COBRAMM 2.0 – A Software Interface for Tailoring Molecular Electronic Structure Calculations and Running Nano-Scale (QM/MM) Simulations

Oliver Weingart[1‡], Artur Nenov[2,3‡], Piero Altoè[2‡], Ivan Rivalta[4], Javier Segarra-Martí[1†], Irina Dokukina[1] and Marco Garavelli[2,3,4]*

[1]Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität Düsseldorf Universitätsstr. 1, 40225 Düsseldorf, Germany.

[2]Dipartimento di Chimica "G. Ciamician", Universita' degli Studi di Bologna, Via Selmi, 2, I - 40126 Bologna, Italy.

[3]Dipartimento di Chimica Industriale, Università degli Studi di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy.

[4]Université de Lyon, École Normale Supérieure de Lyon, CNRS, Université Claude Bernard Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France.

QM, interface, QM/MM, excited states, dynamics, simulation, software, conical intersection, photochemistry, spectroscopy

We present a new version of the simulation software COBRAMM, a program package interfacing widely known commercial and academic softwares for molecular modelling. It allows a problem driven tailoring of computational chemistry simulations with effortless ground and excited state electronic structure computations. Calculations can be executed within a pure QM or combined quantum mechanical/molecular mechanical (QM/MM) framework, bridging from the atomistic to the nanoscale. The user can perform all necessary steps to simulate ground state and photoreactions in vacuum, complex biopolymer or solvent environments. Starting from ground state optimization, reaction path computations, initial conditions sampling, spectroscopy simulation and photodynamics with deactivation events, COBRAMM is designed to assist in characterization and analysis of complex molecular materials and their properties. Interpretation of recorded spectra range from steady state to time resolved measurements. Various tools help the user to setup the system of interest and analyze the results.

1. Introduction

In the last 40 years, electronic structure calculations have become valuable tools in the interpretation and analysis of chemical reactivity. This development is not solely owed to the vast technological advances in computer hardware, but also to the evolution of efficient software to treat electronic structure problems. Numerous commercial and academic packages are available by now, some of them offering general purpose solutions, and some being more specialized to a specific type of electronic structure methodology. The Gaussian series of programs [1] e.g. cover a large range of electronic structure methods, from semi-empirical to Hartree-Fock (HF) and density functional theories (DFT) as well as post HF methods like configuration interaction (CI), coupled cluster (CC), complete active space self-consistent field (CASSCF) and perturbation approaches (Møller-Plesset MP2-MP4 methods, CASMP2). Gaussian combines these methods with efficient tools for obtaining equilibrium and transition state (TS) geometries, enabling also reaction path following in terms of intrinsic reaction

coordinate (IRC) or minimal energy path (MEP) computations in ground and excited states. The Turbomole suite [2] offers a variety of options for ground state and time-dependent excited state density functional computations, second order MøllerPlesset and explicitly correlated (CC) methods, applying efficiently parallelized and stable algorithms. As an extension to Turbomole, the DFT/MRCI program [3, 4] combines the efforts of density functional theory and multireference schemes, delivering accurate excitation energies for organic and inorganic compounds. Molcas [5] is specialized in the computation of CASSCF, restricted active space SCF (RASSCF), multi-state (MS)-CASPT2 and RASPT2 excited state methods including efficient resolution of the identity (RI) routines and frozen natural orbitals (FNO) [6, 7] schemes. MOLPRO[8] offers advanced options for multi-reference CI computations, equation of motion (EOM) CCSD [9], explicitly correlated methods (F12) [10] and local approximations (e.g. LMP2-F12) [11]. The MNDO package [12, 13] collects various programs for the computation with semi-empirical approaches, enabling also excited state calculations with the specially parameterized OMx-MRCI methods [14, 15]. Hence, at this point the user can select between various types of programs, approaches and functionalities. Often, and especially in the process of finding the suitable method to describe the system of interest, it is necessary to perform comparative studies with two or more programs. The variety of programs with different input structures and options and even different implementations of e.g. geometry optimization routines or combined QM/MM approaches can make this a challenging task. In some cases, certain functionalities are implemented in one specific program, but the electronic structure method to perform the desired task or the specific option e.g. for geometry optimization or (CI) gradient calculation is not available. Our implementation of QM and QM/MM interfaces in COBRAMM [16] aims to close the gaps between the existing, commonly applied electronic structure software by providing a common platform with all necessary tools to setup and analyze pure QM and QM/MM calculations in ground and excited states. COBRAMM extends the capabilities and functionalities of the discussed software by providing specific interfaces also *among* the electronic structure programs themselves. A high level of automatization helps the user to define

standard calculations, still allowing advanced program options for more specific cases. In the following, we document our implementation of QM and QM/MM functionalities in COBRAMM.

2. Methods

2.1 General functionality

COBRAMM allows execution of three main tasks on a geometry given in Cartesian coordinates:

1.) Optimization procedures (including TS search, IRC and MEP computations)

2.) Frequency calculations for asserting the stationary nature of optimized geometries

3.) Molecular dynamics simulations (including surface hopping for RASSCF/RASPT2 and OMx-MRCI methods)

These tasks can be performed within a pure QM or a combined QM/MM environment through the COBRAMM built-in interfaces with various QM software (Table 1), as well as the Amber [17] software for the molecular mechanics part in a QM/MM calculation (details in section 2.5).

For optimization tasks, connection to the very efficient and robust Gaussian Berny-Schlegel optimization routine [18] is provided using the Gaussian "external" communication facilities. This adaption enables usage of most provided facilities therein, i.e. computation of reaction paths (IRC), transition state optimization, coordinate scans. Furthermore, the use of generalized redundant internal coordinates for constrained optimizations is facilitated.

Frequency calculations are performed by numerical differentiation of energy gradients. To this aim, gradients computed at each displaced geometry are processed by Gaussian (through the "external" communication facility) which generates and diagonalizes the Hessian in order to

obtain normal modes and associated frequencies. Within a QM/MM environment, frequency computations are accelerated by performing explicit QM calculations only when QM atoms are displaced, whereas otherwise MM gradients are employed.

Energy gradients and non-adiabatic couplings (required for conical intersection (CoIn) optimizations and non-adiabatic dynamics simulations) can be computed numerically for methods that lack analytical implementation (such as SS- and MS-CASPT2). This is achieved by a built-in routine in COBRAMM. All numerical computations can be performed in a parallel environment using efficient task distribution routines handled by external shell scripts, allowing facile porting to various HPC environments. More details are given in section 2.1.5.

Molecular dynamics simulations can be performed in two flavors: adiabatic and non-adiabatic (employing Tully's fewest switches surface hopping algorithm) with velocity Verlet [19] and Rattle [20] integrator schemes. Their functionalities and implementation will be discussed in section 2.2. Currently, surface hopping is available for CASSCF, SS- and MS-CASPT2 and OMx-MRCI methods, with details in section 2.2.2.

The implementation and functionality of the QM/MM routines are given in section 2.3.

2.1.1 Interfaces to QM programs and Amber

Table 2 lists the current programs that can be used with COBRAMM and the availability of surface hopping and parallel computing options for the listed modules. The COBRAMM set of routines are written in a mostly object-oriented manner and entirely coded in Python (version ≥2.7). Interaction between the subroutines and interfaces is realized using the Python shelve functionality, providing a communication file to store and modify all active variables, vectors and matrices. This strategy enables simple implementation of new interfaces and subroutines. The *cobram* main routine (Table 2) reads the user input files and generates a list of tasks, which

are executed through the central CBF module. The latter decides, which interfaces and further subroutines need to be accessed to perform the requested tasks.

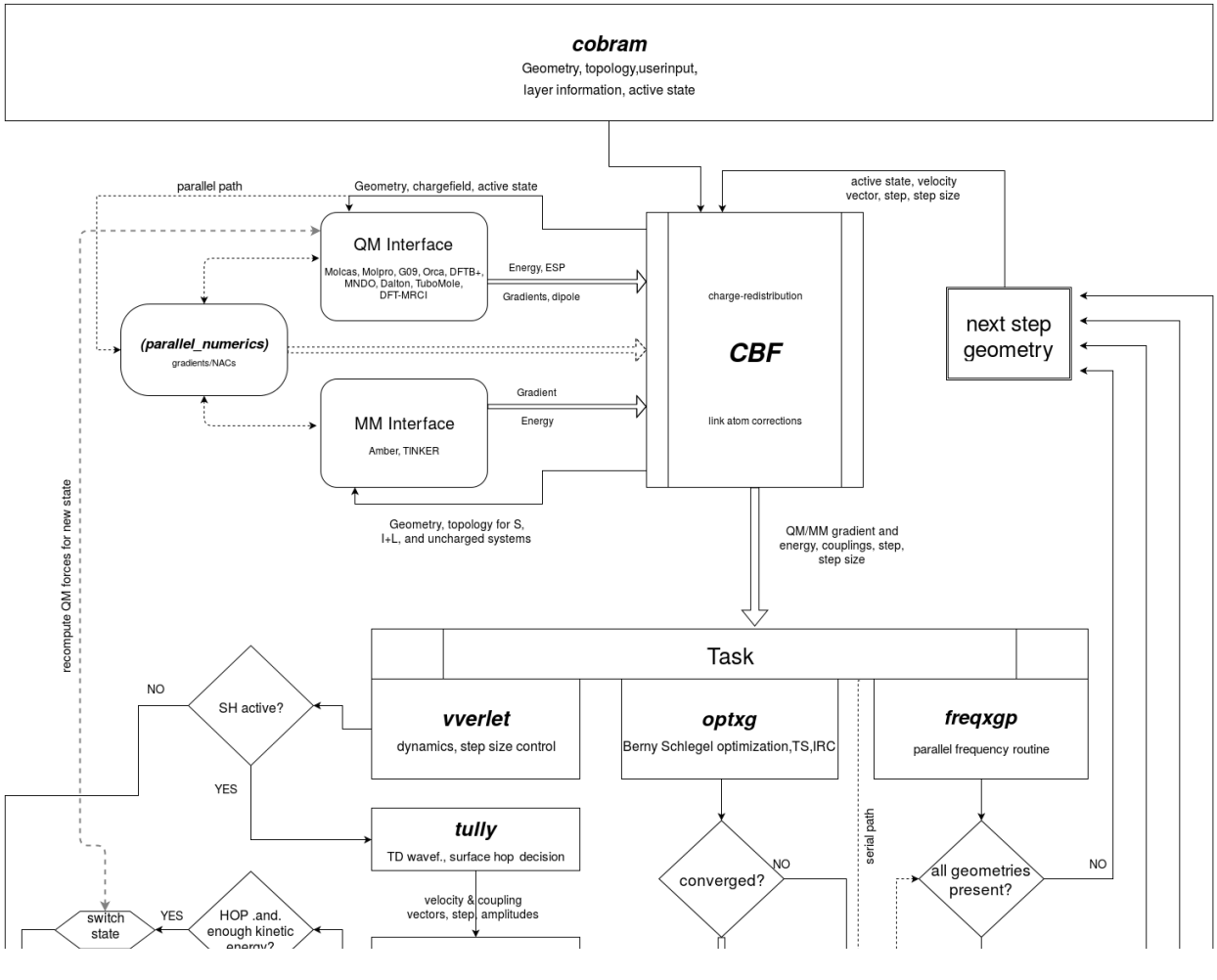
Table 1. Cobramm interfaces to QM and MM programs and functionality

Interface	*Parallel option*	*Berny-Optimizer*	*Velocity Verlet*	*Tully SH*	*Development status*
Gaussian	SMP/Linda®	yes	yes	no	active
Molcas	partly MPI	yes	yes	yes	active
MOLPRO	partly MPI	yes	yes	yes	active
ORCA	MPI	yes	yes	no	basic
DFTB(+)	No	yes	yes	no	basic
Turbomole	SMP/MPI	yes	yes	no	active
DFT/MRCI	SMP/MPI	yes	yes	no	basic
Dalton	No	yes	yes	no	basic
MNDO	No	yes	yes	yes	active
Amber	MPI/GPU	yes	yes	-	active

In the separate QM and MM interfaces, input files for the corresponding programs are created and the programs are executed. Their outputs are analysed and the computed quantities (energies and gradients for MM computations and additionally coupling vectors, dipole moments, ESP fitted charges and point charge gradients for the QM part) are stored and delivered back to the CBF module, where combined QM/MM quantities for gradients, energies and charges are produced. These are, depending on the chosen task, subjected either to the optimizer (optxg) to create a new QM/MM geometry for an optimization, to the dynamics module (vverlet), to compute the next MD geometry and velocities or to the frequency module (freqxg) to produce another elongated geometry for numerical frequency analysis. The flow chart in Scheme 1 provides an overview of the tasks and the data flow between different interfaces and subroutines.

Table 2. Main subroutines and functions

Subroutine/Call	*Function*
Cobram	Main routine, initialisation and task creation
CBF	COBRAMM basic functions, invoke QM and MM interfaces, compute QM/MM energies and gradients, correct link-atoms, distribute charges
freqxg	Numerical frequencies as implemented in Gausssian03/09 via g09 external command
optxg	Interface to Berny-Schlegel G09 algorithm as implemented in Gausssian03/09 via the *external* command
parallel_numerics	Controls the numerical computation of energy gradients, second derivatives and non-adiabatic coupling vectors (parallelized)
product	Matrix and vector products, decoherence correction, velocity scaling
Tully	Tully fewest switches surface hopping algorithm, exponential propagator for time-dependent wavefunction
Vverlet	Velocity Verlet and RATTLE propagators



Scheme 1. Schematic task and data flow in the QM/MM program COBRAMM. Subroutines are denoted in bold italics. For a clear view, explicit communication lines between *CBF*, *cobram* and the individual tasks are not shown.

2.1.2 Optimization Procedures

The required gradients are supplied to Gaussian in Cartesian coordinates through the "*external*" module. MEP computations can be initiated in internal reaction coordinates from the Franck-Condon (FC) point or from a TS. In the former case, the gradient is automatically selected as the direction of motion. At a TS, the normal mode vector associated with the imaginary frequency of a preceding vibrational analysis will be chosen. The force constants required to perform a TS optimization can be either estimated using a valence force field or taken from a previous frequency computation. CoIn optimizations are performed employing Bearpark's gradient projection method [21]:

$$\vec{g} = 2(E_1 - E_2)\frac{\vec{X}_1}{\left|\left|\vec{X}_1\right|\right|} + \mathbf{P}_{IS}(\nabla_R E_2) \tag{1.1}$$

$E_{1/2}$ are the adiabatic energies, X_1 and X_2 are the gradient difference and the derivative coupling vectors, $\nabla_R E_2$ is the gradient on the higher state, while P_{IS} is the projection on the *3N-8* complement orthogonal to the intersection space spanned by X_1 and X_2. The effective gradient is computed inside COBRAMM and forwarded to the Gaussian optimizer. In the case of a QM/MM environment the algorithm includes the effect of the MM atoms on the structure and energy of the CoIn implicitly via the one-electron Hamiltonian (electrostatic embedding) and explicitly through the difference forces acting on the movable MM atoms in the two states subject to optimization. The only approximation is the neglect of structural deformations in the movable MM sub-system, which couple the states as the derivative coupling for the MM part is zero due to the missing QM wavefunction description.

2.1.3 Basis set implementation

The interfaces of COBRAMM with Molcas and Molpro offer the capability of a per-atom basis set definition. In the case of generally contracted basis sets, as e.g. the ANO-S and ANO-L available in Molcas, the user can choose the contraction of each atom separately. In the same manner, the interface with Gaussian accepts user-defined and non-standard basis sets specified through a separate section of the COBRAMM input file using the Gaussian native format of the keyword *Gen*. Turbomole basis sets are selected via the "define" setup routine, which already provides these features.

2.1.5 Numerical routines and parallel environment

COBRAMM's interface with Molcas includes a couple of numerical features. In particular, gradients and NACs can be obtained for methods that do not possess analytical implementation, as e.g. RASSCF, SS- and MS-CASPT2, RICD-MP2 and RICD-CASSCF (prior to Molcas 8). Furthermore, the atomic displacements required for numerical frequency computations can be handled through COBRAMM Individual displacement steps are stored and can be read in again for a simple restart.

Numerical computations in COBRAMM are performed through finite differences implemented in first (3N displacements) or second (6N displacements) order. To ensure total energy conservation, built-in routines monitor the quality of numerical gradients during MD simulations. Gradients are considered acceptable when the ratio of the lengths of consecutive gradients does not overshoot a (user-defined) threshold. In the case of a detected failure the time step is automatically reduced or enlarged for one step in order to avoid the problematic region. Failures can occur for various reasons: failed wavefunction convergence, divergence of the CP-MCSCF equations, wavefunction instabilities due to strong mixing in state-average MCSCF

computations. In the vicinity of CoIns the numerics is further complicated by possible state swapping. This is counteracted by monitoring the overlap between the wavefunctions at the reference point and at the displaced geometry.

All numerical computations can be performed also in parallel. COBRAMM uses a simple parallelization scheme which relies on the subprocess module of Python which allows to spawn processes and outsources the handling of the individual computations to external (Bash) scripts. This allows users to adapt COBRAMM's performance to the architecture of a HPC cluster. A default Bash script (freqext) is hard-linked in COBRAMM which allows parallel execution on single shared memory multi-processor nodes. freqext can be replaced by more elaborate scripts which handle distribution among multiple nodes without affecting COBRAMM's functionality.

2.1.6 Spectroscopy analysis

COBRAMM provides several tools that help in creating and analyzing spectral data from QM or combined QM/MM computations. The script-based tools range from static IR, (resonance) Raman (see ref. [22]) and UV/Vis analyses towards time-resolved techniques as e.g. fs-UV/Vis (see section 3.1 for an example). Two-dimensional excited state spectroscopy (2DES) is a nonlinear optical technique which emerged in the last decade. It is based on sequences of ultrashort laser pulses with high temporal and spectral resolution. This method provides fundamental insight into coherent excited state dynamics, interstate population transport, solvent reorganization timescales, and noncovalent interactions (π-stacking) [23]. Experimental 2D electronic spectra contain a wealth of (congested) information associated with time-dependent and state-specific signals or fingerprints, such as stimulated emission and excited state absorption, calling for theoretical interpretations that face several computational challenges. COBRAMM connects to the recently developed computational protocol SPECTRON of

Mukamel and co-workers for simulation of linear and non-linear spectra of complex systems like peptides or poly-nucleotides [24–27]. It is based on the sum-over-states (SOS) approach [28], where an *ab initio* characterization of the excited state manifolds of multichromophoric systems within a quantum mechanics/molecular mechanics (QM/MM) scheme is coupled with nonlinear response theory [23], an in-depth description of the method can be found in ref. [29]. An example is given in section 3.2.

2.2 Molecular dynamics implementation

2.2.1 General overview

For molecular trajectories, the classical nuclear motions are propagated in time using a velocity Verlet scheme [19]. In photo-excited systems high frequency motions may occur, as the system acquires kinetic energy in the exited state and returns, vibrationally hot, to the ground state. These oscillations require small time steps to properly represent their molecular motions. Often these motions do not significantly influence the reaction under study but rather act on the stability and energy conservation of the calculation [30]. To freeze out such motions, an iterative Rattle scheme [20] was implemented. Such a scheme is urgently required, when rigid solvent models as e.g. the TIP3P model are used within the mobile MM layer (Section 2.2). The MM parameters of this model are modified to obtain reliable structures in geometry optimizations, and propagation of the corresponding forces would lead to unphysically high frequencies for the O-H bond stretch and H-O-H bending vibrations (~4000-5000 cm-1). The Rattle algorithm therefore fixes O-H and H-H distances, and the water molecule moves as one rigid entity. Details on the implementation of velocity Verlet and Rattle schemes can be found in the SI.

2.2.2 Tully Surface hopping

The transition probability from the current state towards the neighbouring state is computed with the method introduced by Tully [31, 32]. His approach elegantly connects the individual

probabilities for surface hopping with the electronic properties of all molecules in the ensemble. According to Tully's formulation, the squared coefficients for state *i* at time *t* will always reflect the proportion of trajectories that reside in this state:

$$N_i(t) = c_i^*(t)c_i(t)N \tag{2.1}$$

This relation ensures that the average electronic population of the ensemble is always equal to the ratio of trajectories in ground and excited states. Inserting this additional information into the total wavefunction equation over states (see SI) and into the time-dependent Schrödinger equations leads to Tully's surface hopping probability for a hop from state *i* to *j*:

$$P_{ij} = 2\frac{\int_t^{t+\Delta t} Re\left(c_i^*(t)c_j(t)\dot{R}\cdot d_{ij}\right)dt}{c_i^*(t)c_i(t)} \tag{2.2}$$

For integration of equation (2.2) between two consecutive nuclear MD steps, a unitary propagator is applied. The computed non-adiabatic coupling vectors are linearly interpolated between *t* and *t+Δt*, with a typical step size of *Δt/100*. The sign of the coupling vector depends on the sign of the molecular orbitals, which may be arbitrary in consecutive MD steps. It is therefore monitored and adapted accordingly to avoid artificially large numerical changes in between two time steps.

The computed probability P_{ij} for each possible hop from the current state will be compared with a uniform random number. If the random number falls below the corresponding hopping probability, a hop to that state will be invoked, with the molecule propagating from now on the new potential energy surface. Otherwise, the molecule stays in its current state. In COBRAMM the Tully surface hopping is available both in the pure QM and hybrid QM/MM regimes for an arbitrary number of states. These schemes include also the possibility to switch from a state-averaged (SA) calculation to single state (SS) mode, once the trajectory has left the coupling region and the molecule is in the ground state. [33–35] This strategy will lead to a formal drop in

total energy as the wavefunction description changes, but it can avoid the typical SA convergence problems and it allows the user to follow the formation of the ground state photoproduct with significantly less computational demand.

2.2.3 Coherence effects

An ensemble of independent trajectories cannot fully account for the quantum mechanical properties of a wavepacket. At a conical intersection, a real wavepacket will split, i.e. one part of the wavepacket will reside on the upper surface, while another part moves to the lower surface [36]. The time correlation between the two separate packets is quickly lost, and they start to progress independently on the upper and lower surfaces. A single trajectory will just hop to the lower state, with no part remaining on the upper surface, and no relation to other trajectories in the ensemble. This will lead to a mismatch of populations in equation (2.2), i.e. the average electronic population and the real trajectory state occupation will differ, as the strict hop in each individual trajectory introduces an artificial time/state coherence.

Truhlar and co-workers suggested a scheme for an empirical correction of artificial coherence effects [37]. Such a scheme was later realised by Persico et al. [38] It has been implemented together with the Tully surface hopping routine into the COBRAMM set of programmes. (see SI for further details)

2.2.4 Kinetic energy redistribution

The Tully surface hopping scheme yields stochastic hopping probabilities, but often hops occur at a significant energy difference between the involved states. Excited state trajectory simulations are typically computed within a microcanonical (NVE) ensemble, i.e. the sum of kinetic and potential energies (the total energy) must stay constant. To avoid loss of energy at a hop, the resulting energy difference is redistributed to guarantee energy conservation. Typically, the energy is scaled along a certain coordinate to adjust the kinetic energy of the system. The

coordinate for redistribution is not obvious and therefore a matter of discussion. It has been argued that the non-adiabatic coupling vector should be used for this procedure, as this coordinate provides the most efficient motion to lift the degeneracy at the conical intersection [32, 39]. If this vector is not available, the second branching space coordinate defined through the state gradient difference vector may be used, or one may simply scale the energy difference along the velocity vector [40]. Often, the coordinates defined by the coupling and gradient difference vectors have a decisive character in organic photoreactions: in retinal, the coupling vector at the conical intersection contains hydrogen out-of-plane motions, which are deterministic for photoproduct generation [41–43]. As mentioned before, the sign of the coupling vector is arbitrary, thus a random choice of its direction will ultimately influence the computed product distribution when re-scaling becomes necessary. It is therefore sensible to compare the direction of coupling and velocity vectors to avoid scaling into a direction which would artificially disturb the motion of the molecule, possibly pushing it into a different reaction channel. Comparison can easily be achieved by computing the overlap or the angle between the two vectors.

There are two cases, where the kinetic energy contribution provided by the non-adiabatic or gradient difference vector is not sufficient to account for the state energy difference. A trivial situation includes a predicted hop to a state that lies above the total energy level. Such a hop is classically forbidden and is therefore rejected. In another, more rare case, velocity and scaling vectors are orthogonal. In this particular situation, scaling along positive or negative direction cannot contribute to an increase in kinetic energy, and the hop is then either rejected, or the scaling needs to be performed solely along the velocity vector.

All three models for correcting the kinetic energy are available in COBRAMM for non-adiabatic molecular dynamics simulations.

2.2.5 Evaluation of coupling terms

COBRAMM performs mixed quantum-classical dynamics simulations in the adiabatic wavefunction representation, the native representation of quantum-chemical softwares. Accordingly, propagation of the TDSE (eq. 2.3 in the SI) requires knowledge of the electronic energies and of the non-adiabatic coupling (NAC) vectors. The latter quantities drive the population transfer between electronic states in regions of near-degeneracy (avoided crossings), as well as at points of degeneracy (conical intersections). Hence, the computation of NACs is a central aspect in non-adiabatic mixed-quantum classical dynamics. It is therefore not surprising that in recent years much effort has been invested into deriving analytical formulations for obtaining the required quantities for various quantum-chemical methods like CASSCF [44], MS-CASPT2 [45, 46], MRCI [47], TDDFT [48, 49] and CC [50, 51]. Among the QM software interfaced with COBRAMM, Molpro and Molcas are capable of computing analytical state-average CASSCF NAC vectors, MNDO provides them at the OMx-MRCI level, while numerical NACs can be computed at the CASPT2 level (both single state and multi state) through the interface with Molcas.[§]

In practice, the computation of non-adiabatic coupling terms can be very demanding, when a larger number of excited states are involved, e.g. in a CASSCF state-average scheme. For each non-adiabatic coupling vector, a separate set of CP-MCSCF equations needs to be solved. For a system involving five states, the Hamiltonian matrix which needs to be propagated within the Tully surface hopping scheme takes the form:

[§] Note, that as Molcas does not create SS-CASPT2 wave functions, the overlaps are obtained using the SA-CASSCF wave functions and subsequently re-scaled with the factor $(E_{i,SA\text{-}CASSCF} - E_{j,SA\text{-}CASSCF})/(E_{i,SS\text{-}PT2} - E_{j,SS\text{-}PT2})$ which follows from the definition of the NAC. In the case of MS-CASPT2 the overlaps are computed using the wave functions resulting from the linear combination of the original CASSCF wavefunctions utilizing the eigenvectors obtained by diagonalizing the MS-CASPT2 Hamiltonian. See the SI for further details.

$$\begin{pmatrix} 0 & i\dot{R}d_{12} & i\dot{R}d_{13} & i\dot{R}d_{14} & i\dot{R}d_{15} \\ -i\dot{R}d_{12} & \Delta E_{21} & i\dot{R}d_{23} & i\dot{R}d_{24} & i\dot{R}d_{25} \\ -i\dot{R}d_{13} & -i\dot{R}d_{23} & \Delta E_{31} & i\dot{R}d_{34} & i\dot{R}d_{35} \\ -i\dot{R}d_{14} & -i\dot{R}d_{24} & -i\dot{R}d_{34} & \Delta E_{41} & i\dot{R}d_{45} \\ -i\dot{R}d_{15} & -i\dot{R}d_{25} & -i\dot{R}d_{35} & -i\dot{R}d_{45} & \Delta E_{51} \end{pmatrix} \quad (2.3)$$

The diagonal elements denote the state energy difference to the ground state, and the off-diagonal elements include the scalar products of the time derivative velocity vector *dR/dt* with the corresponding non-adiabatic coupling vector between the two states involved, multiplied by the complex unit *i*. The matrix is antisymmetric, i.e. to propagate the full Hamiltonian within two nuclear dynamic timesteps *t* and *t+Δt*, the computation of ten non-adiabatic coupling vectors would be necessary, each one costing ca. one single nuclear gradient computation. As the computation of excited-state trajectories typically involves many hundreds of single point computations, this would be unaffordable. The resulting amplitudes for those states which are energetically well separated from the active state are, however, expected to be small. They are further damped within the decoherence correction scheme, so that it makes sense to concentrate only on those states, that are significantly close to the current potential, and hence compute only the coupling elements for the close interacting states. For the state selection, an energy criterion specified by the user can be employed. Fig. 1 shows the application of such a scheme to the CASSCF computed photoreaction of pyrrole. The system involves four states, starting on the bright S_3 potential energy surface. With a selector threshold of 30 kcal/mol, the computation of non-adiabatic coupling matrix elements starts after ca 10 fs. Before this period, the state amplitudes are set to one for the active state and zero for all other states. With the applied selection threshold for the energy difference it becomes apparent that no more than three interstate coupling elements have to be computed. The impact on the pyrrole dynamics due to the missing amplitude contributions of the other states is therefore expected to be negligible. The selector threshold must, however, be chosen individually and tested for each system. The presented scheme has been implemented into the COBRAMM set of routines together with a multi-state surface hopping routine.

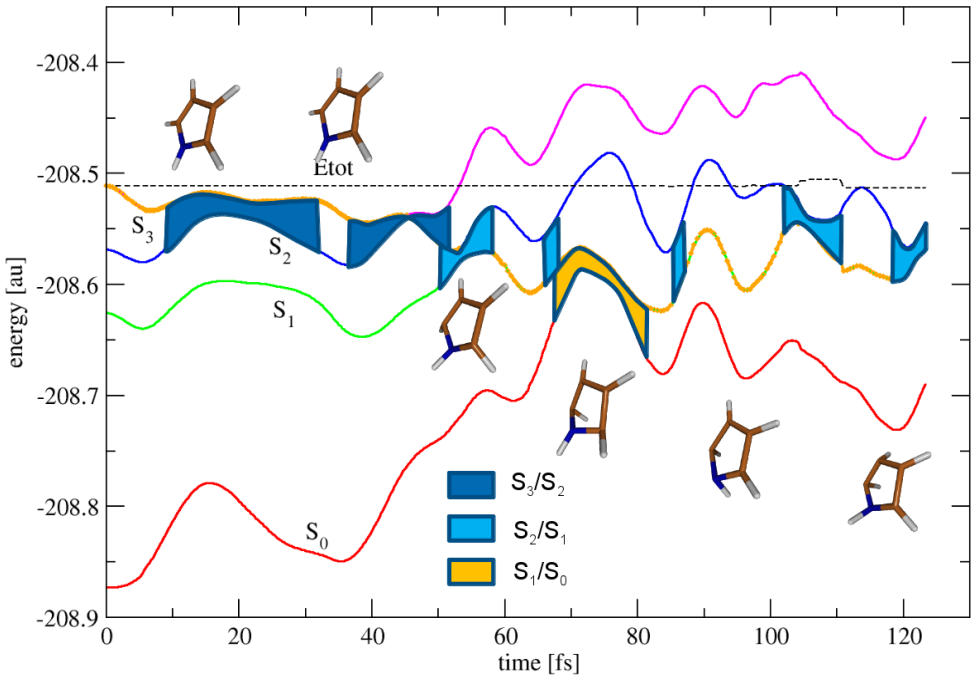


Figure 1. Pyrrole CASSCF photodynamics using the 6-31G* basis set, averaging over four states. The thick orange line denotes the active state, i.e. the PES used to compute the forces. The shaded areas mark the computation of coupling elements.

The description of the population transfer mediated by conical intersections introduces numerical errors when performed in the adiabatic representation, the reason being that real state crossings introduce spikes in the non-adiabatic coupling localized in coordinate space. When the TDSE is numerically integrated with a finite time step and the non-adiabatic coupling exhibits a spike on a time scale shorter than the integration time step the simulation may step over the spike thus missing the coupling region completely. The obvious solution to this issue, namely to use sufficiently short time steps (normally in the range of 0.25 fs), would significantly increase the

computational cost. Using the same rationale as above the energy selector workaround can be employed in COBRAMM to introduce a shorter time step in regions where two states are energetically close.

An alternative to computing the non-adiabatic couplings (NAC) d_{ij} is to compute the time-derivative couplings (TDC) $\sigma_{ij} = \frac{dR}{dt} d_{ij} = \langle \Psi_i(r,t) | \frac{d}{dt} \Psi_j(r,t) \rangle$ employing information about the change of the electronic wave function at the each time step. In practice, the change in the electronic wave functions is resolved by computing overlap integrals between the adiabatic wave functions at different times. This formulation elegantly solves the problem of over-stepping a CI, thus allowing to work with longer time steps even in regions of near-degeneracy. Originally introduced by Hammes-Schiffer and Tully (HST) as an approximation to the analytical scheme, the overlap-based approach has been shown to produce similar probabilities [47].

The TDC scheme is implemented in COBRAMM generalizing the formulation of Barbatti (eq. 9 in ref. [47] to arbitrary time steps Δt:

$$\sigma_{ij} = \tag{2.4}$$

$$\frac{1}{2\Delta t_{old}} \left(1 - \frac{\frac{1}{2}\Delta t_{old} + \Delta t}{\frac{1}{2}\Delta t_{old} + \frac{1}{2}\Delta t} \right) \left(\langle \Psi_i(t - \Delta t) | \Psi_j(t - \Delta t - \Delta t_{old}) \rangle - \langle \Psi_j(t - \Delta t) | \Psi_i(t - \Delta t - \Delta t_{old}) \rangle \right) +$$

$$\frac{1}{2\Delta t} \left(\frac{\frac{1}{2}\Delta t_{old} + \Delta t}{\frac{1}{2}\Delta t_{old} + \frac{1}{2}\Delta t} \right) \left(\langle \Psi_i(t) | \Psi_j(t - \Delta t) \rangle - \langle \Psi_j(t) | \Psi_i(t - \Delta t) \rangle \right)$$

where Δt and Δt_{old} are the time steps used to propagate the TDSE between step *i-1* and *i*, as well as between step *i-2* and *i-1*. Note that eq. 2.4 simplifies to eq. 9 in ref. [47] when $\Delta t_{old} = \Delta t$. In this way, the TDC scheme can be combined with the adaptive time step approach introduced

earlier. In COBRAMM the TDC scheme is available for the single state and multi state flavors of the CASPT2 approach through the interface with Molcas.

2.3 QM/MM Implementation

In many systems, the environment plays a crucial role in ground and excited state chemical reactivity, as it modifies the potential energy landscape of the reactive component and may thus enable different reaction routes compared to vacuum [52]. Involvement of a given protein surrounding or solvent is thus inevitable to understand ground state chemistry, or, for photochemical reactions, the initial excitation process and subsequent energy transfer reactions. A combined quantum mechanical and molecular mechanical (QM/MM) treatment [30, 53–56] applies more accurate quantum mechanical calculations for the reactive component, while the typically much larger environment, which is not directly involved in the initial reactive process, is described by a more approximate strategy based on classical mechanics. A general scheme to treat ground state reactions within a QM/MM scheme has already been presented in the early 70's in the pioneering works of Karplus, and Warshel [57]. Not much later, this model was extended to simulate also excited state dynamics with surface hopping [58]. This approach enabled the accurate prediction of reaction timings and quantum yields for the rhodopsin photosystem with findings that, in general, are still valid at present.

The success of the QM/MM methodology led to a quick development of efficient codes. One may categorize these implementations by the underlying software, i.e. into QM codes with additional MM features, MM programs connecting to specific QM codes or modular software not bound to a specific QM or MM code [53]. The first two variants are extensions to existing QM or MM codes. These are often specific to the underlying software and may not use the full capabilities of the interfaced QM or MM program. Amber e.g. connects to the semi-empirical MOPAC package [59]. The GROMACS MM code [60] has an interface to the Gaussian series of programs; the ONIOM [61] approach in Gaussian allows QM/QM and QM/MM computations within in a subtractive scheme, using all QM methods provided by Gaussian in combination with

the Amber force field. MOLCAS offers a connection to the TINKER MM software [62]. A modular implementation on the other hand does not rely on one specific software package. The ChemShell suite of programs [63] e.g. provides interfaces to the QM software Turbomole [2], Gaussian [1], MOLPRO [8], ORCA [64] and MNDO [12], connecting to the MM force fields used in CHARMM [65], Amber [17], GROMOS [66] and GULP [67] through the DL_POLY [68] interface, offering flexible combinations among the programs with a common, TCL based input language for all modules. Also the Newton-X suite by Barbatti et al. [69] offers QM/MM functionalities using diverse QM programs with the MM program Tinker.

All of the described programs are capable of modelling ground state reactions within a QM/MM environment. A first hurdle for the user is, however, the generation of a proper molecular mechanics setup, which must be performed with the selected MM software. The preparation and analysis work gets more involved when excited state QM/MM dynamics are required to compute e.g. spectral properties or time-resolved spectral traces, although this is generally possible with the described packages. Our QM/MM implementation in the COBRAMM software [5, 70] is a modular approach specifically designed for usage of various excited state QM programs with the Amber force field. In the following, we will describe the general QM/MM implementation and the specific features added to perform and analyze ground and excited state QM/MM computations.

2.3.1 The QM/MM energy expression

COBRAMM uses a subtractive QM/MM scheme, where the total energy is calculated as

$$E_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I + link) - E_{MM}(I + link) \tag{2.5}$$

S is the whole system under study, I forms the inner region plus link atoms (link, see also Fig. 2). To obtain the total QM/MM energy, three separate computation steps are required:

1.) Calculation of the entire systems energy at the MM level of theory

2.) Computation of the inner system and link atom energies using a QM method

3.) Computation of the inner system and link atom energies with an MM method, and finally subtracting this term from the sum of the latter

Different from additive schemes [53], the suggested formalism requires molecular mechanics parameters also for the QM part of the system. These parameters are easily obtained within our setup procedure (see section 2.11). Details on the implemented link atom approach are found in section 2.6.

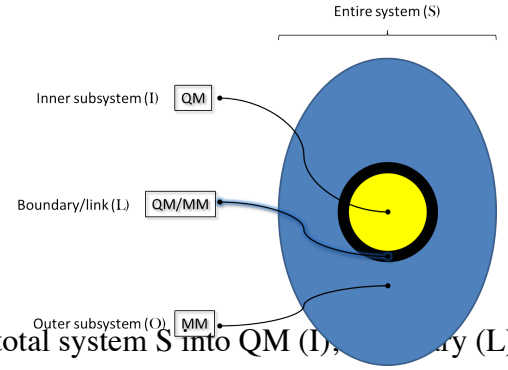


Figure 2. Division of the total system S into QM (I), [illegible]ry (L) and outer (O) regions.

2.3.2 Layer partitioning

The total system is furthermore partitioned into three interacting regions called layers hereafter

- The high layer (H), including those atoms that are treated with a quantum mechanics method. This layer is identical to (I+link) in the energy expression

- The medium layer (M) – this part consists of MM atoms which are movable during optimizations or dynamics runs

- The low layer (L) – the remaining part of the system, also treated at the MM level.

The first two layers build the mobile portion of interest of the system under study. In addition, the low layer may be frozen or optimized/equilibrated *separately* from the two other parts within a microiterative scheme (Fig. 3). This partitioning enables treatment of the core region (H+M) with an advanced optimization scheme like e.g. BFGS [71]. Thereafter, the low layer (L) can be

optimized with a much faster and more "rough" approach like e.g. steepest descent or Newton-Raphson, and the procedure continues until the convergence criteria are fulfilled for both, H+M and L regions.

Figure 3. H, M and L layers in COBRAMM

2.3.3 Embedding of the high layer

Current QM/MM implementations define two main approaches to treat the electrostatic interactions between QM and MM regions: mechanical embedding and electrostatic embedding [53–56, 72, 73].

In a mechanical embedding scheme, the electrostatic effects between inner and outer layers are solely treated at the molecular mechanics level. This means that the quantum system does not directly interact with the charge distribution of its surrounding, but only indirect through MM electrostatics. While mechanical embedding might be a valuable approach in ground state conformational reactions, where the charge fluctuations of the QM system are minor, it is a drastic simplification for situations where bonds are formed, broken or the electronic structure suddenly changes through excitation. In these cases, the missing interaction with the inner system wave function cannot be fully recovered by e.g. simply updating the charges of the QM system in every step of the si

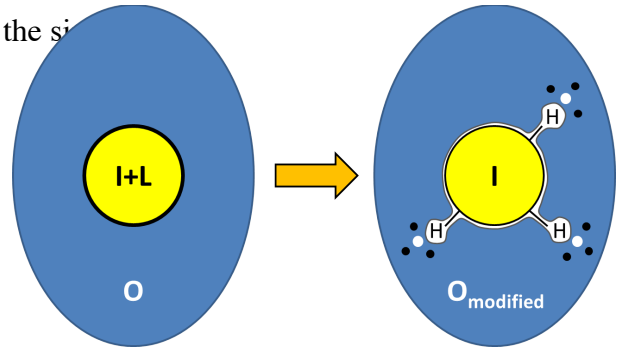


Figure 4. Embedding of the high layer with hydrogen link-atoms. The charge field of the outer region is adapted near the links to avoid over-polarisation of the QM region. Therefore, the charge of the direct link-atom neighbor (M^1) is set to zero (white circles) and the charges are repartitioned among the M^2 atoms (black dots).

In electrostatic embedding schemes (Fig. 4), the point charges of the MM subsystem enter as additional one-electron terms into the QM Hamiltonian and directly influence the electron distribution in the QM system. Both approaches can be selected in the current version of COBRAMM, although for excited state computations the latter scheme should be preferred.

The electrostatic interaction between QM and MM parts in the simulation may lead to overpolarization effects, which especially apply for the boundary region. Such effects can be reduced by a charge repartitioning procedure. Such an approach is discussed, together with the treatment of boundary atoms, in the next section.

2.3.4 The QM/MM frontier

In many biochemical systems, the reactive molecule of interest is covalently attached to its environment. This means, that the bond between the Q^1 and M^1 atoms, which is not considered in the QM system (Fig. 5), would leave a free valence for this part. In the link-atom approach used in COBRAMM, the QM dangling bond between Q^1 and M^1 atoms is saturated with a hydrogen atom. The link-atoms do not appear in the MM description of the system, thus no van der Waals interactions apply for this part. Electrostatic interactions however must be taken care of, since the charge of the close-by M^1 atom may overpolarize the QM system, leading to strong artefacts in the simulation. To avoid this, the charge of the M^1 atom is set to zero (Fig. 4 and 5) [74], and its charge is redistributed among its neighbouring M^2 atoms, weighted by the original charge contribution of the M^2 atoms [70, 75]. The motions of the link-atoms are restricted along the M^1-Q^1 axis. Therefore, the position of a link-atom is expressed as a function of the Q^1 and M^1

positions, and their nuclear gradient components are re-distributed among the Q^I and M^I atoms. At this point, COBRAMM adopts a similar strategy as documented for ONIOM [61]. The placement of a link-atom introduces three additional degrees of freedom, which must be eliminated to avoid artefacts in optimization and MD simulations. Details on the implementation of the discussed link-atom scheme can be found in the SI.

Figure 5. QM and MM regions with link-atom and numbering.

2.3.5. MM interface functionalities

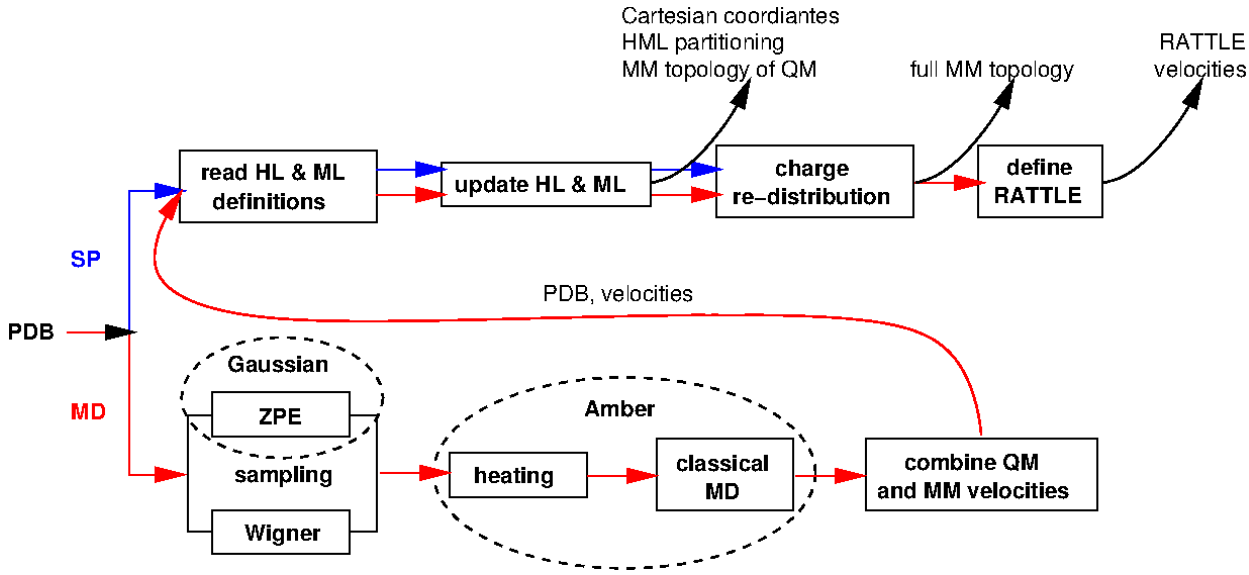
COBRAMM can be used with all Amber distributions starting from version 10, handling also GPU, SMP and MPI parallel options. Pmemd and CUDA versions of the sander program efficiently speed up QM/MM computations in implicit solvation environments. MM gradients are printed using the "dump force" routine of Amber, thus without modification and recompilation of the MM software. In very large MM systems (>20.000 atoms) this strategy can form a significant bottleneck. The process can be accelerated by computing one MD time step at the MM level. The forces can then be retrieved from the velocities of the mobile part. This option can be selected within the new version of COBRAMM.

2.3.6 QM/MM setup and initial conditions for MD simulations

The new version of COBRAMM includes a routine that significantly reduces the user efforts in the generation of the initial QM/MM setup. This script-driven program allows the user to select the QM region of interest and can define in a semi-automatic fashion the mobile region surrounding the QM part through a distance- or H-bonded based criterion. It also takes care for automatic setup of the Rattle scheme, when rigid solvent molecules are used in the M layer. Missing residues in the MM definition can either be added automatically or they may be loaded

from a previously defined library. The distance criterions to define the mobile (M) layer include a) the center of mass of the H or H+M layers or b) the distance to the atoms of the H or H+M layers. H-bonds to H and H+M layers are computed and residues forming an H-bond with atoms of the specific layers can be added automatically to the corresponding layer definitions.

Several options were implemented to generate initial conditions for performing MD simulations. Starting geometries and velocities are either generated through a classical Amber MD simulation taking snapshots from a previous MM dynamics run, or they are created using quantum sampling techniques. Both techniques may also be combined. Quantum sampling requires a prior frequency computation. From the resulting normal modes the user can create a) a zero-point energy (ZPE) sampling [76, 77], b) thermal sampling including also a Boltzmann-temperature distribution [76]. For the latter, we provide an interface to the Gaussian ZPE and thermal sampling options. Furthermore, we have interfaced the Wigner sampling technique applied in the quantum molecular dynamics program JADE [78], which also allows to include temperature effects and the excitation of specific normal modes. For all quantum sampling methods, high frequency modes, which usually do not contribute to the reactions of interest and rather influence the stability of the calculations, may be excluded from sampling. In a combined QM and MM sampling scheme, the quantum mechanically sampled geometry is subjected to an MM equilibration run, as suggested in a study of Rückenbauer et al. describing solvation effects on the photoisomerization of a retinal model [79]. This strategy allows the mobile MM region to adapt to the QM-sampled geometry and delivers velocities and a new geometry for the MM part, which, combined with the quantum sampling velocities and geometry, forms the starting pointing of a new QM/MM dynamics run. Scheme 2 summarizes the capabilities of the implemented routine.



Scheme 2. Schematic task and data flow in the script-assisted preparation of the input files required for QM/MM computations. The setup for single point calculations and optimizations is shown in blue, the route for initial conditions in molecular dynamics simulations is marked with red color.

3 Applications

The QM interfaces implemented in COBRAMM are mainly controlled through their own program specific standard keywords, which are provided as a separate section to the COBRAMM input file. We decided against a common input language for all QM programs, as the functionalities and the availability of specific options in the QM programs largely differ. This way, the user can supply the input in a manner he is already used to and does not need to learn an additional language for this part of the setup. For best convenience, COBRAMM offers the capability of automatic routine-specific QM input completion. In most cases, only the definition of the computational level would fall within the user's responsibilities. Thus, various flavors of geometry optimization (minima, TS, CI), minimum energy path computations (IRC), frequencies and even dynamics simulations can be performed with the same QM input. In this way

COBRAMM lowers significantly the entry barrier for non-experts to a given quantum chemistry software. At the same time, COBRAMM provides experienced users with the freedom to manipulate the QM input according to their needs, hence the user's input takes precedence over the automatic completion routines. The MM input generation is fully automatized and, as a general rule, does not require the user's interference. Nevertheless, in the spirit of the COBRAMM's philosophy also the Amber input can be manipulated in a designated section to the COBRAMM input file. The output of each (QM and MM) program is, in any case, stored, so that additional data not covered by the COBRAMM analysis facilities can be extracted a posteriori. Table 3 gives a quick overview on the structure of the main input file *cobram.command* for two different tasks: a) using Gaussian and the Amber interfaces for QM/MM geometry optimization and b) using Molcas and Amber for excited state surface hopping dynamics. The implementation of the COBRAMM QM/MM routines itself was validated by performing a comparative DFT/Amberff99 study on the rhodopsin photosystem with the Gaussian09 ONIOM program. Geometry parameters and excitation energies of retinal in rhodopsin are reported in the SI. In the following, we discuss some features of newly implemented and updated QM and MM interfaces by means of specific examples.

Table 3. Input sections in *cobram.command* file. a) QM/MM geometry optimization with Gaussian and Amber, b) QM/MM excited state dynamics with surface hopping using Molcas and Amber. The geometry of the system and the layer partitioning are specified in separate files.

a)

```
!keyword
type=optxg qm-type=gauss mm-type=amber
qmem=1000mb nproc=8
?keyword


!sander real
  input for whole system S
  &cntrl
  imin   = 1,
  maxcyc = 1,
  ntb    = 0,
  igb    = 0,
  ntr    = 0,
  ibelly = 1,
  cut    = 12,
  ntxo   = 1
  /
?sander real


!gaussian
#P CAM-B3LYP/TZVP Nosym

0 1
?gaussian
```

b)

```
!keyword
type=mdv qm-type=molcas mm-type=amber
qmem=1000mb surhop=persico tstep=0.25
?keyword


!sander real
  input for whole system S
  &cntrl
  imin   = 1,
  maxcyc = 1,
  ntb    = 0,
  igb    = 0,
  ntr    = 0,
  ibelly = 1,
  cut    = 12,
  ntxo   = 1
  /
?sander real


!molcas
  &RASSCF
  Symmetry = 1
  Spin = 1
  nActEl =  10 0 0
  CIroot =  2 2 1
  Inactive = 40
  RAS2 = 10
  RLXR = 2
  LumOrb
?molcas
```

3.1 Rhodopsin QM/MM photodynamics with CASSCF and OM2-MRCI

The photoreceptor protein rhodopsin triggers vision in mammalians through light induced 11-*cis* to all-*trans* isomerization of the retinal chromophore. Due to the size of the system, the simulation of retinal photodynamics within the protein environment forms a challenging task. By now, simulations were performed with full and reduced active-space CASSCF methodology, describing ensembles with up to 60 individual trajectories with retinal in the QM part of the calculation. In a combined experimental and computational study, we were able to trace the time evolution of spectral signatures after rhodopsin photoexcitation and thus could follow the molecular motions of the retinal chromophore within opsin [80].

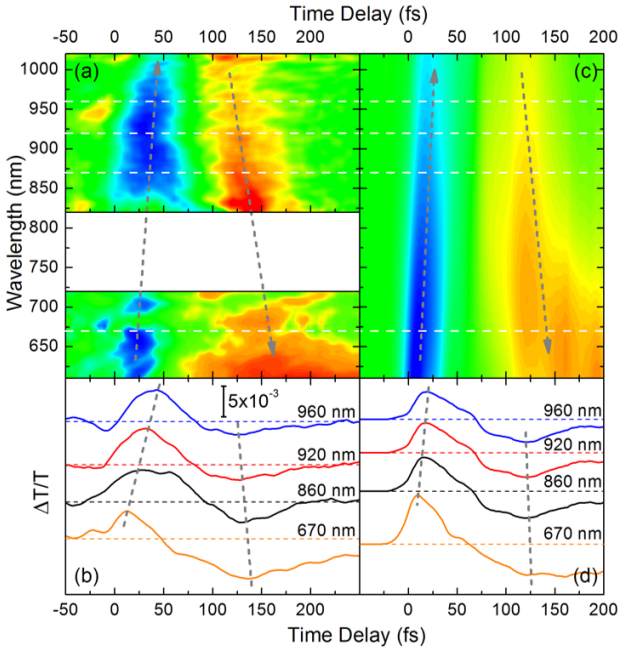


Figure 6. Experimer odopsin. The blue signals in time resolved spec yellow and orange traces denote chromophore photoabsorption. The lower graphs (b) and (d) provide horizontal cuts at selected wavelengths. Figure adapted from ref. [80] with permission from Nature.

Figure 6 shows the results of 38 CASSCF/Amber trajectories in comparison with the experiment. Surface hopping MD computations were performed with the COBRAMM Molpro and Amber

interfaces (see ref. [80] for details on the computational setup). The resulting energy differences and oscillator strengths of the individual trajectories were scaled and processed via Gaussian convolution to reproduce the experimental 15 fs time resolution in a pump probe experiment. In such an experiment, the pump pulse excites the molecules, and a probe pulse induced after a certain delay stimulates the molecules to emit (blue signal in Fig. 6). The emission wavelength then gives information about the distance of the potential energy surfaces. When the molecule is in the ground state, the probe pulse is absorbed (orange/red signatures in Fig. 6). Along the time axis in Fig. 6, the blue emission signal shifts to higher wavelengths, as the chromophore undergoes rotation about the reactive bond and the S_1/S_0 state energy gets smaller. After reaching a conical intersection, the molecule returns to the ground state and the state energy difference drops, this can be seen by the occurrence of the orange signal in the right part of the spectra.

Semiempirical methods can significantly reduce the computational effort, provided a suitable method can be found which is able to reproduce the results of the *ab initio* calculations. The OMx methods in the MNDO program by Thiel et al. [12, 13] are, in combination with efficient GUGA-CI [81] routines, specially parameterized for the computation of excited state properties such as excited state energies, gradients and non-adiabatic couplings used for surface hopping dynamics simulations or geometry optimizations [14]. This has been shown in the latest studies of Thiel and co-workers [15, 82–84]. Orbital overlap procedures make sure that the chosen orbitals remain in the active window during optimization and MD simulation. Within COBRAMM this option is also available for QM/MM simulations. To reduce a possible drift in the total energy of OMx-MRCI dynamics computations, we have furthermore implemented a variant of the adapted time step algorithm suggested by Spörkel et al. [85] into our velocity Verlet scheme.

In the following, we report a benchmark study on the photoisomerization of retinal in the rhodopsin protein using OM2-MRCI methodology.

For computations on rhodopsin, the full retinal system including one CH_2 group of Lys296 was used as the QM part, and the remaining system was described by the Amber ff99 force field. The structure of retinal was re-optimized at the OM2 RHF level with a frozen protein environment. As an initial test, a trajectory starting from the S_1 FC region without initial velocities (at 0 K) was computed. This state also corresponds to the HOMO-LUMO single excitation in OM2/MRCI. The protein environment was kept frozen at the crystal structure coordinates during the whole simulation. As active space for excited-state OM2/MRCI computations, the highest ten occupied and the lowest ten unoccupied orbitals were chosen, corresponding to 20 electrons in 20 orbitals. This selection covers all occupied and unoccupied π-orbitals, but also includes lower lying occupied and higher lying unoccupied σ-orbitals that may become important, when the system starts to twist along the reactive coordinates. Single and double excitations were considered in the MRCI expansion. State energies were computed for the ground and the first two excited singlet states. All related derivative couplings were calculated to obtain state amplitudes and hopping probabilities in the decoherence-corrected Tully surface hopping scheme.

Fig. 7 shows the evolution of several parameters of this MD in comparison to the CASSCF(10,10) and CASSCF(12,12) MD data from earlier CASSCF work reported in ref. [80] We find a reasonable agreement between the two approaches. In both OM2/MRCI and CASSCF computed trajectories, a conical intersection point is reached via the C11=C12 torsion coordinate and the molecule decays to the all-*trans* isomer. The evolution of torsion angles looks very similar; this is also true for the progression and steepness of the S_1 potential. From the FC point to the CI, it drops by ca. 20 kcal/mol in both descriptions. The distance to the S_0 potential is, however, significantly smaller in OM2/MRCI. At the FC point its value reduces from ca. 80 kcal/mol in CASSCF to ca. 60 kcal/mol in the OM2/MRCI description.

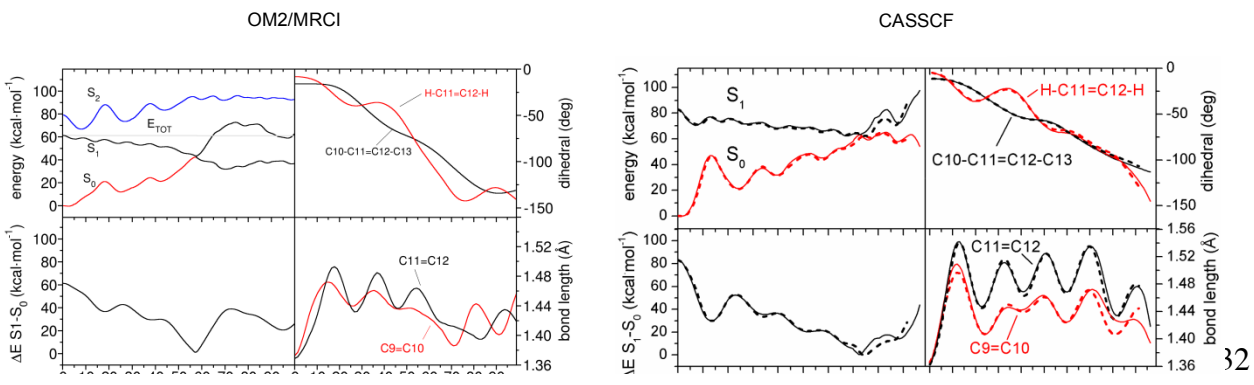


Figure 7. Selected MD calculated parameters for OM2/MRCI (left) and CASSCF (right) trajectories. The left graph additionally reports the evolution of the covalent S_2 state and the total energy. The dotted lines in the right graph denote a CASSCF(12,12) active space, the full lines correspond to the CASSCF(10,10) trajectory.

This corresponds to a vertical excitation energy of 2.66 eV (466 nm), which is much closer to the experimental value of 500 nm than in the CASSCF description. The modified distance to S_0 may also be the reason for the faster decay in OM2/MRCI. With this method, the crossing point is reached already after 57 fs, compared to 73 fs in CASSCF. We also note a smaller amplitude in the C=C stretchings. For OM2/MRCI the first maximum in the C11=C12 coordinate is reached at 1.48 Å and ca. 15 fs, while CASSCF places it at 1.54 Å and 10 fs. A subsequent OM2 numerical frequency analysis delivered the normal modes to create 40 initial conditions using Wigner sampling, constituting a similar ensemble size as in the earlier study. The high frequency modes were not considered in this procedure. OM2/MRCI trajectories starting in the S_1 state were followed for up to 250 fs, using a constant time step of 0.25 fs. Fig. 8 reports the computed UV spectrum and the population decay in the trajectory ensemble. In the UV spectrum from Wigner sampling, a significant red shift of the maximum absorption occurs with respect to the computed vertical excitation. The observed value (492 nm) is even closer to the experiment. In this spectrum, also the S_2 absorption corresponding to HOMO-LUMO double transition becomes apparent, peaking at ca 370 nm (361 nm for the vertical excitation).

The S_1 population sharply drops after the start of the reaction. This drop is connected with a small raise in S_2 population, i.e. according to the OM2/MRCI description this state participates in a small number of ensemble members. The S_0 population begins to rise slowly after ca. 20 fs. After

50 fs it quickly increases until ca. 75 fs, then it grows with smaller slope. After 180 fs, all trajectories are in the ground state.

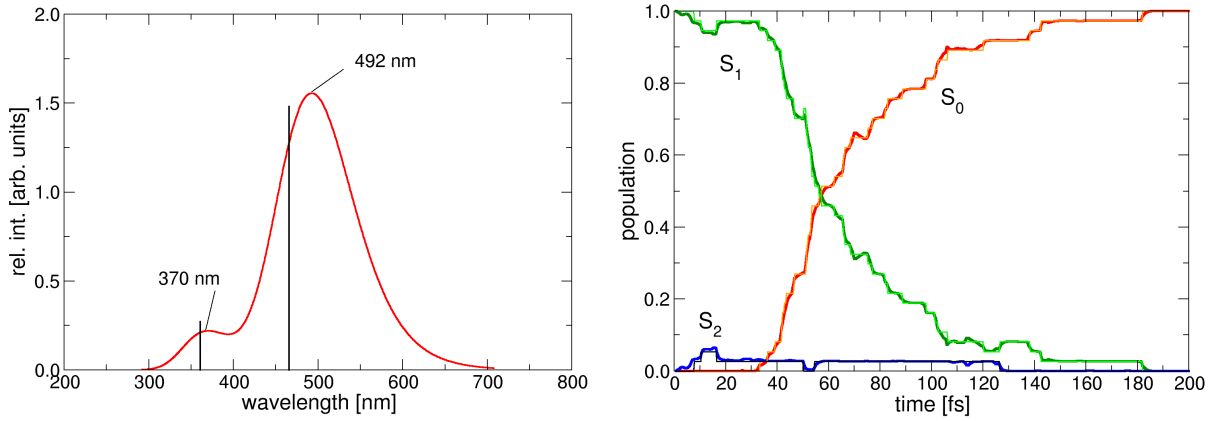


Figure 8. Left: UV spectrum from Wigner sampling including 40 structures. The lines denote the computed vertical excitation energies of the S_1 (466 nm) and S_2 states (361 nm). Right: Populations of S_0, S_1 and S_2 states in the trajectory ensemble.

The average excited-state lifetime was ca. 72 fs, and the computed quantum yield corresponds to 0.65. These values are again in good agreement with the data obtained by earlier CASSCF computations, ranging from 0.58-0.75 for the quantum yield and 80-110 fs for the S_1 lifetime. They also compare well with the experiment. In conclusion, the OM2/MRCI method delivers reasonable results for static and dynamic QM/MM computations of rhodopsin. It may provide further insight into the dynamics of more complex photosystems at significantly reduced computational expenditures.

3.2 Spectral properties of Benzophenone in aqueous solution

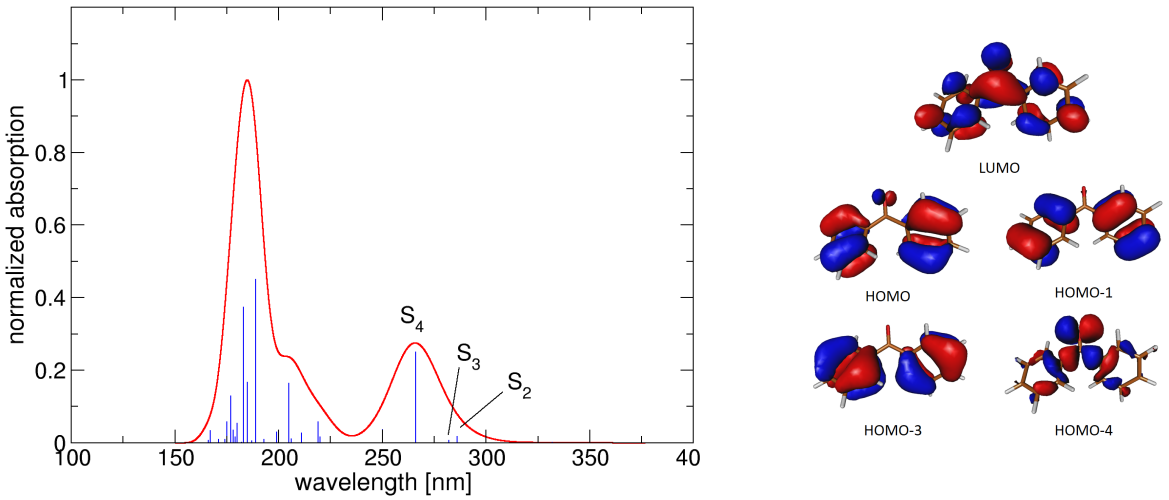
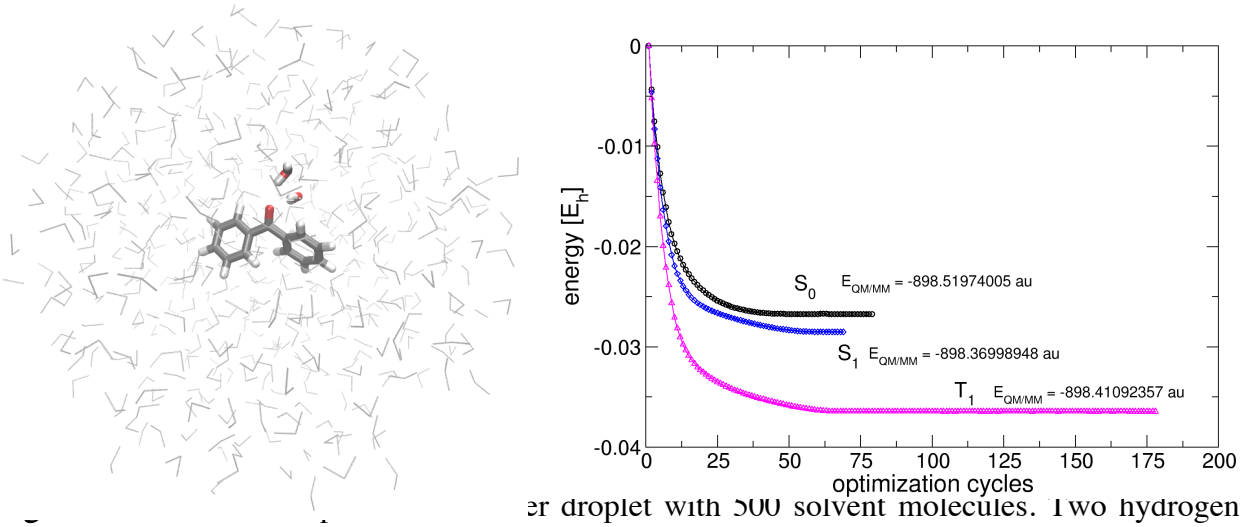


Figure 9. Left: DFT/MRCI computed absorption spectrum of benzophenone in water using the COSMO solvation model [86], convoluted with a Gaussian function of 1500 cm^{-1} (FWHM) and converted to nm. Right: BH-LYP orbitals involved in S_1 and S_4 excitations.

Benzophenone is a well-known triplet sensitizer and was used already for DNA sensitization [87, 88]. Its efficiency in triplet sensitizing is explained by fast internal conversion to S_1, followed by efficient S_1/T_1 intersystem crossing [89]. Fig. 9 shows the DFT/MRCI computed absorption spectrum, obtained from the CAM-B3LYP/TZVP optimized ground state structure in PCM water solution. Its brightest singlet state absorption in the 225-300 nm range (middle UV, MUV) can be attributed to S_0-S_4 excitation, providing a mixture of HOMO-3□LUMO, HOMO-1□LUMO and HOMO□LUMO π-π* excitations. The lowest singlet state (S_1) has n-π* character (HOMO-4□LUMO, see Fig. 9). Experimental evidence for singlet interconversions and S_1/T_1 ISC was provided in the study of Shah *et al.* [89] Optimized geometries for benzophenone in aqueous solution have been obtained using the Turbomole [2] interface. QM/MM computations with this interface require the "define" setup routine to be executed once for the QM part. The chargefield of the MM part is then handled by the internal COBRAMM routines.

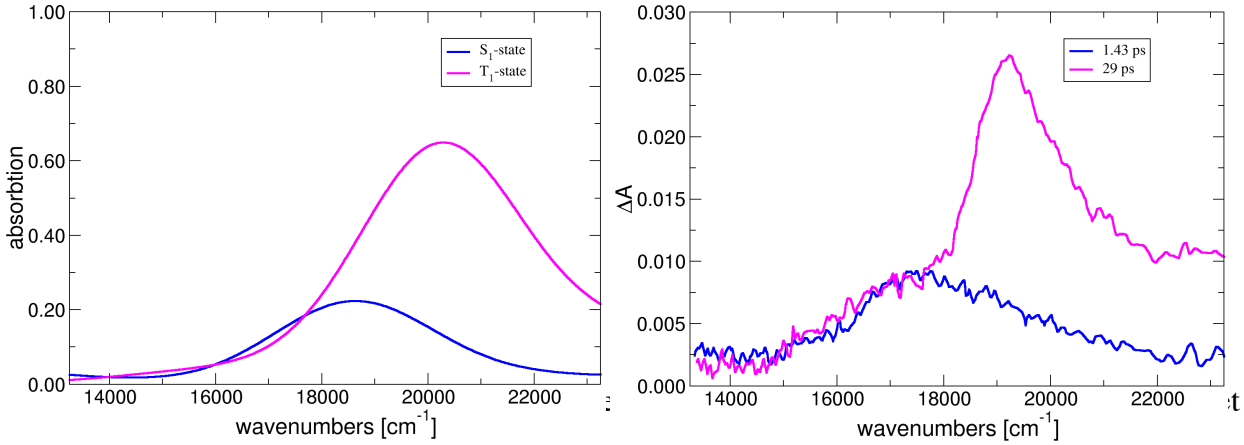


er droplet with 500 solvent molecules. Two hydrogen bonded water molecules (licorice) were included into the high layer and optimized together with

BP. Right: Optimization cycles with Berny-Schlegel optimizer and resulting DFT QM/MM energies.

Fig. 10 depicts S_0, S_1 and T_1 geometry optimizations of benzophenone (BP) in a water droplet consisting of 500 TIP3P water molecules, computed using the BH-LYP functional and the SVP basis set. Two water molecules H-bonded to the oxygen of BP were included in the QM part of the calculation. Geometry data was obtained from a previous Amber ff99 equilibration run with periodic boundary conditions [90]. All optimizations were started from the same, OM2/Amber-pre-optimized ground state geometry. S_1 and S_0 state geometries converge quickly within 70-80 optimization cycles, the T_1 geometry taking significantly longer. The DFT/MRCI [3, 4] interface allows subsequent single point calculations to obtain transient absorption data for benzophenone in the relaxed S_1 (n-π*) and T_1 (π-π*) states.

Fig. 11 shows the corresponding values in comparison with experimental data. We observe a slight overall blue-shift, but the computed data clearly reproduce the trend in short and longer time transient absorption. The short time absorption signal stems from S_1, then the system quickly reaches an S_1/T_1 intersystem crossing, efficiently converting the complete S_1 population towards T_1.



for S_1 and T_1 states, convoluted with a Gaussian function of 1500 cm^{-1} (FWHM). Right: experimental transient absorption spectra measured at 1.43 and 29 ps after excitation. Experimental data from the right-hand side panel was taken from ref. [89] and converted to cm^{-1}.

Further insight into benzophenone photoreactivity in water can be provided by means of 2D electronic spectroscopy. Fig 12 displays a simulated two-dimensional excited state (2DES) spectrum, based on one excited S_1 state trajectory computed with the TD-B3LYP method (see ref. [90] for details). The spectrum itself was created by using the RASPT2 method to obtain estimates for vertical excitation energies and associated transition dipole moments on top of the DFT computed geometries out of the trajectory. The active space included all bonding π valence orbitals and the n_o orbital into the RAS1 subspace, while all anti-bonding valence π* orbitals comprised the RAS3 subspace. Five holes/electrons were allowed into the RAS1/RAS3 subspaces, respectively, leading to RASPT2(5,8|0,0|5,7). Atomic Natural Orbital basis of large type (ANO-L) were employed in their triple-ζ (VTZP) contraction. This level of theory was employed as it was found to be the best compromise between accuracy and computational cost in order to represent the electronic excited state manifold of benzophenone [91].

Figure 12 shows 2D spectral traces right after photoexitation (0 fs) and in three time intervals (100, 200, 300 fs) afterwards. Initially (at time = 0 fs) the signal consists of slightly shifted analogous fingerprints arisen from the $S_1(n_o\pi^*)$ state and thoroughly characterised in gas phase [91], depicting dipole allowed $n_o\pi^* \rightarrow n_o\pi^*$ transitions, with its main fingerprint being placed at ~17500 cm^{-1}. A focus is taken in the Vis probing window as it provides fingerprints that can also be registered with linear-response-based method. [90] After the first 100 fs, this main fingerprint is red-shifted towards the 12-14k cm^{-1} range while blue-shifting the other intense transitions registered at the FC region. This strong shift is encompassed by pronounced carbonyl stretchings occurring along the $S_1(n_o\pi^*)$ relaxation path that stabilise (red-shift) the excited state absorption signal while blue-shifting other higher-lying $n_o\pi^*$ contributions. At later times (200 and 300 fs snapshots) it can be seen how the main intense excited state absorption fingerprint placed at 11-13k cm^{-1} oscillates with the stretching of the carbonyl as well as the appearance of a series of relatively dark signals in the high-energy window associated to doubly excited ππ* states, which are partially dipole allowed due to the geometrical deformations along the trajectory that breaks

the pure character of the $S_1(n_o\pi^*)$ state making them accessible. Monitoring these excited state absorption signals thus provides a useful additional way to track excited state dynamics and where their specific broadening, omitted in that study due to computational demands, can provide further information regarding the specific solvation effects and excited state lifetimes.

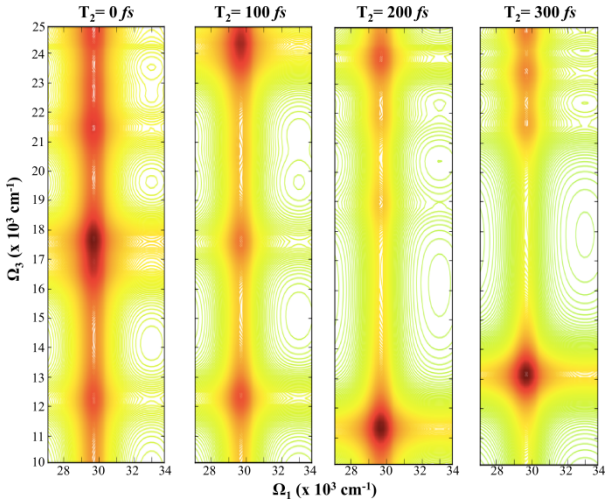


Figure 12. RASPT2 computed 2DES of benzophenone along a representative TD-B3LYP trajectory. Ω_1 refers to the pump frequency whereas Ω_3 is the probe frequency. Red colour denotes positive (excited state absorption) signals whereas blue denotes negative (ground state bleaching and stimulated emission) contributions. Figure reproduced from SI of ref. [90] with permission from the Royal Society of Chemistry.

3.3 Excited state lifetime and quantum yield in animal and human rhodopsins

As a primary photoactive component the chromophore 11-*cis* retinal is also found in human rod, cone and ganglion receptor cells as well as in numerous invertebrates. In this study [92], we have investigated the influence of the different environments on the structural, spectral and dynamic properties of 11-*cis* retinal in the human melanopsin (hMeOp) [93] and squid rhodopsin (sqRh). [94] Although stemming from different species, these two proteins share significant sequence similarities. These analogies allowed the construction of a hMeOp homology model based on the known crystal structure of sqRh [95]. The results of excited state surface hopping molecular dynamics performed with these systems were compared to the photodynamics of bovine rhodopsin. 11-*cis* retinal was geometry-optimized in the static environments of the sqRh and bovine Rh crystal structures as well as in the hMeOp homology model applying CASSCF methodology. The active CAS space comprised 12 electrons and 12 π-orbitals, using the MOLCAS set of routines and the 6-31G* basis set. The QM model of retinal is the same as already described in section 3.1. To describe the fixed protein environment, the Amber FF96 force field was used. The left hand side of Fig. 13 depicts an overlay of the obtained retinal structures. The geometries were aligned along the central C9-C14 fragment. The retinal chromophores in sqRh (yellow) and hMeOp (blue) are significantly bent with respect to bovine Rh (red). In combination with the electrostatic effects provided by the protein environment, these deformations lead to shifts in their UV absorption spectra. MS-CASPT2 excitation energies were computed on top of a room temperature geometry ensemble obtained from thermal sampling using 60 snapshots for each system. The results reproduce the experimental trend: with 481 nm the bright absorption in sqRh is significantly blue-shifted compared to bovine Rh (503 nm). hMeOp (476 nm) is even more blue-shifted.

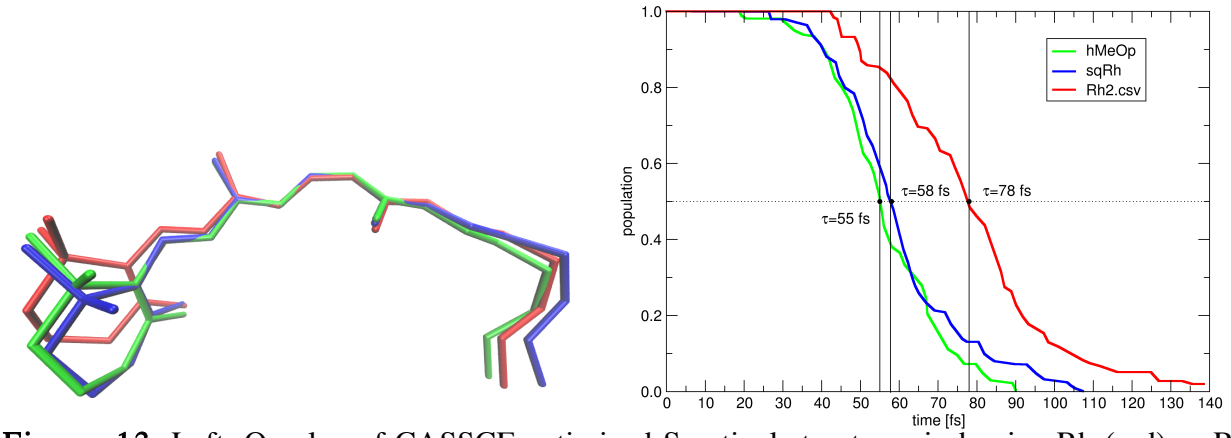


Figure 13. Left: Overlay of CASSCF optimized S_0 retinal structures in bovine Rh (red), sqRh (green) and hMeOp (blue). Right: Decay of S_1 state population in each protein from average electronic population of individual trajectories using the same color coding.

Earlier single trajectory computations [95] revealed an additional difference in photoreactivity: for the blue-shifted opsins a shorter excited state lifetime was observed. According to the one-dimensional Landau-Zener model, this reduction in excited state lifetime should be connected with an increase in quantum yield for the all-*trans* isomer. To test this hypothesis, 60 excited state surface hopping trajectories from the sampled geometries were started using a reduced CASSCF(10,10) active space, averaging equally over the ground and first excited state. These computations were performed by taking advantage of the QM/QM interoperability of COBRAMM. Excited state gradient computations were done with MOLCAS, while for non-adiabatic coupling vectors the resulting wavefunction was converted and the MOLRPO interface was executed.[**] The main results of this study are depicted in the right hand side of Fig 13. The population graph shows the decay of the excited state in all sets. With 78 fs (using the time, where half of the molecules have decayed), Rh has the longest excited state lifetime, sqRh and hMeOp decaying 20-23 fs faster. These findings are in line with earlier single trajectory computations started from the Franck Condon point without initial velocities. Fig. 14 gives more detailed information on the S_1 lifetime and includes statistical data collected in periods of 10 fs. The height of the bar graphs denotes the portion of trajectories leading to the all-*trans* photoproduct (green) and back to the 11-*cis* starting configuration (red) within a particular 10 fs

[**] It should be noted at this point, that the new version of MOLCAS includes analytic derivative couplings for CASSCF wavefunctions.

time frame. As already apparent from the overall all-*trans* retinal quantum yields (QY), which are also given in this figure, the expected increase due to higher reaction speed was not observed. Instead, sqRh and in particular hMeOp feature a reduced QY.

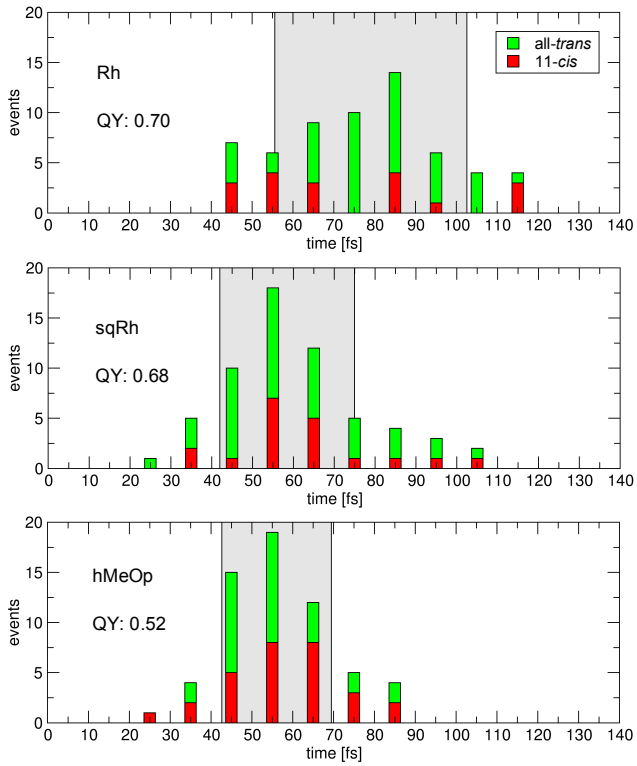


Figure 14. Distribution of excited state lifetimes. The height of the bar graphs denotes the portion of trajectories leading to the all-*trans* photoproduct (green) or to the 11-*cis* reactant (red). The grey area denotes twice the standard deviation of hopping event times.

The lifetime statistics graph also reports the decay time window by means of the standard deviation of the hopping times of all ensemble members (grey areas in Fig. 14 right). These time windows start much earlier and have a smaller width for sqRh and hMeOp than in the corresponding bovine Rh, i.e. the hopping area is much more localized and it is shifted towards shorter times with respect to Rh. This behavior is a consequence of the bend introduced in the ground state structures of sqRh and hMeOp, (Fig. 13) which produces an additional strain in the polyene fragment. This strain is more effectively released and converted to torsional motion along the C11=C12 cis bond in the excited state, thus speeding up the approach towards the

conical intersection. Product formation involves at least two coordinates: the torsion about the C10-C11-C12-C13 carbon fragment and, as the decisive mode, torsion around the corresponding H-C11=C12-H dihedral angle. In other words, the chromophore must be sufficiently twisted around C10-C11-C12-C13 to approach the CI, then the direction of the H-C=C-H motion will define, whether the molecule follows this twist in the ground state to lead to all-*trans*, or whether it will return to the 11-*cis* starting configuration. This rule, which was stated also in earlier studies [41–43], is applicable for the vast majority of our computed trajectories and can explain the shift in quantum yield in hMeOp. In this chromophore, the carbon and hydrogen torsions run out of phase as a consequence of the additional strain in the chromophore.

Fig. 15 right shows averaged values of the corresponding carbon and H-C=C-H torsions denoting the increase in reaction velocity along the C-C-C-C motion for hMeOp and sqRh. A speedup can also be seen for the hydrogen torsion, but the oscillations in this mode are mostly unaffected (see min. and max. lines in Fig. 15). This leads to a mismatch of the two angle velocities, i.e. at those carbon torsion angles spanning the decay window, the H-C=C-H mode reverts in many hMeOp trajectories, apparent by the plateau in Fig. 15 left (blue arrow). In the corresponding time window for Rh (red arrow in Fig. 15), the progression in H-C=C-H is mostly towards larger negative values, similar findings apply to sqRh.

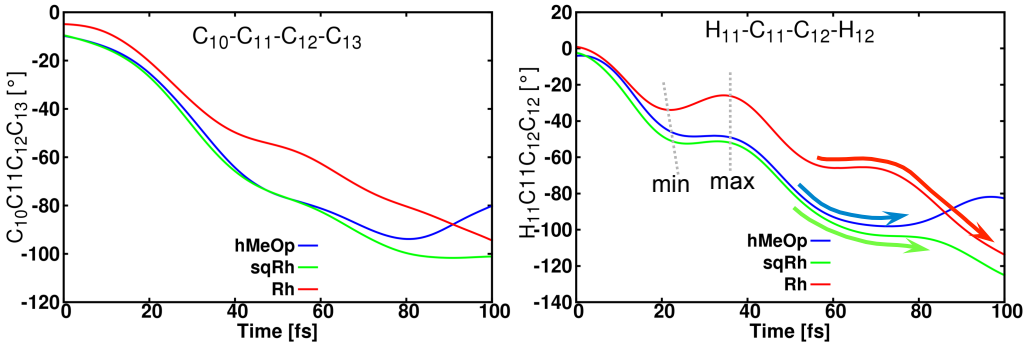


Figure 15. Averaged values of C11=C12 carbon (left) and hydrogen torsion over all trajectories. The arrows denote the direction of the H-C=C-H torsion during the hopping window (Fig. 14).

During these calculations, another important parameter controlling product generation was revealed. This parameter regards the change in wavefunction character in the ground state, which is connected with the bond length alternation coordinate (see ref. [92] for details). In conclusion, the photoreactivity of the studied rhodopsin proteins does not follow a simple one-dimensional Landau-Zener model, which predicts higher quantum yields when the reaction progresses faster along the reactive coordinate. The final outcome is rather related to the phase combination of the decisive modes. A bend in the chromophore structures may speed up the torsion along the carbon fragment, but it has less influence on the involved H-C=C-H motion, leading to phase mismatch and less quantum yield in hMeOp.

3.4 Photoinduced H-transfer in Protochloropyllide Oxidoreductase

The interface of COBRAMM with Gaussian opens the possibility to utilize its versatile set of density functionals for ground (DFT) and excited (TD-DFT) state problems. These include a number of hybrid and double-hybrid functionals, as well as those specifically developed to treat dispersion and long-range correlation effects, allowing for QM/MM partitioning with hundreds of atoms in the QM layer. The capabilities of the COBRAMM/Gaussian interface were recently exploited to refine the ternary structure of the light-driven protochlorophyllide oxidoreductase (LPOR) enzyme of *Thermosynechococcus elongatus*. LPOR plays a crucial role in the light-dependent synthesis of bacteriochlorophyll (BChl) in most photosynthetic organisms. It catalyzes the penultimate step of a complex hydride transfer reaction cascade that reduces protochlorophyllide (Pchlide, PCL) to chlorophyllide (Chlide, CHL). The catalytic reaction involves a light-driven sequential stepwise hydride transfer from the pro-S face of NADPH coenzyme to the C_{17} position of PCL [96–99] followed by the thermally activated protonation at the C_{18} position of PCL with a proton, presumably coming from a highly conserved Tyr residue in the vicinity of PCL [98, 100, 101]. (see Fig. 16). Due to the lack of an LPOR crystal structure a comprehensive understanding of the reaction pathway at room temperature is still far from reach. Surprisingly, there is not yet a consensus even on some fundamental questions about the geometry of the intermediate state(s) and about the structural changes that lie at the origin of the photoactivation process. In the past, all structural/mechanistic interpretations of the available spectroscopic data have been based on different homology models, without ever questioning their validity. We developed a protocol that is in many ways superior to homology modelling as it extends beyond the alignment of related amino acid sequences.

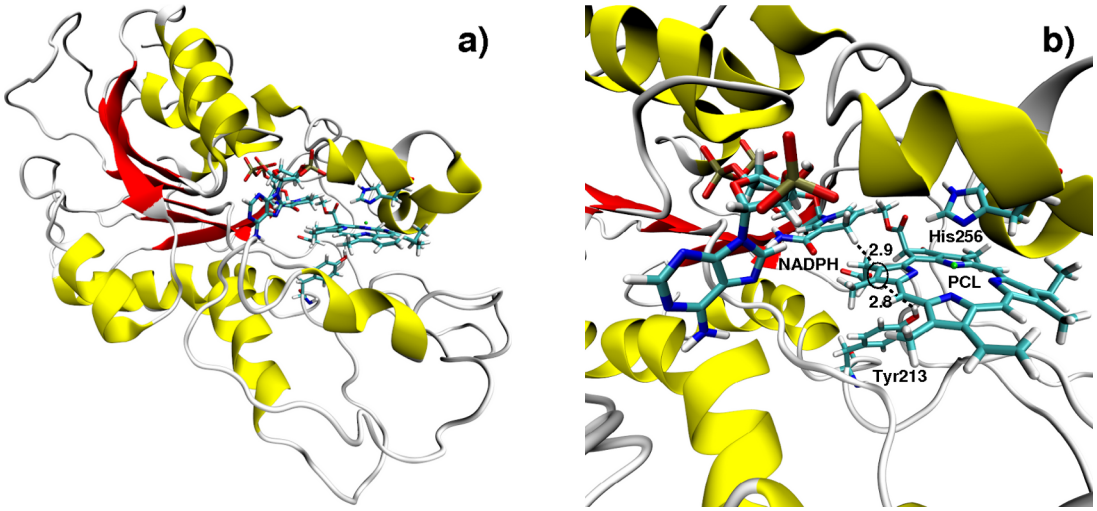


Figure 16. a) Homology model of protochlorophyllide oxidoreductase (LPOR) of *Thermosynechococcus elongatus* in a catalytically competent active site geometry. The structure consists of a central parallel β-sheet (red) surrounded by α-helices (yellow). The structure contains bound NADPH and the docked substrate PCL (shown as licorice representation). b) Catalytically competent active site geometry of PCL in LPOR. Hydride and proton transfer path (PCL(C_{18})-(H)Tyr213 and PCL(C_{17})-(H)NADPH, respectively) have been labeled by dashed lines (distances are given in Å).

Starting from an existing ternary model of NADPH bound LPOR from *Synechocystis sp* (ssPOR) [102] with docked PCL in the active site, the protocol couples various flavours of molecular modelling techniques – molecular mechanics simulations, hybrid quantum mechanics/molecular mechanics calculations, computational spectroscopy – and incorporates the feedback from experiment in a synergetic pipeline in order to assess the fitness of the model on multiple levels and, fundamentally, to suggest improvements. In particular, we made use of the pronounced sensitivity of the high-frequency stretch mode (λ_{max} ~ 1700 cm^{-1}) of the keto group in the cyclopentanone ring (at C_{13}) of PCL to changes in the electronic structure of the conjugated system. The keto-mode has been shown to exhibit a 14 cm^{-1} red-shift upon reduction of the C_{17}-C_{18} double bond in water.[103] The keto group also proves an excellent marker for the hydrogen

bonding environment, exhibiting a remarkable 60 cm^{-1} red-shift going from the polar aprotic tetrahydrofuran (λ_{max} = 1710 cm^{-1}) [104] to the LPOR enzyme environment (λ_{max} = 1650 cm^{-1}) [103], which indicates a strong hydrogen-bond network between PCL and protein residues. Thus, we used the previously published homology model [102] of LPOR, pre-oriented it in a catalytically competent conformation and subjected it to MD simulations and clustering analysis to obtain representative frames of the binding pocket and the protein environment. These were then refined at the QM/MM level and finally used to compute the IR spectrum of LPOR with cofactor and substrate.

The High Layer (QM region) was extended over the pigment (PCL), nicotinamide ring of NADPH, phenol moiety of Tyr213, imidazole ring of HIE256, axial water to the Mg and waters/residues involved in H-bonds with PCL (Figure 16a) to a total of nearly 200 atoms. Protein residues in the vicinity of the QM region as well as waters contributing to the hydrogen bond network were allowed to relax (i.e. movable Medium Layer), while the rest of the system was kept frozen (i.e. Low Layer), thus building a rigid scaffold. The complex QM/MM partitioning and the generation of the MM topology files were achieved in a semi-automatic fashion through a COBRAMM preparatory routine. It took care of the H/M/L partitioning through automatic recognition of hydrogen-bonded amino acids and waters, handled non-standard residues (i.e. PCL, NADPH), charge redistribution of residues shared between QM and MM layers and link atom definitions. Geometry optimizations and frequency computations were performed through the MM(Amber12)/QM(Gaussian09) interface of COBRAMM. The numerical frequency computations were performed on a smaller QM region consisting of PCL, the nicotinamide ring of NADPH, and the phenol moiety of Tyr213 (112 atoms, 1047 basis functions with 6-31G*) by means of massive parallelization over 100 cores using the

COBRAMM native parallel environment in less than one day. Hydrogen-bond analysis of LPOR docked PCL and comparison against free (water-solvated) PCL demonstrated that the LPOR model *does not* offer a more structured hydrogen-bond network around the keto group at C_{13}. On the contrary, the network seems to deteriorate as exemplified by the MD simulations where the keto group has no hydrogen bonding partners in considerable intervals of the simulation time.

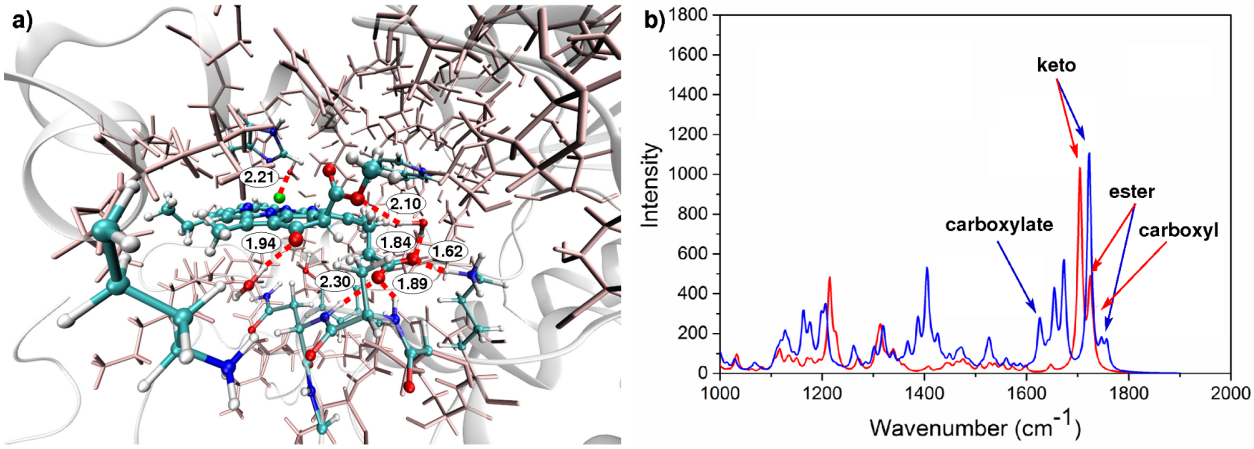


Figure 17. a) Representative snapshot of the catalytic site of the LPOR enzyme with bound NADPH and docked PCL; residues (amino acids and waters) hydrogen-bonded to the carbonyl groups of PCL as well as some relevant distances are highlighted (distances are in Å); residues included in the High, Medium and Low Layers given through CPK, licorice (pink) and tube (white) representation; b) IR spectra of a snapshot of free (water-solvated) PCL (red) and LPOR-bound PCL (blue) with representative hydrogen bonding networks.

Accordingly, we chose a snapshot with an average hydrogen-bond network to simulate the IR spectrum, having three residues Ala168, Asn169 and Lys217 binding to the carboxyl, one water binding simultaneously to the ester and to the carboxyl groups and one water binding to the keto group (Figure 17a). The IR spectrum (Figure 17b) does not show the expected 26 cm^{-1} red-shift

of the keto band (as observed experimentally) with respect to water, instead a blue-shift of 15 cm^{-1} is encountered (1722 cm^{-1}).

The result demonstrates that the modified LPOR enzyme model, while offering a catalytically competent active site, is not able to reproduce the available spectroscopic data. The deficiency of the model is expressed in the lack of a strong hydrogen-bond network around the keto group believed to be responsible for the extraordinary red-shift of its IR frequency. On the basis of these findings we refined the snapshot to introduce a hydrogen bond between the keto group and Lys269, located on the flanking α-helix above the PCL binding site (hereafter referred to as PCL-K269). K269 was selected due its positive charge and proximity to the PCL keto group during the MD simulations. During the refinement of the models minimal deformation of the active site was imperative. Figure 18a documents the changes in the structure (in red) with respect to original snapshot (in blue). Evidently, the largest deformation is exercised on K269 itself and on the water originally involved in the hydrogen bond with the keto group, whereas the nearby residues experience minor deformations. The so obtained snapshot was further refined using the same QM/MM protocol described above and, subsequently, a frequency computation was performed for the equilibrated structure. The IR spectra are depicted in Figure 18b.

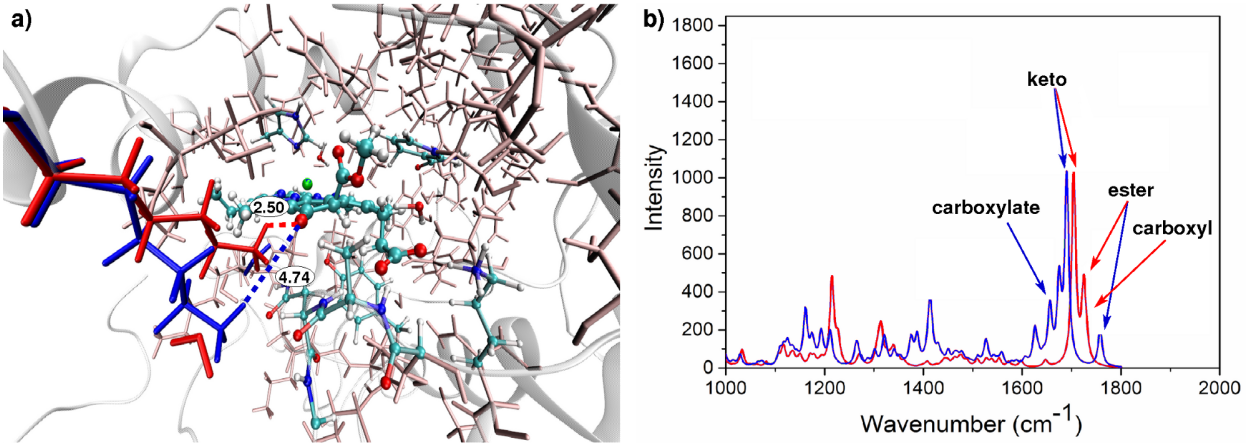


Figure 18. a) Representative snapshot of the catalytic site of the LPOR enzyme with bound NADPH and docked PCL modified to accommodate a hydrogen bond between the keto group at C_{13} and residue K269. Original position (as in Figure 16a) of K269 and proximal water shown in blue, new position shown in red; b) IR spectra of a snapshot of free (water-solvated) PCL (red, same as in Fig. 17b) and the refined LPOR-bound PCL (blue).

Remarkably, a moderate red-shift of 14 cm^{-1} with respect to water is encountered for PCL-K269 (1690 cm^{-1}). This observation makes PCL-K269 a possible candidate for a catalytically competent model of the LPOR enzyme. Yet, quantitative agreement was not reached in IR spectra simulations. Thus, it appears likely that inside the enzyme PCL coordinates a water molecule to the keto group in addition to K269. Indeed, geometries with a close contact of the keto group to both Lys269 and a water molecule are sampled during the MD simulations.

3.5 Deactivation channels in water-solvated pyrimidine nucleosides

COBRAMM connects to various ground and excited state features of the Molcas 8 software, including several multiconfigurational approaches such as CASSCF/RASSCF, their second-order perturbation theory extensions (CASPT2/RASPT2) in both single-state and multi-state (MS and XMS) flavors, as well as a wide range of single-reference methods (MP2, CCSD and CCSD(T) among others), all of those combined with highly efficient resolution of identity schemes framed within the Cholesky decomposition [105, 106] and Frozen Natural Orbitals (FNO) [6, 107]. A recent example showcasing the capabilities of the COBRAMM/Molcas interface are the studies on deciphering the photoinduced deactivation channels in DNA/RNA water-solvated pyrimidine nucleosides oxy-Uridine, deoxy-Thymidine and oxy-Cytidine (nucleobases with a ribose sugar group) [108, 109]. Oxy-Uridine (referred to as Urd hereafter) will be considered as a representative case, as it displays a range of characterized excited state minima, transition states (TSs), conical intersections (CoIns) and intersystem crossings (ISCs) depicting the complex photochemical scenario undergone by this species upon excitation. Whilst these excited-state driven properties have been extensively studied theoretically in the gas-phase, less is known about the concrete molecular motions characterizing them in solution. Moreover, most of the simulations carried out in solution are based on polarizable continuum models, where the atomistic details of the specific solvent to solute (short-range) interactions are lost.

Water has been shown to slow down the excited state decay of the DNA/RNA nucleobases and nucleosides with respect to their *in vacuo* counterparts [110] while also showing a different excited state population branching along the diverse neighboring states, which strongly modulates its overall photophysics. Given most experimental evidence is recorded in water solution due to being the biologically relevant solvent, QM/MM schemes such as those contained in COBRAMM are appealing to treat these sorts of problems. An example is given in Figure 19, where the specific QM/MM set-up employed to study Urd embedded in a water droplet is shown. A closer look is given in Figure 19b, where the specific High/Medium/Low layer partitioning

employed in the simulations is depicted: the high layer treated at the QM level (CASPT2) consists on the uracil nucleobase, whereas the ribose (sugar) group and all water molecules within 5 Å of the nucleobase center of mass are depicted in grey and belong to the mobile Medium layer, the rest of the water molecules being frozen in the Low layer.

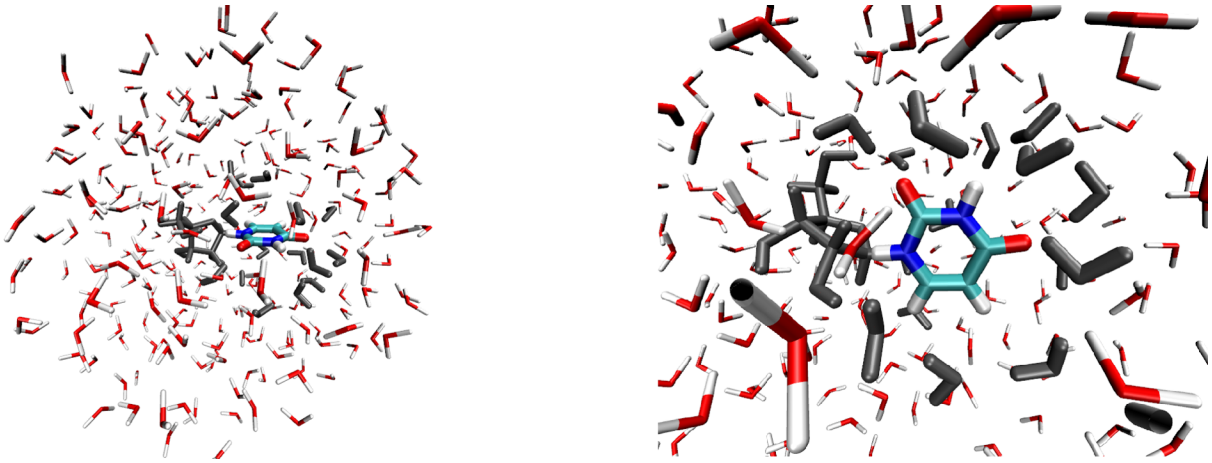


Figure 19. a) Urd with a water droplet cut within a 12 Å radius from the Uracil moiety; b) High/Medium/Low layer partitioning used for MD simulations; sugar and waters in 5 Å distance to center of mass of Uracil (High layer) added to the Medium layer are shown in grey.

In this scheme, the first solvation layer of water is included in the optimization procedures, allowing it to relax along the excited state deactivation of the nucleobase and thus accounting for the short-range interactions present due to the hydrogen bonding motifs formed between the surrounding water molecules and uracil.

Figure 20 displays an overall scheme of the many deactivation channels available to water-solvated Urd. White arrows therein denote deactivation routes attributed to occur in an ultrafast (sub- to few-ps) timescale due to initial excitation into $^1\pi\pi^*$ bright states and subsequent decay to the ground state either through a ring-puckering "ethene-like" conical intersection (right hand side of Figure 20), associated with a barrier of ~0.1 eV (see also ref. [111, 112] for $^1\pi\pi^*$ and $^1n\pi^*$ deactivation channels and ref. [113] for time-resolved IR spectra). Minima, transition states (TSs) and interstate crossing points (CoIns) are characterized at the MS-CASPT2 level

employing COBRAMM's numerical routines for computing gradients and non-adiabatic couplings, described in Section 2.7 and employing the Gaussian optimizer as described in Section 2.3. Frequencies at the MS-CASPT2 level were performed at critical points (minima and TS) to confirm their nature. As mentioned in Section 2.5.5, absorption cross sections can be computed within the COBRAMM interface by means of Wigner sampling [78]. This option is used here to simulate the absorption spectrum of Urd. 500 snapshots were generated at room temperature, thereby excluding high-frequency modes belonging to C-H and N-H stretchings from the sampling. The electronic structure of each snapshot was computed at the MS-CASPT2/SA-CASSCF level using a full valence-π active space comprising 10 electrons in 8 orbitals (i.e. CAS(10,8)), averaging over four states and using the ANO-L basis set adopting contractions 3s2p1d on non-hydrogen and 2s1p on hydrogen atoms

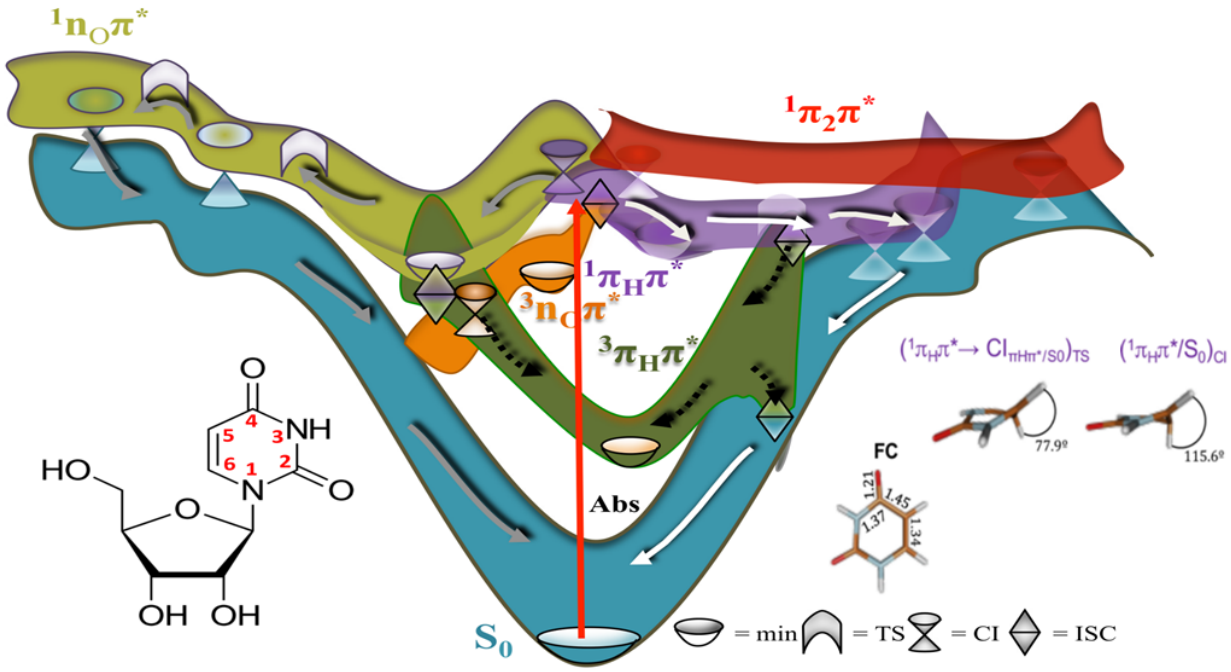


Figure 20. Schematic representation of the photo-excited phenomena triggered in water-solvated Uridine upon UV-light exposure. White arrows represent deactivation channels that are ultrafast (sub- to few-ps), grey denotes long-lived channels based on optically dark 1nπ* states

(several- to dozens of ps) and black shows the longer-lived triplet population (dozens of ps to ns). On the right hand side, concrete critical points characterized along the potential energy surfaces of initially accessed and ultrafast $^1\pi\pi^*$ state deactivations are shown. Figure adapted from ref. [108] with permission from the Royal Society of Chemistry.

The result is shown in Figure 21a). The experimental spectrum of Urd above 220 nm is characterized by a single intense band peaking at 263 nm (magenta line in Figure 21). The full valence-π CAS(10,8) reproduces the bandwidth, yet red-shifts the band significantly by ca. 20 nm (~0.35 eV at this wavelength). In a series of studies on various chromophoric systems we have demonstrated recently, that this effect is general and due to the overestimation of the correlation energy with perturbative methods like CASPT2 with full valence-π active spaces [26, 27, 114, 115]. Increasing the active space by including extra-valence virtual orbitals rectifies the problem as seen in Figure 21b).

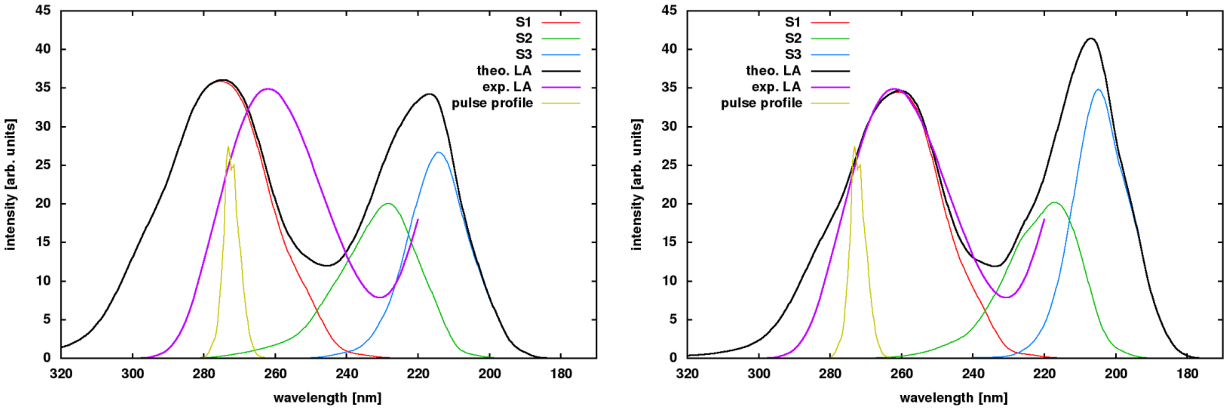


Figure 21. Comparison of the simulated (black) and experimental (magenta) linear absorption spectra of Uridine in aqueous solution. Simulations performed at the MS-4-CASPT2/SA-4-CASSCF(10,8) (left) and MS-4-RASPT2/SA-4-RASSCF(10,8|2,8) (right) levels of theory. Also shown are the individual contributions to the total spectrum from the adiabatic states S_1-S_3 and a typical pump pulse envelope used in time-resolved studies of the systems

In this particular case, by extending the full valence-π active space by an additional eight extra-valence virtual π-orbitals treated at RASSCF level, where up to two simultaneous excitations are allowed, we achieve a near quantitative agreement with the experimental data. The treatment of

such large active spaces is made possible through the restricted active space feature available in Molcas coupled to the Cholesky utility, which reduces the number of RASPT2 terms and speeds up their calculation [7].

The interface of COBRAMM with Molcas allows running trajectory-based non-adiabatic mixed quantum-classical dynamics simulations applying the Tully's fewest switches surface hopping algorithm [116]. In the following we use Urd to showcase these features. As an example, Figure 22 compares simulations at the MS-4-CASPT2/SA-4-CASSCF(10,8) (Figure 22a,d), the SS-CASPT2/SA-2-CASSCF(10,8) (Figure 22b,e) and the SS-CASPT2/SA-2-CASSCF(10,10) (Figure 22c,f) levels of theory run propagated for a total of 200 fs with a time-step of 1.0 fs. To afford this relatively long time-step we kept the high-frequency modes frozen during the sampling procedure. Furthermore, we used the Tully-Hammes-Schiffer (THS) variation [117] of the fewest switches surface hopping scheme which is more suitable to avoid overstepping confined regions in coordinate space with non-adiabatic coupling peaks. The movable layer of the system was increased to include several water layers (on average 45 waters per snapshot) to allow the water solvation shell to adapt to the puckering deformation of the Uracil moiety. Gradients were computed numerically through a two-point finite differentiation formula using the COBRAMM's parallel environment. To speed up the numerical computations of the SS-CASPT2 gradients we made use of the fact that the perturbative correction is applied independently on each root. Thus, while at the reference point all states are corrected at SS-CASPT2 level, during the computation of the gradient only the state that drives the nuclear dynamics (i.e. the photoactive state) was taken into account. Care was taken for possible state swapping at each displaced geometry in case of near-degeneracy at the CASSCF level by following the nature of the wave function. To speed up the numerical computations in the case of the MS-CASPT2 gradients, instead, the MS-CASPT2 eigenvector of the photoactive state at the reference geometry was analyzed and only CASSCF roots that contribute considerably to the

MS-CASPT2 wave function were taken into account. Internal routines monitor the stability of the simulations. If

a) individual contributions to the numerical gradient exceed a threshold (default 0.5)

b) the ratio of the lengths of the gradient vectors at two consecutive steps (i.e. ||g(t)||/||g(t-dt)||) and, at the same time, the deviation of the total energy from its value at the previous step (i.e. Etot(t)-Etot(t-dt)) exceed certain thresholds (defaults 1.3 and 0.5 kcal/mol, respectively)

c) a single point computation fails to converge

the gradient is discarded and the Velocity Verlet step is repeated at the last stable geometry, first with a half step (in this case 0.5 fs) and, if the thresholds are exceeded again, with a double step (in this case 2.0 fs). Furthermore, the wave function of the photoactive state is followed by computing the overlap between two consecutive time steps. A warning is issued in case the overlap is below 0.9 thus indicating possible intruder states (e.g. in this particular case the intrusion of nπ* states due to lone pair rotation in the active space).

A set of trajectories was run with both CASPT2 flavours (SS and MS). The decay to the ground state occurs exclusively through the ring-puckering "ethene-like" CI (compare insets in Figure 22 to structure number 5 in Figure 20b). Thereby, we observe splitting of the wavepacket towards two ring-puckering CIs with contrariwise orientation of hydrogens H9 and H10. In the limited set of trajectories run for this study we find no evidence for decay through bond breaking. The comparison between the two CASPT2 flavors shows a considerably slower dynamics in the case of MS-CASPT2, whereas virtually all trajectories have already decayed to the GS after 200 fs in the case of SS-CASPT2. While in some cases SS and MS simulations agree well (Figure 22a,b), in many cases we observe that the system quickly leaves the FC region in order to roam then for hundreds of fs in the CI region at the MS-CASPT2 level (Figure 22d) opposite to trajectories run at the SS-CASPT2 level which decayed in a ballistic fashion (Figure 22e). The reason for this discrepancy is an artifact of the MS-CASPT2 approach with small

active spaces that gives rise to relatively large off-diagonal S_0-S_1 elements in the MS-CASPT2 Hamiltonian (an order of magnitude larger than the value of 0.002 suggested in the seminal paper on MS-CASPT2 [118]). While not posing a problem in the FC region where the S_0-S_1 energy gap is large these become important in the case when the two states come energetically close. In a direction in which the CASSCF wave functions mix, the MS-CASPT2 routine over-estimates the level of un-mixing thus pushing the two states apart, introducing an artificial barrier in the excited state and preventing an efficient decay. The only way for the excited state to decay to the ground state is in the form of a diabatic crossing where the two states do not mix.

The MS- and SS-CASPT2 dynamics with the full valence-π AS are overall stable, showing a tolerable increase of the total energy over the course of the simulation which reaches ca. 10 kcal/mol. There are several factors that play a role in this issue. Apart from the 1.0 fs time step, which is insufficient to describe fast stretching dynamics or the large amplitude dynamics of the "hot" ground state, Figure 22c,f exemplifies the issues arisen due to active space size. Increasing the full valence-π active space by two extra-valence orbitals reduces substantially the deviation of the total energy. The improved accuracy comes at the price of an increased cost in the computation, only affordable with restricted active spaces (i.e. RASPT2). We note, however, that the two sets of trajectories, SS-CASPT2/SA-2-CASSCF(10,8) (red) and SS-CASPT2/SA-2-CASSCF(10,10) (black), behave qualitatively similarly, i.e. the active space increase does not affect the overall photophysics displayed by the system.

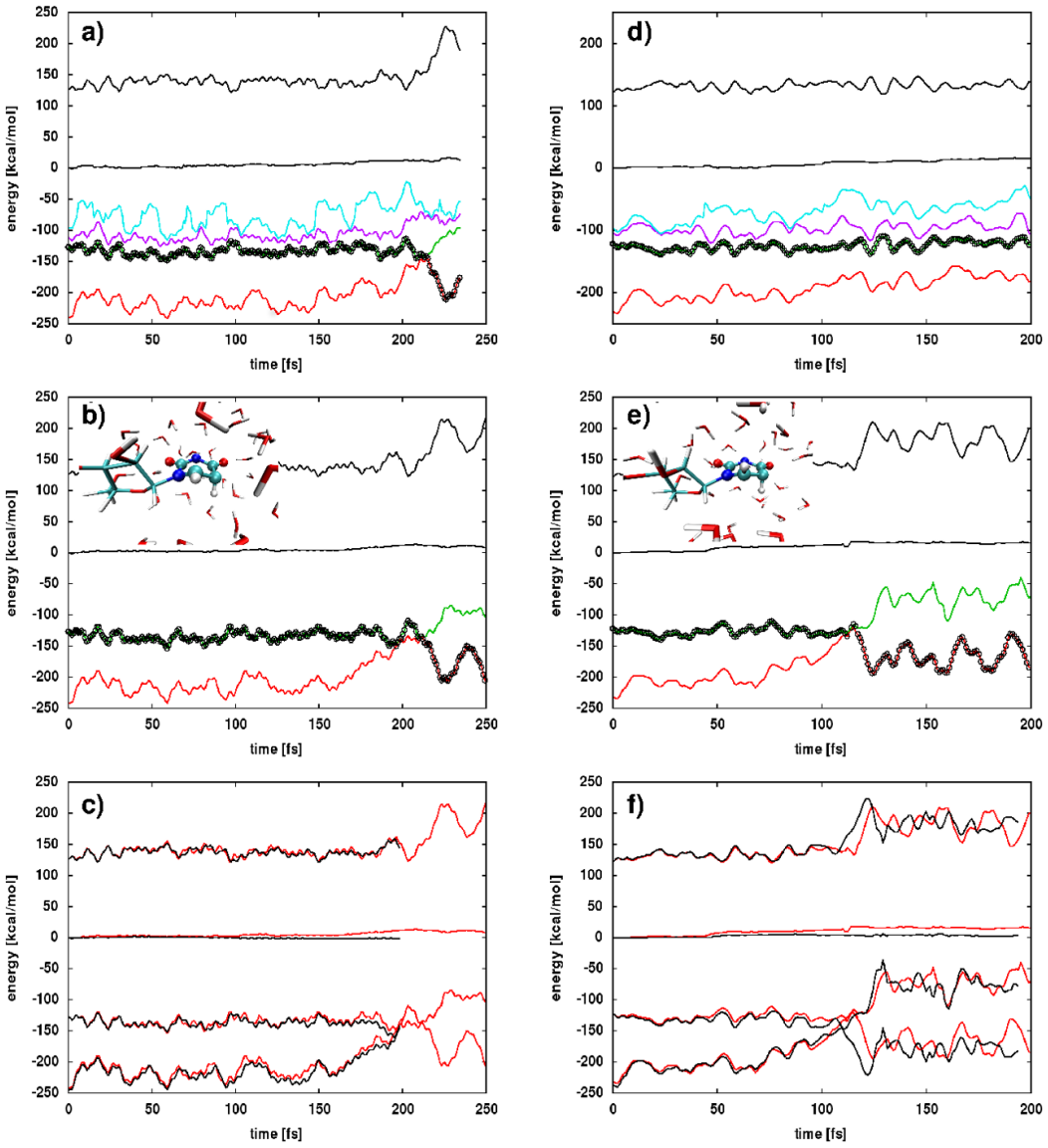


Figure 22. Comparison of simulations run at the MS-4-CASPT2/SA-4-CASSCF(10,8) (a,d), the SS-CASPT2/SA-2-CASSCF(10,8) (b,e) and the SS-CASPT2/SA-2-CASSCF(10,10) (c,f) levels of theory for two snapshots. Color code for a,b,d and e: S_0 (red), S_1 (green), S_2 (magenta), S_3 (cyan), photoactive state (black dotted), kinetic and total energy (black); Color code for c and f: SS-CASPT2/SA-2-CASSCF(10,8) level (red), SS-CASPT2/SA-2-CASSCF(10,10) (black). The insets show the Uridine conformation at the crossing points.

4 Conclusions

We have reported on new developments and additions to our modular QM and QM/MM interface program COBRAMM, enabling efficient electronic structure computations in ground and excited states. The new developments include algorithms to perform surface hopping dynamics with an arbitrary number of states and decoherence correction using CASSCF, CASPT2 and OMx-MRCI methods. A selector scheme allows the computation of only those gradients and derivative couplings strictly necessary for the current step of the MD. Efficient numerical procedures allow the parallelization of numerical frequency calculations for all implemented methods and speed up gradient and coupling calculations for single state and multistate CASPT2 methods. The modified atom-link implementation restricts the movement of the link atom to the bond axis of the QM/MM boundary. A Rattle version of the velocity Verlet algorithm can freeze out high frequency motions like e.g. C-H or O-H, thus reducing total energy fluctuations and enabling the use of larger time steps. Several tools were added to simplify the QM/MM setup. A script driven routine helps to select QM and MM portions as well as to define mobile and static regions. Within these tools, computations defining the initial conditions for MD simulations can be performed. Further routines allow simplified generation of static and time-resolved absorption spectra.

The capabilities of the implementations have been demonstrated with a number of examples. CASSCF/Amber studies of the rhodopsin photoisomerization could reproduce time resolved spectral traces of a pump-probe experiment. A benchmark calculation applying OMx-MRCI methods to the same reaction revealed good overall agreement with the CASSCF QM/MM trajectory computations. In combined DFT and DFT/MRCI studies we find indications for a fast singlet and triplet crossing in benzophenone. Photodynamics of the human melanopsin protein and squid rhodopsin were simulated by combining features from different QM programs: wavefunction computations with Molcas and non-adiabatic couplings from Molpro, demonstrating the interoperability of these routines. Both, sqRh and hMeOp show faster decay in

comparison to bovine rhodopsin, but lead to a lower quantum yield. The drastic reduction in hMeOp quantum yield stems from a phase mismatch of the product generating H-C=C-H motion at the isomerizing double bond. DFT-based QM/MM computations assisted in refining the structure of the LPOR enzyme. The results indicate an advanced hydrogen bonded network that involves a keto-group of the PCL chromophore, a water molecule and a neighboring lysine residue. The numerical routines available within the MOLCAS interface have been extensively used to revisit the photophysics and photochemistry of water-solvated Uridine, enabling the unprecedented mapping of potential energy hypersurfaces at the MS-CASPT2 level within a QM/MM scheme that considers environmental effects. These have also allowed the simulation of non-adiabatic molecular dynamics within different schemes. The latter feature a series of on-the-fly checks that ensure the quality of the simulation and simplify the tasks required by the user, facilitating their use for non-specialists and highlighting COBRAMM's philosophy of extending the applicability of the different features available in a user-friendly manner.

AUTHOR INFORMATION

Corresponding Author

*Marco Garavelli, email: marco.garavelli@unibo.it, address: Dipartimento di Chimica Industriale, Università degli Studi di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy.

Present Addresses

† Department of Chemistry, Imperial College London, SW7 2AZ London, UK

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Funding Sources

I.D. is grateful for financial support through "Strategischer Forschungsfonds" of the Heinrich-Heine-University Düsseldorf, project no. F 2013 – 442-9. J.S.M. acknowledges support from the European Commission through the Marie Curie actions (FP8-MSCA-IF, grant nº 747662). I.R., M.G. and J.S.M thank the Agence National de la Recherche project FEMTO-2DNA (ANR-15-CE-29-0010).

Notes

The COBRAMM set of routines can be obtained free of charge upon request to the corresponding author.

ACKNOWLEDGMENT

The COBRAMM project was started by Piero Altoé and Marco Stenta under guidance of Marco Garavelli and Andrea Bottoni at the university of Bologna. The current development of COBRAMM is led by Marco Garavelli and his group in Bologna, in collaboration with Oliver Weingart, Heinrich-Heine University of Düsseldorf. Major contributions to the code were provided by Piero Altoé, Marco Stenta, Oliver Weingart, Artur Nenov, Irina Dokukina, Ivan Rivalta, Javier Segarra-Martí, Emiliano Poli and Salvatore F. Altavilla. We thank Baptiste Demoulin, Irene Conti, Ana Julieta Pepino and Mohsen El-Tahawy for extensive testing and help

with debugging. We furthermore acknowledge the kind help of Francesco Aquilante when implementing certain options of the MOLCAS interface.

I.R. gratefully acknowledges the use of HPC resources of the “Pôle Scientifique de Modélisation Numérique” (PSMN) at the ENS-Lyon (France).

ABBREVIATIONS

2DES	Two-dimensional electronic spectroscopy
2DUV	Two-dimensional UV spectrsoscopy
B3LYP	Becke's 3 Parameter exchange in combination with Lee, Young and Parr correlation functional
BChl	Bacteriochlorophyll
BFGS	Broydon-Fletcher-Goldfarb-Shannon
BH-LYP	Becke's half-and-half exchange in combination with LYP correlation functional
BP	Benzophenone
CAM	Coulomb attenuated method
CAS	Complete active space
CC	Coupled Cluster
CCSD	Coupled cluster singles and doubles
CHL	Chlorophyllide
CI	Configuration interaction
CoIn	Conical intersection
CP-MCSCF	Coupled perturbed multiconfigurational SCF
DFT	Density functional theory
EOM	Equation-of-motion
FC	Franck-Condon
FNO	Frozen natural orbital
Fs	femtosecond
HF	Hartree-Fock
hMeOp	Human Melanopsin

HOMO	Highest occupied molecular orbital
HPC	High performance computing
IR	Infrared
IRC	Intrinsic reaction coordinate
ISC	Intersystem crossing
LMP2	Local MP2
LPOR	Protochlorophyllide oxidoreductase
LUMO	Lowest unoccupied molecular orbital
MD	Molecular dynamics
MEP	Minimal energy path
MM	Molecular mechanics
MNDO	Modified neglect of differential overlap (a semi-empirical method and a program package distributed by Thiel et al.)
MP	Moller Plesset (e.g. MP2, MP4)
MRCI	Multi-reference CI
MS	Multi state
NAC	Non-adiabatic coupling
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NVE	microcanonical ensemble (constant number of particles, volume and energy)
OMx	Orthogonalization method x
PCL	Protochlorophyllide
PCM	Polarizable continuum model
Ps	picosecond
QM	Quantum mechanics
QY	Quantum Yield
RAS	Restricted active space
Rh	Rhodopsin
SA	State-average
SCF	Self-consistend field
SOS	Sum-over-states

SqRh	Squid rhodopsin
SS	Single state
ssPOR	LPOR from Synechocystis sp
SVP	Single zeta valence basis set including polarization functions
TCL	Tool command language
TD	Time-dependent
TDC	Time derivative coupling
TDSE	Time-dependent Schrödinger equation
THS	Tully-Hammes-Schiffer (surface hopping scheme)
TIP3P	Three-site transferrable intermolecular potential (MM water model)
TS	Transition state
Urd	Oxy-uridine
ZPE	Zero-point energy

REFERENCES

1. Frisch MJ, Trucks GW, Schlegel HB, et al Gaussian 09 Revision E.01

2. TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

3. Grimme S, Waletzke M (1999) A combination of Kohn – Sham density functional theory and multi-reference configuration interaction methods. J Chem Phys 111:5645–5655 . doi: 10.1063/1.479866

4. Lyskov I, Kleinschmidt M, Marian CM (2016) Redesign of the DFT/MRCI Hamiltonian. J Chem Phys 144:34104 . doi: http://dx.doi.org/10.1063/1.4940036

5. Aquilante F, Autschbach J, Carlson RK, et al (2016) Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. J Comput Chem 37:506–541 . doi: 10.1002/jcc.24221

6. Segarra-Martí J, Garavelli M, Aquilante F (2015) Multiconfigurational Second-Order Perturbation Theory with Frozen Natural Orbitals Extended to the Treatment of

Photochemical Problems. J Chem Theory Comput 11:3772–3784 . doi: 10.1021/acs.jctc.5b00479

7. Aquilante F, Malmqvist P-Å, Pedersen TB, et al (2008) Cholesky Decomposition-Based Multiconfiguration Second-Order Perturbation Theory (CD-CASPT2): Application to the Spin-State Energetics of CoIII(diiminato)(NPh). J Chem Theory Comput 4:694–702 . doi: 10.1021/ct700263h

8. Werner H-J, Knowles PJ, Knizia G, et al (2010) MOLPRO, version 2010.1, a package of ab initio programs

9. Watts JD (2008) An Introduction to Equation-of-Motion and Linear-Response Coupled-Cluster Methods for Electronically Excited States of Molecules. In: Shukla MK, Leszczynski J (eds) Radiation Induced Molecular Phenomena in Nucleic Acids. Springer Netherlands, Dordrecht, pp 65–92

10. Kong L, Bischoff FA, Valeev EF (2012) Explicitly Correlated R12/F12 Methods for Electronic Structure. Chem Rev 112:75–107 . doi: 10.1021/cr200204r

11. Werner H, Köppl C, Ma Q, Schwilk M (2017) Explicitly Correlated Local Electron Correlation Methods. In: Fragmentation. John Wiley & Sons, Ltd, Chichester, UK, pp 1–79

12. Dewar MJS, Thiel W (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. J Am Chem Soc 99:4899–4907 . doi: 10.1021/ja00457a004

13. Kolb M, Thiel W (1993) Beyond the MNDO model: Methodical considerations and numerical results. J Comput Chem 14:775–789 . doi: 10.1002/jcc.540140704

14. Tuttle T, Thiel W (2008) OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. Phys Chem Chem Phys 10:2159–2166 . doi: 10.1039/B718795E

15. Tuna D, Lu Y, Koslowski A, Thiel W (2016) Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States. J Chem Theory Comput 12:4400–4422 . doi: 10.1021/acs.jctc.6b00403

16. Altoè P, Stenta M, Bottoni A, Garavelli M (2007) A tunable QM/MM approach to

chemical reactivity, structure and physico-chemical properties prediction. Theor Chem Acc 118:219–240 . doi: 10.1007/s00214-007-0275-9

17. Case DA, Babin V, Berryman JT, et al (2014) Amber 14 OR - University of California, San Francisco

18. Hratchian HP, Schlegel HB (2005) Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces. In: Frenking G, Kim KS, Scuseria G (eds) Theory and Applications of Computational Chemistry: The first 40 years. Elsevier, Amsterdam, pp 195–249

19. Verlet L (1967) Computer experiments on classical fluids. Phys Rev, 159:98–103

20. Andersen HC (1983) Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. J Comput Phys 52:24–34 . doi: http://dx.doi.org/10.1016/0021-9991(83)90014-1

21. Bearpark MJ, Robb MA, Bernhard Schlegel H (1994) A direct method for the location of the lowest energy point on a potential surface crossing. Chem Phys Lett 223:269–274 . doi: https://doi.org/10.1016/0009-2614(94)00433-1

22. Dokukina I, Weingart O (2015) Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin. Phys Chem Chem Phys 17: . doi: 10.1039/c5cp02650d

23. Mukamel S (1995) Principles of Nonlinear Optics and Spectroscopy. Oxford Unversity Press, Oxford, UK

24. Nenov A, Rivalta I, Cerullo G, et al (2014) Disentangling Peptide Configurations via Two-Dimensional Electronic Spectroscopy: Ab Initio Simulations Beyond the Frenkel Exciton Hamiltonian. J Phys Chem Lett 5:767–771 . doi: 10.1021/jz5002314

25. Nenov A, a Beccara S, Rivalta I, et al (2014) Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study. ChemPhysChem 15:3282–3290 . doi: 10.1002/cphc.201402374

26. Nenov A, Rivalta I, Mukamel S, Garavelli M (2014) Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles. Comput Theor Chem 1040–1041:295–303 . doi: https://doi.org/10.1016/j.comptc.2014.03.031

27. Nenov A, Giussani A, Segarra-Martí J, et al (2015) Modeling the high-energy electronic state manifold of adenine: Calibration for nonlinear electronic spectroscopy. J Chem Phys 142:212443 . doi: 10.1063/1.4921016

28. Ivan R, Artur N, Giulio C, et al (2013) Ab initio simulations of two□dimensional electronic spectra: The SOS//QM/MM approach. Int J Quantum Chem 114:85–93 . doi: 10.1002/qua.24511

29. Segarra-Martí J, Nenov A, Mukamel S, et al (2018) Towards Accurate Simulations of Two-Dimensional Electronic Spectroscopy. Top Curr Chem submitted

30. Weingart O (2017) Combined Quantum and Molecular Mechanics (QM/MM) Approaches to Simulate Ultrafast Photodynamics in Biological Systems. Curr Org Chem 21:586–601 . doi: 10.2174/1385272821666161108150421

31. Tully JC, Preston RK (1971) Trajectory Surface Hopping Approach to Nonadiabatic Molecular Collisions: The Reaction of H+ with D2. J Chem Phys 55(2):562–572

32. Tully JC (1990) Molecular dynamics with electronic transitions. J Chem Phys 93:1061–1071

33. Hayashi S, Tajkhorshid E, Schulten K (2009) Photochemical Reaction Dynamics of the Primary Event of Vision Studied by Means of a Hybrid Molecular Simulation. Biophys J 96:403–416 . doi: 10.1016/j.bpj.2008.09.049

34. Weingart O, Garavelli M (2012) Modelling vibrational coherence in the primary rhodopsin photoproduct. J Chem Phys 137:523 . doi: 10.1063/1.4742814

35. Polli D, Weingart O, Brida D, et al (2014) Wavepacket splitting and two-pathway deactivation in the photoexcited visual pigment isorhodopsin. Angew Chemie - Int Ed 53: . doi: 10.1002/anie.201309867

36. Quenneville J, Martı TJ (2000) Ab Initio Multiple Spawning□: Photochemistry from First Principles Quantum Molecular. J Phys Chem A 104:5161–5175 . doi: 10.1021/jp994174i

37. Zhu C, Jasper AW, Truhlar DG (2005) Non-Born-Oppenheimer Liouville-von Neumann Dynamics. Evolution of a Subsystem Controlled by Linear and Population-Driven Decay of Mixing with Decoherent and Coherent Switching. J Chem Theory Comput 1:527–540 . doi: 10.1021/ct050021p

38. Granucci G, Persico M (2007) Critical appraisal of the fewest switches algorithm for surface hopping. J Chem Phys 126:134114 . doi: 10.1063/1.2715585

39. Doltsinis NL (2002) Nonadiabatic Dynamics: Mean-Field and Surface Hopping. In: Grotendorst J, Marx D, Muramatsu A (eds) Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes, NIC series. John von Neumann Institute for Computing, Jülich, Germany, pp 377–397

40. Fabiano E, Keal TW, Thiel W (2008) Implementation of surface hopping molecular dynamics using semiempirical methods. Chem Phys 349:334–347 . doi: 10.1016/j.chemphys.2008.01.044

41. Weingart O (2008) The role of HOOP-modes in the ultrafast photo-isomerization of retinal models. Chem Phys 349(1):348–355

42. Schapiro I, Ryazantsev MN, Frutos LM, et al (2011) The ultrafast photoisomerizations of rhodopsin and bathorhodopsin are modulated by bond length alternation and HOOP driven electronic effects. J Am Chem Soc 133:3354–3364 . doi: 10.1021/ja1056196

43. Weingart O, Altoè P, Stenta M, et al (2011) Product formation in rhodopsin by fast hydrogen motions. Phys Chem Chem Phys 13:3645–3648 . doi: 10.1039/c0cp02496a

44. Fdez. Galván I, Delcey MG, Pedersen TB, et al (2016) Analytical State-Average Complete-Active-Space Self-Consistent Field Nonadiabatic Coupling Vectors: Implementation with Density-Fitted Two-Electron Integrals and Application to Conical Intersections. J Chem Theory Comput 12:3636–3653 . doi: 10.1021/acs.jctc.6b00384

45. Park JW, Shiozaki T (2017) Analytical Derivative Coupling for Multistate CASPT2 Theory. J Chem Theory Comput 13:2561–2570 . doi: 10.1021/acs.jctc.7b00018

46. Liu L, Liu J, Martinez TJ (2016) Dynamical Correlation Effects on Photoisomerization: Ab Initio Multiple Spawning Dynamics with MS-CASPT2 for a Model trans-Protonated Schiff Base. J Phys Chem B 120:1940–1949 . doi: 10.1021/acs.jpcb.5b09838

47. Pittner J, Lischka H, Barbatti M (2009) Optimization of mixed quantum-classical dynamics: Time-derivative coupling terms and selected couplings. Chem Phys 356:147–152 . doi: https://doi.org/10.1016/j.chemphys.2008.10.013

48. Ou Q, Bellchambers GD, Furche F, Subotnik JE (2015) First-order derivative couplings

between excited states from adiabatic TDDFT response theory. J Chem Phys 142: . doi: http://dx.doi.org/10.1063/1.4906941

49. Zhang X, Herbert JM (2014) Analytic derivative couplings for spin-flip configuration interaction singles and spin-flip time-dependent density functional theory. J Chem Phys 141:64104 . doi: http://dx.doi.org/10.1063/1.4891984

50. Faraji S, Matsika S, Krylov AI (2018) Calculations of non-adiabatic couplings within equation-of-motion coupled-cluster framework: Theory, implementation, and validation against multi-reference methods. J Chem Phys 148:44103 . doi: 10.1063/1.5009433

51. Tajti A, Szalay PG (2009) Analytic evaluation of the nonadiabatic coupling vector between excited states using equation-of-motion coupled-cluster theory. J Chem Phys 131:124104 . doi: 10.1063/1.3232011

52. Toniolo A, Olsen S, Manohar L, Martinez TJ (2004) Conical intersection dynamics in solution: The chromophore of Green Fluorescent Protein. Faraday Discuss 127:149–163 . doi: 10.1039/b401167h

53. Senn HM, Thiel W (2007) QM/MM Methods for Biological Systems. Top Curr Chem 268:173–290 . doi: 10.1007/128

54. Bo C, Maseras F (2008) QM/MM methods in inorganic chemistry. Dalton Trans 22:2911–2919

55. van der Kamp MW, Mulholland AJ (2013) Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. Biochemistry 52:2708–2728 . doi: 10.1021/bi400215w

56. Lin H, Truhlar DG (2006) QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Acc 117:185–199 . doi: 10.1007/s00214-006-0143-z

57. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103:227–249 . doi: http://dx.doi.org/10.1016/0022-2836(76)90311-9

58. Warshel A (1976) Bicycle-pedal model for the first step in the vision process. Nature 260:679–683

59. Stewart JJP (2016) MOPAC2016

60. Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25 . doi: http://dx.doi.org/10.1016/j.softx.2015.06.001

61. Dapprich S, Komáromi I, Byun KS, et al (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J Mol Struct 461–462:1–21 . doi: http://dx.doi.org/10.1016/S0166-1280(98)00475-8

62. Ponder Jay W, Richards Frederic M (2004) An efficient newton□like method for molecular mechanics energy minimization of large molecules. J Comput Chem 8:1016–1024 . doi: 10.1002/jcc.540080710

63. Metz S, Kästner J, Sokol AA, et al (2014) ChemShell-a modular software package for QM/MM simulations. Wiley Interdiscip Rev Comput Mol Sci 4:101–110 . doi: 10.1002/wcms.1163

64. Neese F (2012) The ORCA program system. Wiley Interdiscip Rev Comput Mol Sci 2:73–78 . doi: 10.1002/wcms.81

65. Brooks BR, Brooks CL, Mackerell AD, et al (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30:1545–1614 . doi: 10.1002/jcc.21287

66. Reif MM, Winger M, Oostenbrink C (2013) Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. J Chem Theory Comput 9:1247–1264 . doi: 10.1021/ct300874c

67. Gale JD, Rohl AL (2003) The General Utility Lattice Program (GULP). Mol Simul 29:291–341 . doi: 10.1080/0892702031000104887

68. Smith W, Yong CW, Rodger PM (2002) DL_POLY: Application to molecular simulation. Mol Simul 28:385–471 . doi: 10.1080/08927020290018769

69. Barbatti M, Ruckenbauer M, Plasser F, et al (2014) Newton-X: a surface-hopping program for nonadiabatic molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci 4:26–33 . doi: 10.1002/wcms.1158

70. Altoé P, Stenta M, Bottoni A, Garavelli M (2007) A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. Theor Chem

Acc 118(1):219–240

71. Fletcher R (1987) Practical methods of optimization, 2nd editio. John Wiley & Sons, New York

72. Friesner RA, Guallar V (2005) Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Ezymatic Catalysis. Ann Rev Phys Chem 56:389–427 . doi: 10.1146/annurev.physchem.55.091602.094410

73. Brunk E, Rothlisberger U (2015) Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. Chem Rev 115:6217–6263 . doi: 10.1021/cr500628b

74. Eurenius KP, Chatfield DC, Brooks BR, Hodoscek M (1996) Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. Int J Quantum Chem 60:1189–1200 . doi: 10.1002/(SICI)1097-461X(1996)60:6<1189::AID-QUA7>3.0.CO;2-W

75. König PH, Hoffmann M, and Th. Frauenheim, Cui Q (2005) A Critical Evaluation of Different QM/MM Frontier Treatments with SCC-DFTB as the QM Method. J Phys Chem B 109:9082–9095 . doi: 10.1021/jp0442347

76. Sloane CS, Hase WL (1977) On the dynamics of state selected uni-molecular reactions: Chloroacetylene dissociation and predissociation. J Chem Phys 66(4):1523–1533

77. Lan Z, Fabiano E, Thiel W (2009) Photoinduced nonadiabatic dynamics of pyrimidine nucleobases: on-the-fly surface-hopping study with semiempirical methods. J Phys Chem B 113:3548–55 . doi: 10.1021/jp809085h

78. Du L, Lan Z (2015) An On-the-Fly Surface-Hopping Program JADE for Nonadiabatic Molecular Dynamics of Polyatomic Systems: Implementation and Applications. J Chem Theory Comput 11:1360–1374 . doi: 10.1021/ct501106d

79. Ruckenbauer M, Barbatti M, Müller T, Lischka H (2013) Nonadiabatic photodynamics of a retinal model in polar and nonpolar environment. J Phys Chem A 117:2790–9 . doi: 10.1021/jp400401f

80. Polli D, Altoè P, Weingart O, et al (2010) Conical intersection dynamics of the primary photoisomerization event in vision. Nature 467:440–443 . doi: 10.1038/nature09346

81. Koslowski A, Beck ME, Thiel W (2003) Implementation of a General Multireference Configuration Interaction Procedure with Analytic Gradients in a Semiempirical Context Using the Graphical Unitary Group Approach. J Comput Chem 24:714 –726

82. Fabiano E, Thiel W (2008) Nonradiative Deexcitation Dynamics of 9H-Adenine: An OM2 Surface Hopping Study. J Phys Chem A 112:6859?6863

83. Silva-Junior MR, Thiel W (2010) Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. J Chem Theory Comput 6:1546–1564 . doi: 10.1021/ct100030j

84. Sanchez-Garcia E, Doerr M, Thiel W (2010) QM/MM study of the absorption spectra of DsRed.M1 chromophores. J Comput Chem 31:1603–1612 . doi: 10.1002/jcc.21443

85. Spörkel L, Thiel W (2016) Adaptive time steps in trajectory surface hopping simulations. J Chem Phys 144:194108 . doi: http://dx.doi.org/10.1063/1.4948956

86. Klamt A (1995) Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. J Phys Chem 99:2224–2235 . doi: 10.1021/j100007a062

87. Marazzi M, Mai S, Roca-Sanjuán D, et al (2016) Benzophenone Ultrafast Triplet Population: Revisiting the Kinetic Model by Surface-Hopping Dynamics. J Phys Chem Lett 7:622–626 . doi: 10.1021/acs.jpclett.5b02792

88. Cuquerella MC, Lhiaubet-Vallet V, Cadet J, Miranda MA (2012) Benzophenone Photosensitized DNA Damage. Acc Chem Res 45:1558–1570 . doi: 10.1021/ar300054e

89. Shah BK, Rodgers MAJ, Neckers DC (2004) The S2 □ S1 Internal Conversion of Benzophenone and p-Iodobenzophenone. J Phys Chem A 108:6087–6089 . doi: 10.1021/jp049615z

90. Zvereva E, Segarra-Martí J, Marazzi M, et al (2018) The effect of solvent relaxation in the ultrafast time-resolved spectroscopy of solvated benzophenone. Photochem Photobiol Sci 17: . doi: 10.1039/c7pp00439g

91. Segarra-Martí J, Zvereva E, Marazzi M, et al (2018) Resolving the Singlet Excited State Manifold of Benzophenone by First-Principles Simulations and Ultrafast Spectroscopy. J Chem Theory Comput. doi: 10.1021/acs.jctc.7b01208

92. El-Tahawy M, Nenov A, Weingart O, et al (2018) On the relationship between excited state lifetime and isomerization quantum yield in animal rhodopsins: beyond the one-dimensional Landau-Zener model. J Phys Chem Lett submitted

93. Arendt D (2003) Evolution of eyes and photoreceptor cell types. Int J Dev Biol 47:563–571

94. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature 453:363–367 . doi: 10.1038/nature06925

95. Rinaldi S, Melaccio F, Gozem S, et al (2014) Comparison of the isomerization mechanisms of human melanopsin and invertebrate and vertebrate rhodopsins. Proc Natl Acad Sci 111:1714 LP-1719

96. Archipowa N, Kutta RJ, Heyes DJ, Scrutton NS (2018) Stepwise Hydride Transfer in a Biological System: Insights into the Reaction Mechanism of the Light-Dependent Protochlorophyllide Oxidoreductase. Angew Chemie Int Ed 57:2682–2686 . doi: 10.1002/anie.201712729

97. Begley TP, Young H (1989) Protochlorophyllide reductase. 1. Determination of the regiochemistry and the stereochemistry of the reduction of protochlorophyllide to chlorophyllide. J Am Chem Soc 111:3095–3096 . doi: 10.1021/ja00190a071

98. Heyes DJ, Ruban A V, Wilks HM, Hunter CN (2002) Enzymology below 200 K: The kinetics and thermodynamics of the photochemistry catalyzed by protochlorophyllide oxidoreductase. Proc Natl Acad Sci 99:11145 LP-11150

99. Heyes DJ, Heathcote P, Rigby SEJ, et al (2006) The First Catalytic Step of the Light-driven Enzyme Protochlorophyllide Oxidoreductase Proceeds via a Charge Transfer Complex. J Biol Chem 281:26847–26853

100. Heyes DJ, Ruban A V, Hunter CN (2003) Protochlorophyllide Oxidoreductase:□ “Dark” Reactions of a Light-Driven Enzyme. Biochemistry 42:523–528 . doi: 10.1021/bi0268448

101. Heyes DJ, Hunter CN (2004) Identification and Characterization of the Product Release Steps within the Catalytic Cycle of Protochlorophyllide Oxidoreductase. Biochemistry 43:8265–8271 . doi: 10.1021/bi049576h

102. Townley HE, Sessions RB, Clarke AR, et al (2001) Protochlorophyllide oxidoreductase:

A homology model examined by site-directed mutagenesis. Proteins Struct Funct Genet 44:329–335 . doi: 10.1002/prot.1098

103. Sytina OA, Alexandre MT, Heyes DJ, et al (2011) Enzyme activation and catalysis: characterisation of the vibrational modes of substrate and product in protochlorophyllide oxidoreductase. Phys Chem Chem Phys 13:2307–2313 . doi: 10.1039/C0CP01686A

104. A Sytina O, Van Stokkum I, J Heyes D, et al (2010) Protochlorophyllide Excited-State Dynamics in Organic Solvents Studied by Time-Resolved Visible and Mid-Infrared Spectroscopy

105. Aquilante F, Pedersen TB, Lindh R (2007) Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. J Chem Phys 126:194106 . doi: 10.1063/1.2736701

106. Ke S-H, Baranger HU, Yang W (2007) Electron transport through single conjugated organic molecules: Basis set effects in ab initio calculations. J Chem Phys 127:144107 . doi: 10.1063/1.2770718

107. Rohrdanz MA, Herbert JM (2008) Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory. J Chem Phys 129:34107 . doi: 10.1063/1.2954017

108. Pepino AJ, Segarra-Marti J, Nenov A, et al (2018) UV-induced long-lived decays in solvated pyrimidine nucleosides resolved at the MS-CASPT2/MM level. Phys Chem Chem Phys 20:6877–6890 . doi: 10.1039/C7CP08235E

109. Pepino AJ, Segarra-Martí J, Nenov A, et al (2017) Resolving Ultrafast Photoinduced Deactivations in Water-Solvated Pyrimidine Nucleosides. J Phys Chem Lett 8:1777–1783 . doi: 10.1021/acs.jpclett.7b00316

110. Improta R, Santoro F, Blancafort L (2016) Quantum Mechanical Studies on the Photophysics and the Photochemistry of Nucleic Acids and Nucleobases. Chem Rev 116:3540–3593 . doi: 10.1021/acs.chemrev.5b00444

111. Kong W, He Y, Wu C (2008) Decay Pathways of Pyrimidine Bases: From Gas Phase to Solution BT - Radiation Induced Molecular Phenomena in Nucleic Acids: A Comprehensive Theoretical and Experimental Analysis. In: Shukla MK, Leszczynski J (eds). Springer Netherlands, Dordrecht, pp 301–321

112. Hare PM, Crespo-Hernández CE, Kohler B (2007) Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution. Proc Natl Acad Sci 104:435 LP-440

113. Hare PM, Middleton CT, Mertel KI, et al (2008) Time-resolved infrared spectroscopy of the lowest triplet state of thymine and thymidine. Chem Phys 347:383–392 . doi: https://doi.org/10.1016/j.chemphys.2007.10.035

114. Nenov A, Mukamel S, Garavelli M, Rivalta I (2015) Two-Dimensional Electronic Spectroscopy of Benzene, Phenol, and Their Dimer: An Efficient First-Principles Simulation Protocol. J Chem Theory Comput 11:3755–3771 . doi: 10.1021/acs.jctc.5b00443

115. Nenov A, Borrego-Varillas R, Oriana A, et al (2018) UV-Light-Induced Vibrational Coherences: The Key to Understand Kasha Rule Violation in trans-Azobenzene. J Phys Chem Lett 9:1534–1541 . doi: 10.1021/acs.jpclett.8b00152

116. C. Tully J (1998) Mixed quantum-classical dynamics. Faraday Discuss 110:407–419 . doi: 10.1039/A801824C

117. Hammes-Schiffer S, Tully JC (1994) Proton transfer in solution: Molecular dynamics with quantum transitions. J Chem Phys 101:4657–4667 . doi: 10.1063/1.467455

118. Serrano-Andrés L, Merchán M, Lindh R (2005) Computation of conical intersections by using perturbation techniques. J Chem Phys 122:104107 . doi: 10.1063/1.1866096