature of the computed QM/MM profile may be seen as a clue of that. Under this point of view, this result highlights the importance of the cooperative effect of the surrounding protein environment in triggering the finals steps (i.e., product release) of the reaction.

3.5 The "set-up" problem

As seen above, a QM/MM code is generally based on many assumptions and approximations and not all the molecular system can be successfully studied under these limitations. For example, extreme caution is needed in choosing the QM-MM boundary because most programs (like our own) may give unphysical results if the broken bond is not a $C(sp^3)$–$C(sp^3)$ one. Thus, cutting a bond more polar than an aliphatic one gives problems in the charge redistribution and choosing a Q1 atom that is different from a carbon atom may affect the validity of the hydrogen atom link approach. Some technical features can be adopted to partially solve these problems, but a good idea is to find the best compromise between the right boundary choice and the size of the QM sub-region. This decision can be taken on the basis of simple considerations: (a) we can extend the QM frontier until the boundary satisfies the conditions above, but one have to remember the proportionality between the size of the QM region and the cost of its calculation; (b) the better system should have a small localized reactive region and some

aliphatic QM-MM boundary bonds near (but not in contact with) the reactive core.

Another important issue concerns the availability of MM parameters for the system under investigation, including the QM part. In fact, given the subtractive approach used for energy computation [see Eq. 1], all QM-MM bonded and non-bonded cross interaction terms (except for the electrostatic effects for which the electrostatic embedding procedure is adopted) are computed at the MM level. Thus, in order to properly describe these QM-MM cross interaction terms, it is important to assign good MM parameters also to the QM atoms, even if the energy and force contributions of the *high* layer are computed by mean of a QM calculation. For example, it is of fundamental importance to have the correct Lennard–Jones parameters for all the (QM) atoms so that the Van der Waals terms are correctly accounted for. Anyway, bonded cross terms (i.e., stretching, bending and torsional contributions) for a bond placed at the frontier involve only up to one, two and three atoms of the *high* (QM) layer, respectively. Thus, only the frontier atoms (and its —one or two bonds— adjacent neighbors) are important in the calculation of bonded cross interactions and MM parameters of bonded terms are not necessary for the other QM atoms. This is important since, while the intimate nature of the atoms of the reactive core may change during the reactive process, frontier atoms are not expected (hopefully) to change, which allows the use of the same MM parameters for these atoms all throughout the reactive process.

On the basis of all these considerations, in order to get meaningful results, it would be preferable to put the frontier at least four bonds from the reacting region, if possible. Unfortunately, this is often hard to do, due to the overwhelming size of the corresponding QM region and one has to make a compromise.

Charges represent another delicate point in system set-up. Since the electrostatic QM-MM cross interactions are calculated at the QM level (i.e., electrostatic embedding), the results may be very sensitive to the charge distribution, even away from the reactive region. Many residues in an enzyme have an acidic or basic character and, according to their protonation state (which is strongly influenced by the close environment), they can carry a net negative or positive charge. This may lead to a system with a net (non-zero) charge. Two main questions arise: is it possible to systematically assign the protonation state of the residues? And what is the best way to deal with a net charge in the system? The answers are not unique and strongly depend on the system under study. In general, the easiest procedure it to manually assign the protonation state of the residues that are part of the active site (by means of some available experimental data, for example), which are the most important ones, while some more general procedures may be used to assign the protonation state of all other residues (e.g., based on the pH of the surrounding media or of the local environment, such as using the **H++** algorithm [57]): they are easy to use but care is needed for the active site region and a supervision of the final result is highly encouraged to correct for some incompetence of the software. Often, after this protonation phase, a net (non-zero) charge exists in the system. To avoid problems in QM/MM calculations some authors have proposed to scale the charges on titratable groups in order to avoid the presence of a net charge on the system. We think this can be a good procedure if these residues do not play an interesting electrostatic effect on the reactive process under study; but it is not a general method. Another procedure, widely used in MM molecular dynamics, is to assign a right number of counterions (usually Na^+ or Cl^-) at the surface of the protein, in order to correct for the charge imbalance and have an electroneutral system. Anyway, this procedure suffers from partiality in placing the counter-ions. In the section below we present an alternative procedure that we have recently developed for the investigation of an enzymatic reaction.

3.5.1 Handling a system bearing a net charge

The treatment of electrostatic interactions within proteins has been extensively reviewed and discussed in some recent papers [87–89] and herein we neither want to present a further comprehensive analysis of this problem nor a method of general validity. Nevertheless, in the enzymatic study reported here particular care and attention must be devoted to properly handle electrostatic effects (due to charged residues) that suggest a novel non-standard approach to the problem. Specifically, in this section we present the preliminary studies (i.e., the model set-up) for the investigation of the reaction mechanism of the TcPRAC [90] enzyme. This is the first eukaryotic proline racemase, which was found in the pathogenic parasite *Trypanosoma cruzi*, responsible of the endemic Chagas' disease. TcPRAC catalyzes the reversible stereoinversion between the *L*-proline and *D*-proline, making no use of cofactors. Two residues of cysteine (Cys130 and Cys300) are involved in the catalytic activity, as proved by mutagenesis experiments [91]. Iodometric titrations also showed that one of these residues is unprotonated bearing a net negative charge. On the basis of recent crystallographic data for an inhibitor/enzyme complex [92], a possible mechanism has been proposed: the two Cys residues can play a concerted general acid–base catalysis, by the deprotonation of the C^α of the *L*-Pro substrate (Cys130) and the following protonation (Cys300) on the opposite face to give the product *D*-Pro. Since the mechanism and its details are still unknown, we decided to investigate the reaction path, in order to discriminate between a concerted or a step-wise pathway and to know if the co-crystallized inhibitor [92] mimics a transition state or a labile intermediate. To set the system up we used the 1W61 pdb data file. We used the **H++** [57] program to add

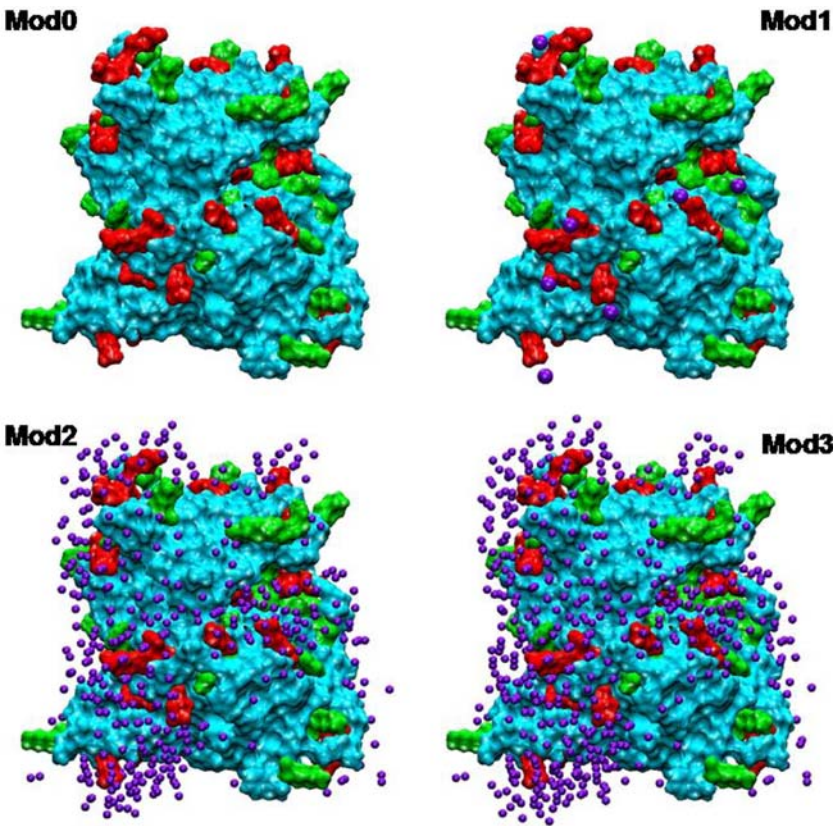


Fig. 16 Distribution of the added point charges (*purple spheres*) around the TcPRAC enzyme. Negatively and positively charged residues are represented, respectively, with *red* and *green* surfaces. In Mod0 the total charge of the enzyme is not counterbalanced by any added charge

missing hydrogens (on the basis of the assigned pH $\approx$ 7.0 the program assigns the protonation state of titratable groups). In order to verify the above-mentioned mechanistic hypothesis, we removed the thiolic hydrogen atom of Cys130, eventually leading to a total charge for the system of -10. Using a tool ("leap") of the Amber8 package, an appropriate number of positive ions was placed to neutralize the overall charge. This program places the counterions in a shell around the enzyme using a grid of Coulombic potential. Three different ways of balancing the 10 negative charges were tried, using (1) 10 Na^+ ions (Mod1), (2) a set of 500 (Mod2) and (3) a set of 1,000 (Mod3) pseudo-ions (with Na^+ Lennard–Jones parameters) whose partial charges were of +0.02 and +0.01, respectively. Finally, also the unbalanced (i.e., not neutralized) -10 charged system (Mod0) was considered for reference (see Fig. 16 Table 4).

The crystallographic structure of the system was refined via MM (*ff03*) [32] optimizations by adding no counterions but using the implicit Born solvation method [37] to simulate the aqueous medium. We then studied the model systems at the QM(DFT/B3LYP/DZVP)/MM(*ff03*) [32] level, assigning to the *high* layer the substrate and the side chains of Cys130 and Cys300 and introducing a *medium* layer to contain the active site; all the remaining residues were assigned to the *low* layer (Fig. 17). The systems were optimized by keeping frozen (at their MM optimized position) all the residues

Table 4 Number and value of the point charges added to the model system used for the QM/MM study of TcPRAC

Model name	Overall charge	Added charges	Charge value
Mod1	0	10	+1
Mod2	0	5,00	+0.02
Mod3	0	1,000	+0.01
Mod0	-10	–	–

on the surface, including the added ions or pseudo-ions. Two minima were identified on the PES for each model system, corresponding to the reactant (ENZ-[*L*-Pro]) and the product (ENZ-[*D*-Pro]) complex called, respectively, M1 and M2. We evaluated the equilibrium between the two minima by computing the relative energy of M1 and M2 for each set-up: the three electro-neutral models plus the charged system.

In Table 5 and Fig. 18 we report the computed $\Delta E_{\mathrm{M2-M1}}$ values. It is noticeable that in all cases M2 is less stable than M1. We think this is a clue for the lower affinity of the enzyme for *D*-Pro and that this can be the basis for the protein conformational changes that allow the final release of the product to the aqueous medium. Under this respect, $\Delta E_{\mathrm{M2-M1}}$ can be correlated to the binding constant between the enzyme and the two enantiomers of proline.

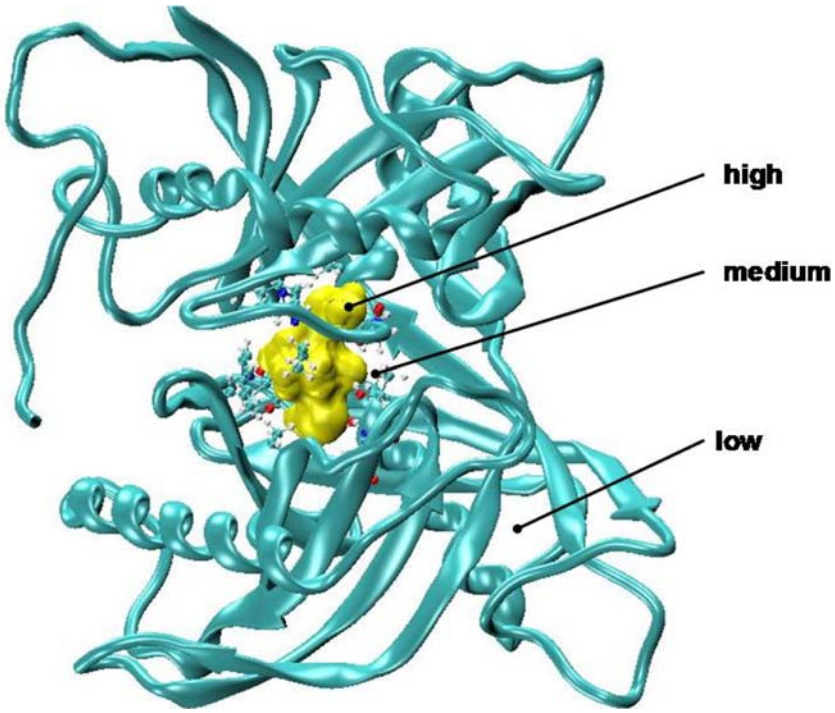


Fig. 17 Model system used to study the reactant–product equilibrium in the TcPRAC enzyme

Table 5 Energy difference between the two minima: $\Delta E_{\mathrm{M2-M1}} = E_{\mathrm{M2}} - E_{\mathrm{M1}}$; E_c represents the electrostatic contribution to ΔE due to the added external charges (ions or pseudo-ions); E_{Enz} represents the contribution from the enzyme itself

Model name	$\Delta E_{\mathrm{M2-M1}}$ $(kcal/mol)$	E_c (kcal/mol)	E_{Enz} (kcal/mol)
Mod1	10.14	7.87	2.27
Mod2	3.37	1.20	2.17
Mod3	2.44	0.46	1.98
Mod0	2.68	0	2.68

It is worth noting that $\Delta E_{\mathrm{M2-M1}}$ decreases (see Table 5) when increasing the number of the added charges to reach a value that is almost identical to the one obtained for the unbalanced system (Mod0). We noticed that the contribution to $\Delta E_{\mathrm{M2-M1}}$ due to the enzyme residues is almost constant in all considered cases, and that the difference is mainly due to the added charges (see Appendix 1 for the details). As we can see in Fig. 18, the effect of point charges decreases when increasing their number (keeping constant the total charge value). This behavior can be explained if we consider the nature of the chemical reaction under study and the method we used to place the charges. As the reaction takes place, passing from M1 to M2, a net charge moves from Cys130 to Cys300 (Fig. 19) thus changing the dipole moment of the molecule; since the ions (which are placed on the basis of the electrostatic potential at the surface) are not allowed to move, they stabilize the initial minimum (M1) better than the other (M2). When using only few point charges (Mod1), the effect of stabilizing the M1 minimum characterized by a net charge near the surface is very strong, but when the number of charges increases (Mod2 and Mod3), and the single value of each charge diminishes, then this stabilizing effect decreases, because the charge cloud is isotropically spread all around the enzyme. For this reason, we preferred to select the unbalanced system (Mod0) for the study of the enzymatic process; in fact we believe the standard method of balancing the net charge of the system using external charges can lead to an error in this case. Of course, an even better model would be to consider the enzyme in a solvent box of explicit molecules with the right number of Na^+ ions fully solvated and equilibrated (i.e., after an appropriate MD run).

Finally, it is worth to point out that here we do not want to present our approach as a generally valid procedure. Anyhow, our data strongly suggest the need of an improved description of medium/long-range electrostatic interactions in order to improve the physical reliability of the simulations in this specific case. This clearly reveals the delicateness of this issue, and an alternative approach to the problem has been shown.

4 Conclusions

A versatile approach for QM/MM computations (i.e., geometry optimizations as well as molecular dynamics) has been illustrated here, which allows up to a three layers partitioning of the investigated molecular system. This partitioning does correspond to different levels of accuracy both in energy/gradients computations and in the optimization (or MD) procedure. It results that the current implementation is a general

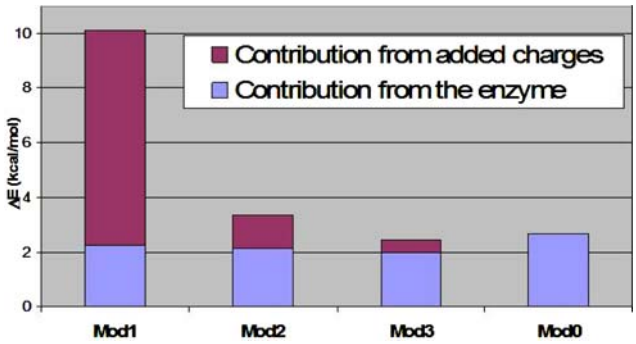


Fig. 18 Bar diagram representing the energy difference between the M2a and the M1a minima for all the models studied. The contribution of added point charges is reported in each case as a red portion of the whole bar

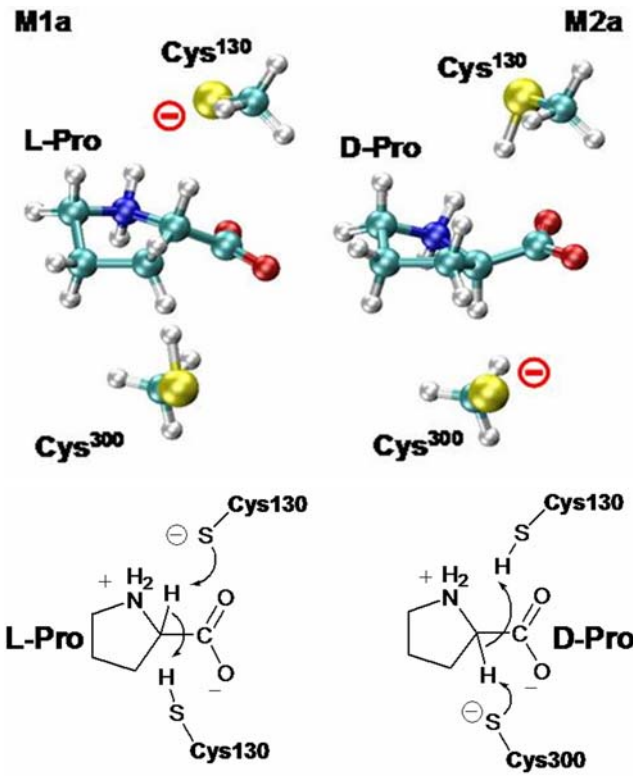


Fig. 19 QM portion (i.e., *model-H*) of the QM/MM model system used for the study of TcPRAC

hybrid approach with a modular structure that is able to integrate some specialized programs, thus increasing the flexibility/efficiency of QM, MM and QM/MM computations.

The flexibility and reliability of this implementation has been supported by several test examples, which span the wide and diversified area of chemistry, from ground to excited states topics: biochemical, photobiological and super/supramolecular applications have been presented for this purpose.

Acknowledgments. Funds for Selected Research Topics, MURST PRIN 2005 (project: "SINTESI E STEREOCONTROLLO DI MOLECOLE ORGANICHE PER LO SVILUPPO DI METODOLOGIE INNOVATIVE DI INTERESSE APPLICATIVO."), FIRB (RBAU01L2HT) and Bologna University (Progetti Strategici d'Ateneo 2005, Project name: CompReNDe) are gratefully acknowledged.

Appendix 1 the direct fingerprint analysis

This analysis is performed at the QM/MM optimized geometries. It gives semi-quantitative information about the capability of a certain residue to stabilize/destabilize a specific critical point with respect to another.

The analysis on two generic critical points A and B is performed by means of a series of single point QM calculations (SPc) on the QM region (i.e., *model-H*, see Fig. 4) of A and B (basing on the optimized QM/MM geometry).

A QM calculation in vacuo on the QM region (i.e., *model-H*) alone gives the energy reference values E_0^{A} and E_0^{B}. Then, N SPc (where N is the total number of residues we want to analyze) are performed, where *model-H* is surrounded by the atomic point charges of the ith residue only, which are placed according to the coordinates of the atoms of the ith residues; this gives NE_i^{A} and NE_i^{B} energy values, as well as the corresponding self energy of the charges e_i^{A} and e_i^{B} (which represent the pure electrostatic contribution among the point charges of the ith residue, only). Since the first term comprises the second, the difference $C_i^{\mathrm{A}} = E_i^{\mathrm{A}} - e_i^{\mathrm{A}} - E_0^{\mathrm{A}}$ and $C_i^{\mathrm{B}} = E_i^{\mathrm{B}} - e_i^{\mathrm{B}} - E_0^{\mathrm{B}}$ gives only the electrostatic (i.e., coulombian) contribution C_i of the ith residue on the QM region (i.e., *model-H*) of A and B, which is the term we want to evaluate (note that non-bonded terms comprise both electrostatic and Van der Waals contributions and, in principle, also the latter should be estimated; nevertheless, Van der Waals terms are far smaller than the former, even for residues close to the QM region and can be neglected). If we compare for each ith residue the values of C_i^{A} and C_i^{B}, we obtain the stability parameter $S_i = C_i^{\mathrm{B}} - C_i^{\mathrm{A}}$ for the ith residue. If $S_i > 0$ (i.e., $C_i^{\mathrm{B}} > C_i^{\mathrm{A}}$) then the ith residue stabilizes A more than B, while if $S_i < 0 (i.e., C_i^{\mathrm{B}} < C_i^{\mathrm{A}})$ the ith residue stabilizes B more than A.

In the specific problem seen above (Sect. 4.1), A and B are represented by the minima M1 and M2. All the ions (or pseudo-ions) are considered as a unique residue, in order to evaluate the electrostatic contribution of all the added charges at a time (purple portion in the bar diagrams of Fig. 18). By difference with the total energy $\Delta E_{\mathrm{M2-M1}}$ we estimate the contribution due to the enzyme residues only (blue portion in the bar diagrams of Fig. 18).

Appendix 2. The construction of the *real* system

A particular care is needed in redistributing charges if the boundary region cuts one or more covalent bonds. Specifically, MM computations need a topology file for the whole system (*real*) that contains the connectivity data and the

atomic point charges assigned on the basis of the adopted force field. The sum of all the atomic charges gives the total charge of the *real* system (Q^{real}), which is necessarily a net charge and must be a constant value all through QM/MM calculations. Thus, before starting calculations, atomic point charges for the whole system atoms have to be assigned and their sum must be a net constant value; but while the charges for the atoms in the MM region will remain constant throughout the course of all QM/MM calculations (at least this is our case, since we do not use a polarizable force field), charges for the QM atoms are generated at each step from the wave function and may change accordingly, which may lead to a variable fractional total charge for the *model* system (note that when the *real* system is split in *model* and *pod* (see Fig. 4), the corresponding total charges, $Q^{\mathrm{model}'}$ and $Q^{\mathrm{pod}'}$, are not necessarily net constant values although it is their sum $Q^{\mathrm{real}} = Q^{\mathrm{model}'} + Q^{\mathrm{pod}'}$). QM calculations are performed on *model-H*, whose net charge $Q^{\mathrm{model-H}}$ and multiplicity are imposed by the user. To prepare a suitable topology file and grant invariability of the system total net charge, in the current implementation we force the total charge of *model* (Q^{model}) to be equal to that of *model-H* by redistributing the charge difference $\Delta Q = Q^{\mathrm{model-H}} - Q^{\mathrm{model}'}$ over the QM atoms (we do that proportionally to the magnitude of each QM atom charge, but alternative strategies are also possible). Thus:

$$Q^{\mathrm{model}'} + \Delta Q = Q^{\mathrm{model-H}} = Q^{\mathrm{model}}$$

Obviously, to conserve the total charge of the system (Q^{real}), also the total charge of *pod* (C^{pod}) has to be modified by redistributing an opposite charge value ($-\Delta Q$) over the MM atoms:

$$\begin{aligned} Q^{\mathrm{pod}} &= Q^{\mathrm{real}} - Q^{\mathrm{model}} = Q^{\mathrm{real}} - (Q^{\mathrm{model}'} + \Delta Q) \\ &= (Q^{\mathrm{real}} - Q^{\mathrm{model}'}) - \Delta Q = Q^{\mathrm{pod}'} - \Delta Q - \Delta Q \\ &= Q^{\mathrm{pod}} - Q^{\mathrm{pod}'} \end{aligned}$$

For *pod* (i.e., the MM atoms) this is done only once, i.e., for the starting geometry during the initial set-up stage and, as stated above, these charges do not change any more. This charge redistribution process may be accomplished according to different strategies since the choice is not unique (mostly depending an the system under study). Here, we can draw some general guidelines on the basis of our experience. When studying an enzymatic system the procedure is quite complex, because all the residues have a net charge. In this case, $-\Delta Q$ is equally redistributed only on the MM atoms of *pod* belonging to the residues crossed by the boundary line (thus leaving a net charge on ach one) and these new atomic point charges of *pod* will not change anymore during QM/MM calculations (i.e., we do not use a polarizable force field); then, *emb* is generated from *pod* at each step (see Fig. 4 and discussion in Sect. 2), to be included in the QM calculation according to the electrostaic embedding scheme [7].

Another crucial point is how to handle charges in the QM region. In many cases the charges of this region are not correctly parameterized in the adopted force field, or they are completely unavailable. In these cases it is necessary to get a guess of the charges to perform the first calculation step and different approaches are possible for accomplishing this task; for example, we found that computing the wave function of *model-H* at the AM1-BCC [53,54] (or a low ab initio) level gives a good starting point. Anyway, during the QM/MM calculation, the wave function of *model-H* is computed at the higher level (including the effect of the electrostatic embedding) and its charges reevaluated. Thus, according to the flowchart of our code (Scheme 1), at the end of each cycle the complete set of charges of *real* is reassembled using for *model* the charges computed at the high QM level. It is worth repeating that QM calculations are performed on *model-H* and the obtained charge includes H link atoms. According to the procedure described above, at every cycle we redistribute the charges of the H link atoms (ΔQ) on the *model* (QM) atoms. Thus, by adopting all these precautions, the total charge of the system (Q^{real}), as well as the total charge of *model* ($Q^{\mathrm{model}} = Q^{\mathrm{model-H}}$) and *pod* ($Q^{\mathrm{pod}}$) is preserved. The atomic point charges of *pod* are preserved all over the QM/MM calculations, while the charges of *model* do change at every step, according to the computed QM electronic distribution.

References

1. Leach AR (2001) Molecular modelling: principles and applications. Pearson Education EMA, UK, pp 1–744
2. Jensen F (1999) Introduction to computational chemistry. Wiley, UK, pp 1–429
3. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
4. Sherwood P (2000) NIC series 3:285–305
5. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
6. Gao J (1995) KB Lipkowitz, DB Boyd (eds) In: Reviews in computational chemistry, VHC Publishers New York, pp 119–185
7. Lin H, Truhlar DG (2006) Theor Chem Acc 117:185–199
8. Gao J, Truhlar DG (2002) Annu Rev Phys Chem 53:467–505
9. Vreven T, Mennucci B, da Silva CO, Morokuma K, Tomasi J (2001) J Chem Phys 115:62–72
10. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357–19363
11. Maseras F, Morokuma K (1995) J Comput Chem 16:1170–1179
12. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, et al. (2003) J Mol Struct 632:1–28
13. Peng C, Ayala PYS, H. Bernhard, Frisch MJ (1996) J Comput Chem 17:49–56
14. Vreven T, Morokuma K, Farkas Ö, Schlegel HB, Frisch MJ (2003) J Comput Chem 24:760–769
15. Klahn M, Braun-Sand S, Rosta E, Warshel A (2005) J Phys Chem B 109:15645–15650
16. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11:700–733
17. Singh UC, Kollman PA (1986) J Comput Chem 7:718–730
18. Ferre N, Olivucci M (2003) J Mol Struct 632:71–82
19. Pu J, Gao J, Truhlar DG (2004) J Phys Chem A 108:632–650
20. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714–4721

21. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) J Comput Chem 15:269–282
22. Breneman CM, Wiberg KB (1990) J Comput Chem 11:361–373
23. Singh UC, Kollman PA (1984) J Comput Chem 5:129–145
24. Besler BH, Jr MKM, Kollman PA (1990) J Comput Chem 11: 431–439
25. Karlstrom G, Lindh R, Malmqvist P-A, Roos BO, Ryde U, et al. (2003) Comput Mater Sci 28:222–239
26. Frisch MJ 2004 Gaussian 03, Revision C.02; Gaussian, Inc., Wallingford CT, 2004.
27. Ahlrichs R, Bar M, Haser M, Horn H, Kolmel C (1989) Chem Phys Lett 162:165–169
28. Neese F. 2006. ORCA An ab initio, DFT and semiempirical SCF-MO package.
29. Sinnecker S, Neese F (2006) J Comput Chem 27:1463–1475
30. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, et al. (2005) J Comput Chem 26:1668–1688
31. Dudek MJ, Ponder JW (1995) J Comput Chem 16:791–816
32. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, et al. (2003) J Comput Chem 24:1999–2012
33. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
34. Gatti FG, Leon S, Wong JKY, Bottari G, Altieri A, et al. (2003) Proc Natl Acad Sci USA 100:10–14
35. Allinger NL, Yuh YH, Lii JH (1989) J Am Chem Soc 111: 8551–8556
36. Cossi M, Barone V (1998) J Chem Phys 109:6246–6254
37. Hawkins GD, Cramer CJ, Truhlar DG (1996) J Phys Chem 100:19824–19839
38. Corchado JC, Sanchez ML, Aguilar MA (2004) J Am Chem Soc 126:7311–7319
39. Miura N, Taniguchi T, Monde K, Nishimura S-I (2006) Chem Phys Lett 419:326–332
40. Momany FA, Appell M, Willett JL, Bosma WB (2005) Carbohydr Res 340:1638–1655
41. Momany FA, Appell M, Willett JL, Schnupf U, Bosma WB (2006) Carbohydr Res 341:525–537
42. Geerlings P, De Proft F, Langenaeker W (2003) Chemical Reviews 103:1793–1873
43. Becke AD (1993) J Chem Phys 98:1372–1377
44. Cossi M, Rega N, Scalmani G, Barone V (2003) J Comput Chem 24:669–681
45. Woods RJ, Dwek RA, Edge C, Fraser-Reid JB (1995) J Phys Chem 99:3832–3846
46. He X, Bell AF, Tonge PJ (2002) J Phys Chem B 106:6056–6066
47. Nielsen SB, Lapierre A, Andersen JU, Pedersen UV, Tomita S, Andersen LH (2001) Phys Rev Lett 87:228102
48. Tsien RY (1998) Annu Rev Biochem 67:509–544
49. Andersen LH, Nielsen IB, Kristensen MB, ElGhazaly MOA, Haacke S, et al. (2005) J Am Chem Soc 127:12347–12350
50. Freedman KA, Becker RS (1986) J Am Chem Soc 108:1245–1251
51. Sinicropi A, Andruniow T, Ferre N, Basosi R, Olivucci M (2005) J Am Chem Soc 127:11534–11535
52. Altoe P, Bernardi F, Garavelli M, Orlandi G, Negri F (2005) J Am Chem Soc 127:3952–3963
53. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) J Comput Chem 21:132–146
54. Jakalian A, Jack DB, Bayly CI (2002) J Comput Chem 23: 1623–1641
55. Martin ME, Negri F, Olivucci M (2004) J Am Chem Soc 126: 5452–5464
56. Feese MD, Faber HR, Bystrom CE, Pettigrew DW, Remington SJ (1998) Structure 6:1407–1418
57. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) Nucleic Acids Res 33:368–371
58. Creemers TMH, Lock AJ, Subramaniam V, Jovin TM, Volker S (1999) Nat Struct Mol Biol 6:557–560
59. Cembran A, Bernardi F, Olivucci M, Garavelli M (2004) J Am Chem Soc 126:16018–16037
60. Vreven T, Bernardi F, Garavelli M, Olivucci M, Robb MA, Schlegel HB (1997) J Am Chem Soc 119:12687–12688
61. Cembran A, Gonzalez-Luque R, Altoe P, Merchan M, Bernardi F, et al. (2005) J Phys Chem A 109:6597–6605
62. Andruniow T, Ferre N, Olivucci M (2004) Proc Natl Acad Sci USA 101:17908–17913
63. Cembran A, Bernardi F, Olivucci M, Garavelli M (2005) Proc Natl Acad Sci USA 102:6255–6260
64. Houjou H, Inoue Y, Sakurai M (1998) J Am Chem Soc 120:4459–4470
65. Houjou H, Sakurai M, Inoue Y (1996) Chem Lett 1075–1076
66. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2000) J Mol Biol 301:1207–1220
67. Davies DR (1990) Annu Rev Biophys Biophys Chem 19:189–215
68. Dreyer GB, Metcalf BW, Tomaszek TA, Carr TJ, Chandler AC, et al. (1989) Proc Natl Acad Sci USA 86:9752–9756
69. Piana S, Bucher D, Carloni P, Rothlisberger U (2004) J Phys Chem B 108:11139–11149
70. Larsson PE, Marti S, Moliner V, Andres J (2004) Abstracts Papers Am Chem Soc 228:U291
71. Hensen C, Hermann JC, Nam K, Ma S, Gao J, Holtje HD (2004) J Med Chem 47:6673–6680
72. Cecconi F, Micheletti C, Carloni P, Maritan A (2001) Proteins: Struct, Funct, Genet 43:365–372
73. Warshel A (1998) J Biol Chem 273:27035–27038
74. Liu H, Muller-Plathe F, van Gunsteren WF (1996) J Mol Biol 261:454–469
75. Lee H, Darden TA, Pedersen LG (1996) J Am Chem Soc 118: 3946–3950
76. Piana S, Sebastiani D, Carloni P, Parrinello M (2001) J Am Chem Soc 123:8730–8737
77. Harte WEB, Jr. David L (1993) J Am Chem Soc 115:3883–3886
78. Rodriguez EJ, Angeles TS, Meek TD (1993) Biochemistry 32:12380–12385
79. Hyland LJ, Tomaszek TA, Roberts JGD, Carr SA, Magaard VW, et al. (1991) Biochemistry 30:8441–8453
80. Hyland LJ, Tomaszek TA, Meek JTD (1991) Biochemistry 30:8454–8463
81. Cavalli A, Carloni P, Recanatini M (2006) Chem Rev 106: 3497–3519
82. Pillai B, K KK, Hosur, VM (2001) Proteins: Struct Funct Genet 43:57–64
83. Bottoni A, Lanza CZ, Miscione GP, Spinelli D (2004) J Am Chem Soc 126:1542–1550
84. Bottoni A, Miscione GP, De Vivo M (2005) Proteins: Struct Funct Bioinf 59:118–130
85. Scott WRP, Schiffer CA (2000) Structure 8:1259–1265
86. Piana S, Carloni P, Parrinello M (2002) J Mol Biol 319:567–583
87. Antoniou D, Basner J, Nunez S, Schwartz SD (2006) Chem Rev 106:3170–3187
88. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM (2006) Chem Rev 106:3210–3235
89. Soriano A, Silla E, Tuñón I, Martí S, Moliner V, Bertrán J (2004) Theor Chem Acc V112:327–334
90. Chamond N, Gregoire C, Coatnoan N, Rougeot C, Freitas LH, et al. (2003) J Biol Chem 278:15484–15494
91. Tonelli RR, Silber AM, Almeida-de-Faria M, Hirata IY, Colli W, Alves MJM (2004) Cell Microbiol 6:733–741
92. Buschiazzo A, Goytia M, Schaeffer F, Degrave W, Shepard W, et al. (2006) Proc Natl Acad Sci USA 103:1705–1710
93. Schellenberg P, Johnson E, Esposito AP, Reid PJ, Parson WW (2001) J Phys Chem B 105:5316–5322