

# Photochemical reactions in biological systems: probing the effect of the environment by means of hybrid quantum chemistry/molecular mechanics simulations<sup>†</sup>

**Martial Boggio-Pasqua,<sup>c</sup>, Carl F. Burmeister,<sup>a</sup>, Michael A. Robb,<sup>b</sup> and Gerrit Groenhof<sup>\*a</sup>**

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 200X

DOI: 10.1039/b000000x

Organisms have evolved a wide variety of mechanisms to utilize and respond to light. In many cases, the biological response is mediated by structural changes that follow photon absorption in the protein complex. The initial step in such cases is normally the photoisomerization of a highly conjugated prosthetic group. To understand better the factors controlling the isomerization, we perform atomistic molecular dynamics simulations. In this perspective article we briefly review the key theoretical concepts of photochemical reactions and present a practical simulation scheme for simulating photochemical reactions in biomolecular systems. In our scheme, a multi-configurational quantum mechanical description is used to model the electronic rearrangement for those parts of the system that are involved in the photon absorption. For the remainder, typically consisting of the apo-protein and the solvent, a simple force field model is used. The interactions in the systems are thus computed within a hybrid quantum/classical framework. Forces are calculated on-the-fly, and a diabatic surface hopping procedure is used to model the excited-state decay. To demonstrate how this method is used we review our studies on photoactivation of the photoactive yellow protein, a bacterial photoreceptor. We will show what information can be obtained from the simulations, and, by comparing to recent experimental findings, what the limitations of our simulations are.

## Introduction

Photobiological processes, such as vision or photosynthesis, in which sunlight is used as the energy source to bring about chemical reactions, provide valuable templates to create tools for nanotechnology, biomolecular imaging, information technology and renewable energy. However, before one can mimic such processes, one needs a complete understanding of the underlying molecular dynamics. As the relevant time and spatial resolution are notoriously hard to access experimentally it is difficult to get detailed information about the mechanism by experiment alone. Computer simulations, on the other hand, provide such information in atomic detail, and can thus complement experiments in unraveling the details of photochemical processes in biological systems.

In this perspective we will show how computer simulations may be used to shed light on photochemical reactions in complex environments. As an illustration, we will discuss our work on the photoactive yellow protein (PYP), a bacterial photoreceptor, which is believed to be responsible for negative phototactic response to blue light of *Halorhodospira halophila* bacteria. Since the main goal of our work is to explore the effect of the protein environment on the excited-state dynamics of the embedded chromophore, we investigate the photochemical properties of the chromophore in different environments by means of hybrid QM/MM simulations. In these simulations the chromophore is described at the ab initio level (QM), while the environment is modeled by a molecular mechanics force field (MM).

By comparing the chromophore in isolation, in solution and in the protein, the effect of the different environments can be revealed. Each level in this hierarchy has its advantages and disadvantages. In vacuum, the level of ab initio theory can be increased systematically, but experimental data are difficult to obtain. In solution, reliable experimental data on lifetimes are often available, but the effect of the solvent is more difficult to model. At the protein level, experimental data is often more difficult to interpret, while at the same time the complexity of these systems poses the most serious challenges to the

<sup>†</sup> Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

<sup>a</sup> Department of Theoretical and Computational Biophysics, Max Planck Institute for biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. Fax: +495512012302; Tel. +495512012321; E-mail: [ggroenh@gwdg.de](mailto:ggroenh@gwdg.de)

<sup>b</sup> Laboratoire de Chimie et Physique Quantiques - IRSAMC, CNRS et Université de Toulouse, 31062 Toulouse, France.

<sup>c</sup> Chemistry department, Imperial College, London SW7 2AZ, United Kingdom.

theory. Therefore, it is not always easy to validate the simulations against experiment. Nevertheless, the atomistic insights available through simulations may stimulate new experiments that will ultimately lead to a better understanding as well as improvements on the theory.

Before we start our presentation of simulating photochemical processes in photobiological systems and in PYP in particular, we want to point out that the idea of combining quantum chemistry with molecular mechanics to perform molecular dynamics simulations of photo-induced processes is not new, but dates back more than three decades, when Warshel used this approach for the first time to simulate the photoisomerization of retinal in rhodopsin<sup>1</sup>. Since then, he and co-workers have systematically refined their description, and obtained more detailed information about the influence of the protein environment on retinal photoisomerization in rhodopsin<sup>2</sup> and bacteriorhodopsin<sup>3–5</sup>. Since these pioneering studies, other researchers, including ourselves, have applied the method to investigate the excited-state dynamics in different biological systems. Garavelli and co-workers have applied QM/MM in conjunction with diabatic surface hopping to uncover the details of retinal isomerization in bovine rhodopsin<sup>6</sup>. The photoisomerization step of this protein had been addressed before by QM/MM dynamics simulations by the Olivucci<sup>7</sup> and Schulten groups.<sup>8</sup> Schulten and co-workers have also studied the retinal isomerization in bacteriorhodopsin<sup>9</sup>. Morokuma and co-workers have used the QM/MM method to uncover the photoswitching in Dronpa<sup>10</sup>, a reversible switchable fluorescent protein<sup>11</sup>. In addition to photoisomerization, also light-induced electron and proton transfer reactions in biological systems have been studied by means of molecular dynamics, as demonstrated by the works of Warshel and co-workers on photochemical charge separation in photosynthetic reactioncenters<sup>5,12–14</sup>, or the excited state proton transfer event in DNA<sup>15</sup>. Finally, radiationless decay processes of DNA bases in DNA have been studied by Thiel and co-workers,<sup>16</sup> and Lischka and co-workers.<sup>17</sup> These applications show that there is a broad interest in applying hybrid QM/MM molecular dynamics techniques to understand photochemical processes in condensed phase systems.

This article is organized as follows: we first provide a concise introduction into the simulation methodology that we use to model condensed phase photochemistry. Then, we demonstrate how we applied such methods to reveal the initial response of PYP to photon absorption and how the protein environment controls the dynamics in the excited state. We will compare the outcome of the simulations to experimental data and use these comparisons to discuss the limitations of the method. We end with an outlook on where the theory can be improved and what can be expected from the simulations in the future.

## Modeling excited-state dynamics in biological systems

The size and complexity of a typical photobiological system, together with the timescales that must be reached, necessitate the use of classical molecular dynamics (MD) for the nuclear degrees of freedom. In MD simulations Newton's equations of motion are solved numerically to obtain a trajectory of the dynamics of a molecule over a period of time. To model the electronic rearrangement upon excitation a quantum mechanical description (QM) is required for the parts of the system that are involved in photon absorption, usually a chromophore. For the remainder, typically a hydrated protein, a simple molecular mechanics force field model (MM) suffices. The interactions in the systems are thus computed within a hybrid QM/MM framework<sup>18</sup>.

### Conical intersections

The Born-Oppenheimer approximation forms the basis for modern quantum chemistry calculations. Based on the small electron/nucleus mass ratio of about one to two thousand, it is assumed that the motions of electrons and nuclei are uncoupled and nuclear dynamics is restricted to a single potential energy surface corresponding to an adiabatic electronic state. This assumption is valid as long as the separation between the electronic energy levels is large compared to the separation between the vibrational energy levels. For photochemical reactions, however, this is usually not true.

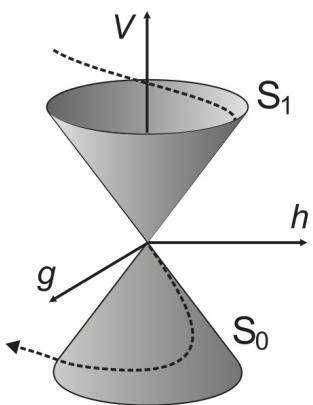
During a photochemical reaction the system samples regions of the configuration space where the energy gaps between electronic states are of the same magnitude as the energy gaps between the vibrational states of the nuclei. Under such conditions resonance occurs between vibrational and electronic transitions. Thus, nuclear and electronic degrees of freedom are strongly coupled in these regions and the approximation that the nuclear wave function is restricted to a single adiabatic electronic potential energy surface breaks down. Instead, nuclear dynamics induces population transfers between adiabatic electronic states that are close in energy. If the coupling is strong enough, the adiabatic potential energy surfaces can even intersect. These surface crossings provide efficient funnels for radiationless deactivation of the excited state and therefore play a crucial role in photochemistry.

The degeneracy between two potential energy surfaces, *e.g.* the ones of the first singlet excited state ( $S_1$ ) and the ground state ( $S_0$ ), is lifted in the two dimensional space spanned by the gradient difference vector

$$\mathbf{g} = \nabla_{\mathbf{R}} (E_{S_1} - E_{S_0}), \quad (1)$$

and interstate coupling vector

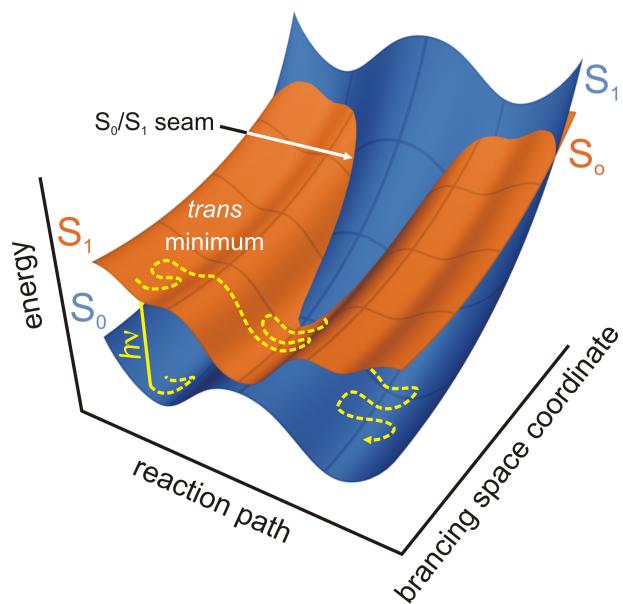
$$\mathbf{h} = \nabla_{\mathbf{R}} \langle \psi_{S_0} | \hat{H} | \psi_{S_1} \rangle, \quad (2)$$



**Fig. 1** A conical intersection plotted in the 2-dimensional branching space that is spanned by the gradient difference vector ( $g$ ) and the derivative coupling vector ( $h$ ). Motion away from the intersection in the  $g-h$  plane lifts the degeneracy between the two electronic states  $S_1$  and  $S_0$ . The dotted line shows a path of a nuclear trajectory passing from one electronic state ( $S_1$ ) to another ( $S_0$ ) through the intersection funnel.

where  $\mathbf{R}$  is the  $3N$ -dimensional vector of the positions of the nuclei in the system,  $\hat{H}$  the electronic Hamiltonian,  $E_{S_1}$  ( $E_{S_0}$ ) and  $\psi_{S_1}$  ( $\psi_{S_0}$ ) the energy and electronic wave function of the excited (ground) state, respectively. The space spanned by these two vectors is referred to as the branching space, or  $g-h$  plane. When projected onto this plane, the intersecting surfaces exhibit the typical double cone topology, with the point of degeneracy at the apex (Figure 1). Orthogonal to the two-dimensional branching space exists the so-called intersection space (or seam space), in which the energies of the two states remain degenerate to first order. In a molecule with  $M$  internal degrees of freedom, the intersection space thus forms an  $(M-2)$  dimensional seam, each point of which is a conical intersection (Figure 2).

The conical intersection seam is the central mechanistic feature in a photochemical reaction. The conical intersection provides a funnel for efficient radiationless decay between electronic states (Figure 2). To illustrate the relationship between a surface crossing and photochemical reactivity, we draw a parallel with the transition state in ground-state chemistry. The transition state forms the dynamical bottleneck through which the reaction must pass on its way from reactants to products. A transition state separates the reactant and product energy minima along the reaction path. A conical intersection provides a locus of points that separate the excited-state branch of the reaction path from the ground-state branch. The crucial difference between a conical intersection and a transition state is that, while the transition state must connect the reactant minimum to a single product minimum via a single reaction path, an intersection is a singularity on the ground-state energy sur-



**Fig. 2** Schematic overview of a photochemical reaction pathway (dashed line). After photon absorption, evolution takes place on the excited-state potential energy surface (red) until the system hits the  $S_1/S_0$  intersection seam. At the intersection, a radiationless transition to the ground state occurs (blue). After the decay, the system continues evolving in the ground state. Adapted from Groenhof *et al.*<sup>19</sup>

face and thus connects the excited-state reactants to one, two or even more products on the ground state via several reaction paths. More detailed information about conical intersections and the role they play in photochemistry can be found in recent perspectives and reviews<sup>20-23</sup>.

### Excited-state molecular dynamics

The desire to understand photochemical processes has sparked the development of a variety of methods for treating non-adiabatic effects in classical molecular dynamics simulations. Most of these methods are based on surface hopping: nuclei move on a single potential energy surface and non-adiabatic transitions are included by allowing the trajectory to hop from one surface to another<sup>1,24-27</sup>. Notable exceptions are: (i) the full-multiple spawning method by Martínez and co-workers<sup>28</sup> that exploits the benefits of a *local* description of the adiabatic potential energy surface on the one hand, and a *global* description of the nuclear wavefunction on the other; (ii) the density matrix evolution approach by Parson and Warshel<sup>29</sup> that is based on integrating a stochastic Liouville equation; (iii) the mean-field surface hopping approach by Prezhdo and Rossky, which combines surface hopping with Ehrenfest dynamics in regions of non-adiabatic coupling,<sup>30</sup> and (iv) the station-

ary phase surface hopping method, by Webster *et al.*<sup>31,32</sup>, in which the time-evolution of the quantum (electronic) and classical (nuclear) degrees of freedom are iterated until self-consistency during a characteristic coherence interval.<sup>33</sup>

The most popular implementation of the surface hopping method is Tully's fewest-switches algorithm (FSA)<sup>25,34</sup>. The FSA is based on the minimization of the number of transitions needed to maintain the self-consistency between the quantum and the classical populations. In its practical implementation the FSA requires knowledge of the non-adiabatic couplings and the coherent integration of quantum amplitudes along the entire trajectory. Because these requirements cannot be easily fulfilled in applications on large systems, the FSA is rarely applied in these studies. Instead, we describe the relationship between the magnitude of the non-adiabatic coupling and the hopping probability by a simpler model.

Our hopping algorithm is based on the one-dimensional Landau-Zener equation, which relates the probability of a transition between two electronic states  $\psi_2$  and  $\psi_1$  to the non-adiabatic coupling, via

$$P_{2 \rightarrow 1} = \exp\left(-\frac{1}{4}\pi\xi\right) \quad (3)$$

In this equation  $\xi$  is the Massey parameter, defined as

$$\xi = \frac{\Delta E}{\hbar \frac{\partial \mathbf{Q}}{\partial t} \cdot \mathbf{g}(\mathbf{Q})}, \quad (4)$$

where  $\Delta E$  is the energy gap between the adiabatic states,  $\mathbf{Q}$  represents a one dimensional nuclear reaction coordinate, and

$$\mathbf{g}(\mathbf{Q}) = \langle \psi_1 | \nabla_{\mathbf{Q}} \psi_2 \rangle \quad (5)$$

is the derivative coupling vector. If we differentiate  $\psi_2$  with respect to  $t$  via  $\frac{\partial}{\partial t} \langle \psi_1 | \frac{\partial \psi_2}{\partial t} \rangle$ , we can rewrite  $\xi$  as

$$\xi = \frac{\Delta E}{\hbar \langle \psi_1 | \frac{\partial \psi_2}{\partial t} \rangle}. \quad (6)$$

To decide when to undergo a transition to a different potential energy surface, one would in principle need to compute  $\langle \psi_1 | \frac{\partial}{\partial t} \psi_2 \rangle$  at every time step ( $\Delta t$ ) of the simulation. In practice, however, it is possible to approximate the derivative coupling vector  $\langle \psi_1 | \frac{\partial}{\partial t} \psi_2 \rangle$  as  $\langle \psi_1(t) | \psi_2(t + \Delta t) \rangle$ , i.e., the overlap between the excited-state wave function at the current time step and the ground-state wave function at the previous time step. Since in our simulations we use the complete active space self-consistent field (CASSCF) method<sup>35</sup> with state-averaged (SA) orbitals to model the wave function, we compute the overlap as the inner product of the corresponding SA-CASSCF eigenvectors  $\mathbf{C}_1$  and  $\mathbf{C}_2$ :

$$\langle \psi_1(t) | \psi_2(t + \Delta t) \rangle = \mathbf{C}_1^t \cdot \mathbf{C}_2^{t+\Delta t} \quad (7)$$

Calculating the energy gap  $\Delta E$  and  $\mathbf{C}_1^t \cdot \mathbf{C}_2^{t+\Delta t}$  at every time step is straightforward, and we can use the Landau-Zener formula to calculate the probability of a transition to the other surface. In principle, the transition probability can be used to spawn a new trajectory on the other surface. However, since this procedure would lead to multiple trajectories that have to be computed simultaneously, spawning is too demanding in practice. We therefore restrict hopping to situations where the transition probability approaches unity. This happens at the conical intersection seam, where  $\Delta E \sim 0$  and  $\mathbf{C}_1^t \cdot \mathbf{C}_2^{t+\Delta t} \sim 1$ .

Because we allow hopping only at the conical intersection seam, our classical trajectories never leave the diabatic surface. Therefore, energy and momentum are obviously conserved. In principle, this strict diabatic hopping criterion could lead to an underestimation of the population transfer probability, because a surface hop in regions with strong non-adiabatic coupling far from the intersection is prohibited. In practice, however, the high dimensionality of the seam ensures that all trajectories encounter such regions of high transfer probability. A major advantage of restricting hopping to the seam is that we obtain information on the location of the seam in our trajectories. The latter is important to understand how the interactions between a chromophore and its (protein) environment alter the topology of the surfaces and the seam and thereby control the outcome of the photochemical process. The Landau-Zener model is clearly an approximation, but can help to keep a proper physical insight, which is crucial for understanding complex systems.

## Excited-state wave functions

Although highly accurate methods for computing excited-state electronic wave functions have become available over the past years, they are usually too time-consuming for systems larger than a few atoms. Therefore, most of these methods are not yet applicable in on-the-fly molecular dynamics simulations of large biomolecular systems. Alternatively, simple force fields or existing semi-empirical methods that are computationally efficient may be used, but their applicability is limited, unless properly re-parameterized<sup>1,36–38</sup>. Therefore, for on-the-fly molecular dynamics, a compromise between cost and accuracy has to be made.

Computationally feasible approaches to describe excited-state electronic structures are time-dependent density functional theory (TD-DFT) and equation-of-motion coupled cluster (EOM-CCSD), and have been used in excited-state molecular dynamics simulations<sup>39,40</sup>. Both approaches however suffer from deficiencies of the underlying monoconfigurational DFT or CCSD description of the ground state in regions of bond breaking and bond formation. Moreover, TD-DFT is known to encounter severe problems in describing valence states of molecules exhibiting extended  $\pi$  systems, doubly ex-

cited states, charge-transfer excited states<sup>41</sup>, and conical intersections between ground and excited states<sup>42,43</sup>.

The problems associated with methods that are based on a single reference configuration, such as EOM-CCSD and TD-DFT, demonstrate that for computing excited states, multi-configurational methods are required to provide wave functions that are sufficiently flexible to describe bond rearrangements, electronic state mixing, and electronic reorganizations. In addition, to calculate molecular dynamics trajectories analytical energy gradients are necessary. Since the CASSCF method fulfills these requirements<sup>35</sup>, it has often been used in the framework of excited-state dynamics simulations<sup>44–48</sup>.

In CASSCF, a judicious set of occupied and virtual orbitals is chosen, the so-called active-space orbitals<sup>35</sup>. In this active space, a full configuration interaction calculation is performed, while the other orbitals are being kept doubly occupied or empty in all configurations. The occupied orbitals are optimized such that the electronic energy of the state considered is minimal. Alternatively, the average energy of the states under study is minimized (state-average (SA) approach), if state bias or root-flipping has to be avoided<sup>49</sup> that occur near surface crossing regions<sup>50</sup>.

The CASSCF method captures to a large extent so-called static electron correlation. However, due to the necessary truncation of the active space, it does not recover dynamic electron correlation completely. Dynamic correlation is known to play a key role in the quantitative description of barrier heights and excitation energies. Thus, a higher-level treatment that includes dynamic electron correlation effects is desirable. Unfortunately, methods that resolve both static and dynamic correlations tend to be computationally too demanding and furthermore often lack the required analytical energy gradients.

Even at the CASSCF level of theory, the calculation of energies and gradients at every step of the simulation places a severe demand on computational resources. We are therefore forced to use reduced active spaces. These rather small active spaces (typically 6 electrons distributed over 6 orbitals) have to be calibrated against higher-level methods before the simulations can be performed. Validation is usually done by comparing the energies of stationary points on the CASSCF potential energy surfaces to the single-point CASPT2 (complete active space with second order perturbation theory)<sup>51</sup> energies at these geometries.

### Hybrid quantum chemistry/molecular mechanics simulations

Molecular dynamics computer simulations of biological systems have come of age. Since the first application of MD on a photobiological process more than three decades ago<sup>1</sup>, advances in computer power, algorithmic developments and improvements in the accuracy of the used interaction functions

have established MD as an important and predictive technique to study dynamic processes at atomic resolution<sup>52</sup>.

In the interaction functions, the so-called molecular mechanics (MM) force field, simple chemical concepts are used to describe the potential energy of the system<sup>53</sup>:

$$V_{\text{MM}} = \sum_i^{N_{\text{bonds}}} V_i^{\text{bond}} + \sum_j^{N_{\text{angles}}} V_j^{\text{angle}} + \sum_l^{N_{\text{torsions}}} V_l^{\text{torsion}} \\ \sum_i^{N_{\text{MM}}} \sum_{j>i}^{N_{\text{MM}}} V_{ij}^{\text{Coul}} + \sum_i^{N_{\text{MM}}} \sum_{j>i}^{N_{\text{MM}}} V_{ij}^{\text{LJ}}, \quad (8)$$

where  $N_{\text{MM}}$  is the number of atoms in the system. Bonds and angles ( $V^{\text{bond}}$ ,  $V^{\text{angle}}$ ) are normally modeled by harmonic functions, and torsions by periodic functions ( $V^{\text{torsion}}$ ). The pairwise electrostatic interaction between atoms with a partial charge ( $Q_i$ ) is given by Coulomb's law:

$$V_{ij}^{\text{Coul}} = \frac{e^2 Q_i Q_j}{4\pi\epsilon_0 R_{ij}}, \quad (9)$$

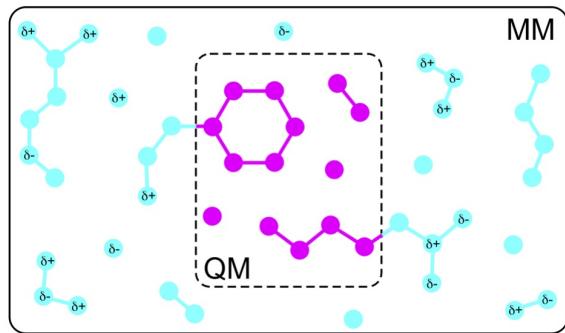
in which  $R_{ij}$  denotes the interatomic distance,  $e$  the elementary charge, and  $\epsilon_0$  the vacuum di-electric constant. Short-range Pauli repulsion and long-range dispersion attraction are most often described by a single Lennard-Jones potential:

$$V_{ij}^{\text{LJ}} = \left( \frac{C_{12}^{ij}}{R_{ij}} \right)^{12} - \left( \frac{C_6^{ij}}{R_{ij}} \right)^6, \quad (10)$$

with  $C_{12}^{ij}$  and  $C_6^{ij}$  a repulsion and attraction parameter respectively, that depend on the atom types involved. In the above expression for the MM energy, there are no terms associated with (molecular) polarizability. Instead, the effect of induced dipole - charge and induced dipole - induced dipole interactions are accounted for implicitly via the partial charges and Lennard-Jones parameters. Although different approaches to include the polarizability have been introduced very early on, such as the Protein-Dipoles-Langevin-Dipoles (PDLD) model<sup>18</sup>, the shell<sup>54</sup> and charge-on-a-spring model<sup>55</sup>, or charge equilibration scheme<sup>56</sup>, these methods have not yet found widespread use in the most popular force fields of today (e.g., Gromos<sup>57</sup>, Amber<sup>58</sup>, OPLS<sup>59</sup> or Charmm<sup>60</sup>).

Electrons are thus ignored in molecular mechanics force fields. Their influence is expressed by empirical parameters that are valid for the ground state of a given covalent structure. Therefore, processes that involve electronic rearrangements, such as photochemical reactions, cannot be described at the MM level. Instead, these processes require a quantum mechanical description. As we have discussed above, the computational effort associated with computing excited-state electronic structure puts severe constraints on the size of the system that can be studied. To overcome this limitation for biological systems, which are typically orders of magnitude too large for a complete quantum chemical treatment, methods have been developed that treat a small part of the system

at an appropriate quantum mechanical level (QM), while retaining the computationally cheaper force field (MM) for the remainder. This hybrid QM/MM strategy was introduced by Warshel and Levitt<sup>18</sup> and is illustrated in Figure 3.



**Fig. 3** Illustration of the QM/MM concept. A small region, in which a chemical reaction occurs and therefore cannot be described with a force field, is treated at a sufficiently high level of QM theory. The remainder of the system is modelled at the MM level.

The justification for dividing a system into regions that are described at different levels of theory is the local character of chemical reactions in condensed phases. A distinction can usually be made between a *reaction center* with atoms that are directly involved in the reaction and a *spectator region*, in which the atoms do not directly participate in the reaction. For example, most reactions in solution involve the reactants and the first few solvation shells. The bulk solvent is hardly affected by the reaction, but can influence the reaction via long-range interactions. The same is true for most enzymes, in which the catalytic process is restricted to an active site. The rest of the protein provides an electrostatic background that could facilitate the reaction.

The hybrid QM/MM Hamiltonian contains three classes of interactions: interactions between atoms in the QM region, interactions between atoms in the MM region and interactions between QM and MM atoms

$$H^{\text{hybrid}} = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM/MM}}. \quad (11)$$

The interactions within the QM and MM regions are relatively straightforward to describe, *i.e.*, at the QM and MM level, respectively. The interactions between the two subsystems are more difficult to describe and several approaches have been proposed.

In the most simple approach, the QM subsystem is mechanically embedded in the MM system and kept in place by force field interactions, *i.e.*, bonds, angles, torsions and Lennard-Jones (equation 8). With the exception of these interactions, the two systems are treated independently. Thus a quantum

chemistry calculation is performed on an isolated QM subsystem, while a force field calculation is performed on the MM region. An improvement of the model is to use the isolated electronic wave function to derive partial atomic charges for the QM atoms<sup>61</sup> and use these charges to compute the electrostatic QM/MM interactions with the MM atoms.

In the more popular electronic embedding scheme the (non-polarizable) MM atoms enter the electronic Hamiltonian, as if they were QM nuclei<sup>18,62</sup>:

$$\begin{aligned} H_{\text{QM/MM}}^e &= H^e + \sum_i^{n_e} \sum_K^{N_{\text{MM}}} \frac{e^2 Q_K}{4\pi\epsilon_0 r_{ik}} \\ &= -\frac{\hbar^2}{2m_e} \sum_i^{n_e} \nabla_i^2 + \sum_i^{n_e} \sum_{j>i}^{n_e} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \\ &\quad - \sum_i^{n_e} \sum_A^{N_{\text{QM}}} \frac{e^2 Z_A}{4\pi\epsilon_0 r_{iA}} + \sum_K^{N_{\text{MM}}} \frac{e^2 Q_K}{4\pi\epsilon_0 r_{ik}}, \end{aligned} \quad (12)$$

where  $H^e$  is the original electronic Hamiltonian for the isolated QM system;  $n_e$  is the number of electrons,  $N_{\text{QM}}$  the number of QM nuclei and  $N_{\text{MM}}$  the number of MM atoms;  $Z_A$  and  $Q_K$  are the nuclear and partial charges of QM nucleus  $A$  and MM atom  $K$ , respectively; and  $m_e$  denotes the electron mass. Because the MM atoms enter the Hamiltonian, the electronic wavefunction is polarized by the environment. Simultaneously, the electrons are exerting electrostatic forces on both QM nuclei and MM atoms. Problems may arise if the MM atoms near the QM region have high partial charges. In this case, the electrons are strongly attracted by such MM atoms, and the wave function can become over-polarized. Penetration of electron density into the MM region is an artifact of ignoring the electrons of the MM atoms. A remedy for this spill-out effect is to use gaussian-shaped charge densities rather than point charges to represent partially charged MM atoms<sup>63</sup>. A promising alternative is the constrained or frozen density functional theory method<sup>64–66</sup>, which takes into account the interaction with the (frozen) electron density of the environment explicitly, and thus avoids the problem altogether.

Interactions between the nuclei in the QM region, and between QM nuclei and MM atoms are described by the Coulomb potential:

$$H_{\text{QM/MM}}^{\text{nuc}} = \sum_A^{N_{\text{QM}}} \sum_{B>A}^{N_{\text{QM}}} \frac{Z_A Z_B}{4\pi\epsilon_0 R_{AB}} + \sum_A^{N_{\text{QM}}} \sum_K^{N_{\text{MM}}} \frac{Z_A Q_K}{4\pi\epsilon_0 R_{AK}}. \quad (13)$$

In addition to electrostatics, there are also Van der Waals interactions between the subsystems that are handled at the force field level, as if the QM nuclei were MM atoms. Similar to the previous model, bonded interactions, such as bonds, angles and torsions involving both QM and MM atoms, are described by the respective force field functions.

If the QM and MM subsystems are connected by chemical bonds, care has to be taken when evaluating the QM wave

function. Cutting the QM/MM bond creates one or more unpaired electrons in the QM subsystem. In reality, these electrons are paired in a bonding orbital with electrons belonging to the atoms on the MM side. A number of approaches to remedy the artifact of such open valences have been proposed.

The simplest solution is to introduce a monovalent link atom at an appropriate position along the bond vector between the QM and MM atoms.<sup>62</sup> Hydrogen is most often used, but there is no restriction on the type of the link atom and even complete fragments, such as methyl groups, can be used to cap the QM subsystem. The link atoms are present only in the QM calculation, and are invisible for the MM atoms. In our simulations we use this approach to cap the QM region.

A popular alternative to the link atom scheme is to replace a chemical bond between the QM and MM subsystem by a double occupied molecular orbital. This idea, which dates back to the pioneering work of Warshel and Levitt<sup>18</sup>, assumes that the electronic structure of the bond is insensitive to changes in the QM region. The two most widely used approaches are the localised hybrid orbital<sup>67</sup> or frozen orbital<sup>68</sup> method on the one hand, which introduces orbitals at the QM atom, and the generalised hybrid orbital approach<sup>69</sup> on the other, which places additional orbitals on the MM atom. In all localised orbital approaches, one or more parametrization steps are required. For this reason, the link atom is still the most widely adopted procedure for capping the QM region. Furthermore, studies that compared the accuracy of both methods showed that there is little advantage in using localised orbitals rather than link atoms<sup>70,71</sup>.

The electrostatic embedding QM/MM scheme provides a conceptually intuitive way of including the effect of an environment on a chemical reaction. However, unless polarization is treated explicitly at the force field level, this QM/MM model is not internally consistent. In most force fields polarization is not yet explicitly included, but is implicitly accounted for via the parameters in the other terms, most notably, in the non-bonded interactions. Thus, although the MM region can induce polarization of the QM subsystem, the latter cannot back-polarize the MM region. A related problem arises from the use of standard Lennard-Jones and charge parameters, which implicitly contain polarization. When using these parameters without modification, there is both explicit and implicit polarization of the QM region. Therefore, the total polarization can be overestimated.

To avoid possible artifacts, one has to account for the polarizability of the environment, which, as a matter of fact, was already achieved in the first QM/MM study of Warshel and Levitt by the PDLD model<sup>18</sup>. However, to obtain the total QM/MM energy in the polarizable embedding approach, the MM polarization need to be computed at every step of the self-consistent-field iteration of the QM wave function. Since the polarization is computed in a self-consistent manner as well,

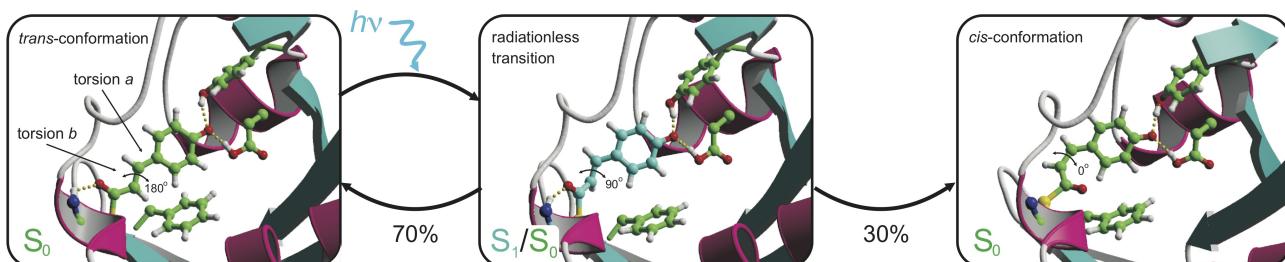
the QM/MM computation can become very cumbersome and demanding. Therefore, only few QM/MM studies on excited states have used a polarizable force field, notably the work of Warshel and Chu on photoisomerization in bR<sup>7</sup> and the calculations of excitation energies of chromophores in different solvent environments by Kongstedt and co-workers<sup>72-74</sup>, including PYP<sup>75</sup>. Alternatively, the non-bonded parameters could be reparametrized for use in QM/MM simulations. However, this procedure requires the optimization of very many parameters that also depend on the level of theory employed for the QM subsystem, and is most often skipped in practice. In our work, we rely on the default force field parameters to describe the QM/MM interactions.

## Photoisomerization of the photoactive yellow protein chromophore

To give the reader a taste of what can be achieved with QM/MM excited-state molecular dynamics simulations, we give a brief overview of our work on photoisomerization in photoactive yellow protein (PYP). Our choice for this system is motivated by the fact that the photochemistry of this protein and its chromophore have been extensively studied by both experiment and computation, so that there is a wealth of data to compare the outcome of our simulations to. We will show that there is agreement, but also discrepancy. We end with a discussion on the possible reasons for the discrepancies, and suggestions on how to overcome these and improve the predictive power of excited-state molecular dynamics simulations in general.

PYP is believed to be the primary photoreceptor for the photoavoidance of the salt-tolerant bacterium *Halorhodospira halophila*. As shown in Figure 4, PYP contains a deprotonated 4-hydroxy-cinnamic acid (or *p*-coumaric acid, pca, figure 5) chromophore that is covalently linked to the  $\gamma$ -sulphur atom of Cys69 via a thioester bond. Upon absorbing a blue-light photon, PYP enters a fully reversible photocycle involving several intermediates on timescales ranging from a few hundred femtoseconds to seconds<sup>76</sup>.

To understand how the protein mediates the photoisomerization of the chromophore, we have performed atomistic simulations of the chromophore in various molecular environments. The results of these computations not only provided a detailed mechanistic picture of the isomerization process, but also revealed the influence of the interactions with the environment. With such detailed information new experiments can be designed, which will ultimately enhance our understanding of PYP's photochemistry.

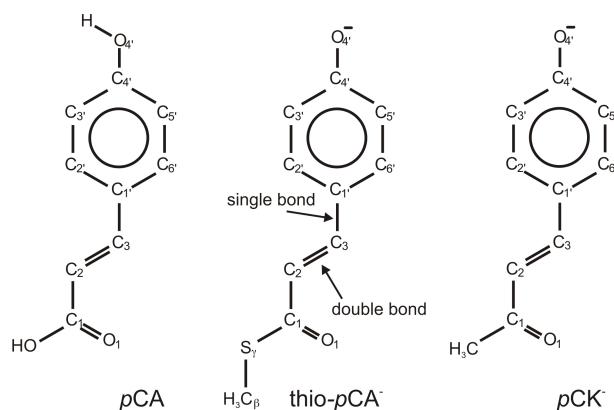


**Fig. 4** Snapshots from excited-state trajectories of wild-type PYP, showing the chromophore ( $\text{thio-pCA}^-$ ) in the active site pocket. The first snapshot is at the excitation. The second shows the configuration at the radiationless transition from  $S_1$  to  $S_0$ . The third snapshot shows the photoproduct, in which the carbonyl oxygen of the thioester linkage has flipped and is no longer hydrogen bonded to the backbone of Cys69. Adapted from Groenhof *et al.*<sup>19</sup>

### Isolated Chromophore

To investigate the intrinsic excited-state dynamics of the deprotonated chromophore, we performed molecular dynamics of two isolated chromophore analogues (thiomethyl *para*-coumaric acid ( $\text{thio-pCA}^-$ ) and *para*-coumaric-ketone ( $\text{pCK}^-$ , Figure 5). Because the chromophore is deprotonated in the protein, we focus our attention on the anionic species. The isolated neutral chromophore, which has been investigated by experiment<sup>77–79</sup> and computation<sup>80,81</sup> is therefore not considered here. In our simulations, the chromophore was described at the CASSCF level, with six electrons distributed in six molecular orbitals, which were expanded in the 3-21G basis set. We remark that in isolation the excited states of the anionic chromophores are autoionizing, *i.e.*, unstable with respect to electron detachment<sup>82</sup>. Furthermore, spectroscopic measurements on several anionic isolated chromophore derivatives have demonstrated that in vacuum, the major excited state decay channels are photodetachment and fragmentation, rather than isomerization.<sup>83–86</sup>. Because we do not include diffuse functions in our calculations (Gromov *et al.* showed that the effect on the vertical excitation energy was found to be less than 0.2 eV<sup>82</sup>), we are not able to model the autoionization and electron detachment processes. Instead, our aim is to understand the effect of different molecular environments on the photoisomerization processes, for which we need the hypothetical chromophore model in vacuum as a reference. For a theoretical study on the electron detachment and dissociation channels in isolated *p*-coumaric acid analogues, we refer the reader to recent work by Krylov and co-workers.<sup>87</sup>

After vertical excitation to  $S_1$ , both chromophores rapidly relax from the Franck-Condon region by twisting around the formal single bond adjacent to the chromophore ring (Figure 5). After reaching the twisted-single bond minimum, however, no decay to the ground state occurred within 5 ps, which was the total time of the simulations. In contrast, Martinez and co-workers observed radiationless decays in their full multi-



**Fig. 5** Schematic drawings of the *p*-coumaric chromophore and analogues used in our simulations.

ple spawning simulations of the  $\text{pCK}^-$  chromophore<sup>88</sup>. Most likely, we underestimate the decay probability too much in our vacuum simulations by restricting hopping to the seam<sup>89</sup>. Nevertheless, the observation of a fast vibrational relaxation into the single-bond twisted  $S_1$  minimum for both  $\text{thio-pCA}^-$  and  $\text{pCK}^-$  is consistent with the topology of the excited-state potential energy surface for these chromophores, as shown below and by others<sup>82,88,90</sup>.

Because no excited-state decay was observed in the molecular dynamics simulations, we resorted to geometry optimizations at the CASSCF/6-31G\* level of theory. In these optimizations, the active space consisted of 12 electrons in 11  $\pi$ -orbitals. In addition to optimizing the local minima on the  $S_0$  and  $S_1$  potential energy surfaces and the barriers that separate them, we also searched for minimum energy  $S_1/S_0$  conical intersections in the vicinity of these minima (Table 1).

The optimizations revealed (Table 1) that in both chromophores there are two minima on  $S_1$ , associated with bond twisting: the single-bond twisted minimum, in which the bond adjacent to the phenol ring is rotated by 90°; and the double-bond twisted minimum, in which the ethylenic bond is twisted

**Table 1** Relative energies (kJ/mol) of critical structures for thio-pCA<sup>-</sup> and pCK<sup>-</sup> with respect to the corresponding ground-state minimum at CASSCF(12,11)/6-31G\* level. Data for pCK<sup>-</sup> taken from Boggio-Pasqua *et al.*<sup>91</sup>.

Structures <sup>a</sup>	thio-pCA <sup>-</sup>		pCK <sup>-</sup>	
	S <sub>0</sub> ΔE <sup>b</sup>	S <sub>1</sub> ΔE <sup>c</sup>	S <sub>0</sub> ΔE <sup>d</sup>	S <sub>1</sub> ΔE <sup>e</sup>
S <sub>0</sub> GS	0.0	328.2	0.0	336.6
S <sub>1</sub> planar	14.6	314.3	16.6	318.9
S <sub>1</sub> DB twist	174.5	294.6	195.0	286.3
S <sub>1</sub> SB twist	132.2	278.0	122.0	293.2
S <sub>1</sub> /S <sub>0</sub> DB twist CI	343.8	344.5	314.7	315.6
S <sub>1</sub> /S <sub>0</sub> hula-twist CI	321.3	321.6	333.5	333.8

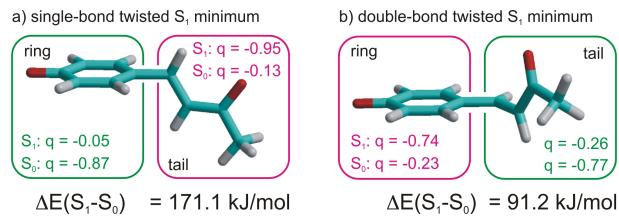
<sup>a</sup> S<sub>0</sub> GS: ground-state minimum, S<sub>1</sub> planar: planar-like minimum, S<sub>1</sub> DB twist: double-bond twisted minimum, S<sub>1</sub> SB twist: single-bond twisted minimum, S<sub>1</sub>/S<sub>0</sub> DB twist CI: double-bond twisted minimum energy conical intersection, S<sub>1</sub>/S<sub>0</sub> hula-twist CI: hula-twist minimum energy conical intersection. <sup>b</sup> Ground-state energies relative to S<sub>0</sub> GS energy of thio-pCA<sup>-</sup>. <sup>c</sup> Excited-state energies relative to S<sub>0</sub> GS energy of thio-pCA<sup>-</sup>. <sup>d</sup> Ground-state energies relative to S<sub>0</sub> GS energy of pCK<sup>-</sup>. <sup>e</sup> Excited-state energies relative to S<sub>0</sub> GS energy of pCK<sup>-</sup>.

by 90° (Figure 6, see also<sup>82</sup>). In the isolated chromophores, there is no barrier for reaching the single-bond twisted S<sub>1</sub> minimum from the Franck-Condon region, whereas there is a significant barrier to double-bond rotation (9.4 kJ/mol for pCK<sup>-</sup> and 14.0 kJ/mol for thio-pCA<sup>-</sup> at the CASSCF(12,11)/6-31G\* level, Table 1). Thus, in agreement with the molecular dynamics simulations, the main relaxation channel after excitation should involve rotation of the single bond to 90°. We furthermore found that the S<sub>1</sub>/S<sub>0</sub> intersection seam lies rather far away from this minimum. Based on these findings we suggest that radiationless decay is not very efficient in vacuum. This observation is in line with the relatively long excited-state lifetime of 52 ps for the isolated pCK<sup>-</sup> chromophore, measured by means of time-resolved photoelectron spectroscopy<sup>92</sup>. In subsequent QM/MM simulations we have probed the effect of different environments on the photochemistry of the chromophore.

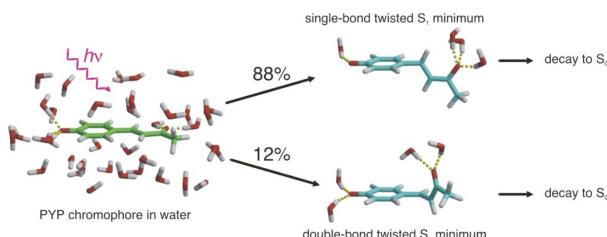
### Chromophore in water

To examine the effect of an aqueous environment, we have performed 91 QM/MM excited-state dynamics simulations of the deprotonated pCK<sup>-</sup> chromophore analogue in water (Figure 7)<sup>91</sup>. The pCK<sup>-</sup> chromophore was selected because experimental data are available both in gas phase<sup>92</sup> and in water<sup>93</sup>. The chromophore was described at the CASSCF(6,6)/3-21G level of theory, while the water molecules were modeled by the SPCE force field<sup>94</sup>.

The results of the simulations demonstrate that radiationless



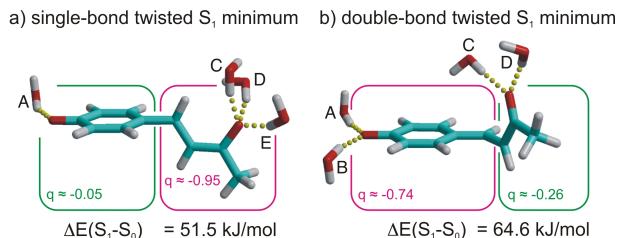
**Fig. 6** Excited-state minimum energy configurations of the pCK<sup>-</sup> chromophore analogue. The distribution of the negative charge over the ring and tail of the chromophore in the ground (S<sub>0</sub>) and excited (S<sub>1</sub>) state are shown. In both the single-bond twisted S<sub>1</sub> minimum (a) and the double-bond twisted S<sub>1</sub> minimum (b) there is a substantial energy gap between the ground and excited state. The distribution of the negative charge in these minima is opposite. The situation in the thio-pCA<sup>-</sup> is similar.



**Fig. 7** In water the chromophore undergoes both single and double bond isomerization. Excited state decay from these minima is very efficient due to stabilization of the chromophores S<sub>1</sub> charge distribution by specific hydrogen bond interactions. Adapted from Boggio-Pasqua *et al.*<sup>91</sup>

decay is very efficient in water<sup>91</sup>. The predominant excited-state decay channel involves twisting of the single bond (88%) rather than the double bond (12%). In contrast to vacuum, decay takes place very near these minima. Inspection of the trajectories revealed that decay is mediated by specific hydrogen-bond interactions with water molecules. These hydrogen bonds are different for the single- and double-bond twisted S<sub>1</sub> minima, which reflects the difference in charge distribution between these minima (Figures 6 and 8). In the single-bond twisted S<sub>1</sub> minimum, the negative charge resides on the alkene moiety of the chromophore (Figure 6). Three strong hydrogen bonds to the carbonyl oxygen stabilize this charge distribution to such an extent that the seam becomes accessible from the single-bond twisted S<sub>1</sub> minimum (Figure 8). In the double-bond twisted S<sub>1</sub> minimum, the negative charge is localized on the phenolate ring (Figure 6). Transient stabilization of this charge distribution by two or more strong hydrogen bonds to the phenolate oxygen brings the seam much closer to this S<sub>1</sub> minimum (Figure 8). Thus, in water the ultrafast excited-state decay is mediated by hydrogen bonds (Figure 7).

To quantify the stabilizing effect of the hydrogen bonds, we have also performed CASSCF/6-31G\* geometry optimiza-



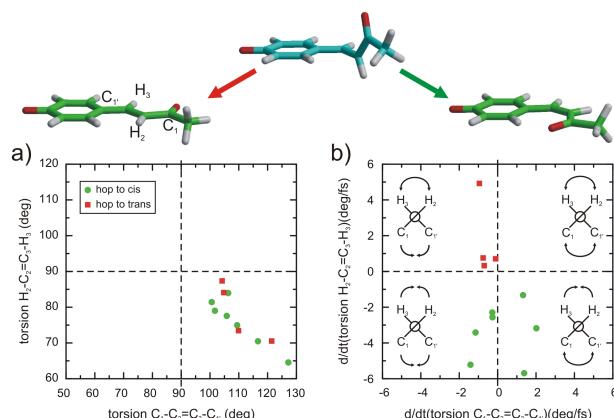
**Fig. 8** Excited-state minimum energy configurations of the pCK<sup>-</sup> chromophore analogue with four explicit water included. Adapted from Boggio-Pasqua *et al.*<sup>91</sup>

tions of the chromophore with and without water molecules. In these optimizations, the complete  $\pi$  system of the chromophore was included in the active space (12e,11o) and the water molecules were included explicitly. In addition to optimizing the local minima on the S<sub>1</sub> potential energy surface, we also searched for conical intersections in the vicinity of these minima.

Without water, there is a significant S<sub>0</sub>-S<sub>1</sub> energy gap at the S<sub>1</sub> minima (see discussion above). In the double-bond twisted structure, the seam lies  $\sim 29$  kJ/mol above the S<sub>1</sub> minimum, while no conical intersection was found near the single-bond twisted S<sub>1</sub> minimum. Including two water molecules near the ring of the chromophore leads to a substantial reduction of the S<sub>0</sub>-S<sub>1</sub> gap at the double-bond twisted S<sub>1</sub> minimum. In this case, a double-bond twisted S<sub>1</sub>/S<sub>0</sub> conical intersection was found only 5 kJ/mol higher in energy than this minimum. The two water molecules have the opposite effect at the single-bond twisted S<sub>1</sub> minimum: the S<sub>0</sub>-S<sub>1</sub> gap increases to 222 kJ/mol, and no conical intersection could be found near this structure. In contrast, including water molecules near the carbonyl oxygen atom reduces the S<sub>0</sub>-S<sub>1</sub> energy gap at the single-bond twisted structure. With three H-bonds to the carbonyl oxygen atom, the S<sub>0</sub>-S<sub>1</sub> gap becomes significantly smaller (25.6 kJ/mol) and reaching the seam from the minimum requires only 4.4 kJ/mol.

In 7 of the 11 trajectories that show twisting of the double bond, the chromophore continued isomerizing towards the cis configuration after decay to S<sub>0</sub>. In the other 4 simulations, the trans configuration was restored. To find out what determines the success of the isomerization in water, we analyzed the geometry changes just before the hop. Following Weingart<sup>95,96</sup>, we collected the torsion angles of the C<sub>1</sub>-C<sub>2</sub>=C<sub>3</sub>-C<sub>1'</sub> and H<sub>2</sub>-C<sub>2</sub>=C<sub>3</sub>-H<sub>3</sub> torsions at the moment of the hop, as well as the time-derivative of these torsion angles (see Figure 5 for definitions). The data, visualized in Figure 9, suggest that the success of double-bond isomerization depends solely on the angular velocity of the H<sub>2</sub>-C<sub>2</sub>=C<sub>3</sub>-H<sub>3</sub> torsion at the moment of hopping.

At the hop, the dihedral angle of the C<sub>1</sub>-C<sub>2</sub>=C<sub>3</sub>-C<sub>1'</sub> tor-



**Fig. 9** Values of torsion angles, and their time-derivatives.

sion is always higher than 90°, while that of the H<sub>2</sub>-C<sub>2</sub>=C<sub>3</sub>-H<sub>3</sub> torsion is always smaller than 90° (Figure 9a). Since hops at these geometries lead to both cis and trans configurations, the hopping geometry does not determine the outcome of the double-bond photoisomerization. In the plot of the time-derivatives of these dihedral angles (Figure 9b) the successful (green dots) and unsuccessful hops (red squares) are separated along the y-axis, which represents the angular velocity of the H<sub>2</sub>-C<sub>2</sub>=C<sub>3</sub>-H<sub>3</sub> torsion. Apparently, the chromophore decays into the cis minimum, if the hydrogen atoms move towards each other (negative gradient,  $y < 0$ ) at the hop, whereas the chromophore reverts to the trans minimum if the hydrogen atoms move away from each other (positive gradient,  $y > 0$ ). The direction in which the C<sub>1</sub> atom moves with respect to C<sub>1'</sub> atom seems not relevant for the outcome of the double-bond photoisomerization process, as hops leading the cis or trans both occur at positive and negative angular velocities of the C<sub>1</sub>-C<sub>2</sub>=C<sub>3</sub>-C<sub>1'</sub> torsion ( $x$ -axis). Statistically the number of trajectories is small, but they nevertheless yield a consistent mechanistic picture. We note that with a similar analysis Weingart and Schapiro have elucidated the role of the hydrogen motion on the isomerization quantum yield of retinal<sup>95-97</sup>.

Weingart *et al.*<sup>96</sup> have provided an explanation for the observation above in terms of  $\pi$ -orbital axis vectors (POAV) theory<sup>98</sup>. In short, the dynamics of the hydrogen atoms cause pyramidalization of the sp<sup>2</sup>-hybridized carbon atoms. Because pyramidalization affects the orientation of the  $\pi$ -orbital axis vectors, the motion of the hydrogen atoms determines the direction in which the  $\pi$ -orbital overlap is maximized after the hop. After the new double bond is formed, the carbon atoms rapidly adjust to the new bonding situation.

Summarizing, the predominant excited-state relaxation process of pCK<sup>-</sup> in water involves a rotation of the single bond, adjacent to the ring (Figure 7), rather than rotation of the dou-

ble bond. In vacuo, only the double-bond rotation can lead to radiationless decay, whereas in water, both channels lead to decay. Because single-bond photoisomerization is strongly favored over double-bond photoisomerization for pCK<sup>-</sup> in water, the probability of finding the chromophore in the cis configuration is very low. This result is in good agreement with the very low isomerization quantum yield, observed by Espagne and coworkers<sup>93</sup>. In their time-resolved fluorescence measurements the excited-state lifetime was determined for a series of chromophore derivatives with different substituents at the *p*-coumaric acid group. The pCK<sup>-</sup> chromophore was found to decay the fastest ( $\sim 1$  ps), albeit with a negligible trans-to-cis photoisomerization quantum yield. Although we have not studied the thio-ester derivatives of the *p*-coumaric acid in water, the lack of isomerization quantum yield at high pH, as observed in time-resolved spectroscopic measurements<sup>99-101</sup>, also suggests the involvement of the unproductive single-bond isomerization channel in these chromophores.

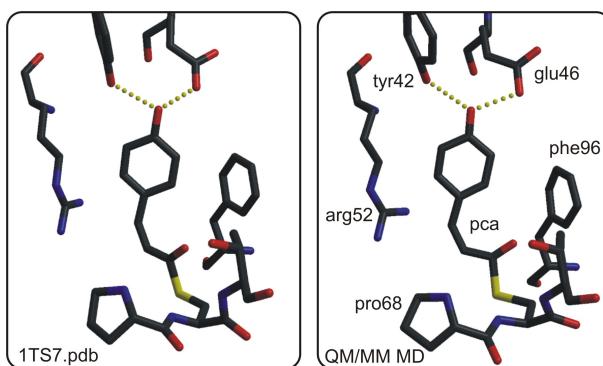
### Effect of the protein environment

By displacing the S<sub>1</sub>/S<sub>0</sub> seam very close to the S<sub>1</sub> minima, hydrogen-bond interactions between the chromophore and water molecules enhance excited-state decay in water<sup>91</sup>. Yet, the major decay channel involves rotation of the single bond, rather than the double bond of the chromophore. To find out the effect of the protein environment on the excited-state dynamics of the chromophore, we have also carried out QM/MM simulations of the wild-type PYP<sup>19</sup>, as well as on the Arg52Gln mutant<sup>102</sup>.

In our first study, we used the CASSCF/3-21G level of theory with 6 electrons in 6  $\pi$  orbitals to describe the chromophore and the Gromos96 force field<sup>57</sup> for the apoprotein, water and ions, to compute 14 trajectories starting from the PYP X-ray structure (PDB ID: 2PHY)<sup>103</sup>. At this level of theory, it is not possible to make quantitative predictions on the effect of the protein on the optical absorption spectrum. This issue, however, has been addressed for instance by Gromov *et al.*,<sup>104</sup> by Coto *et al.*,<sup>105</sup> and very recently, by Rocha-Rinza *et al.*<sup>75</sup>, who used correlated ab initio methods. Figure 4 shows the primary events after photoexcitation in the simulation of the wild-type. In the majority of the trajectories the chromophore rapidly decays to the ground state via a 90° rotation of the double bond (Figure 4), rather than the single bond. During this photoisomerization process, the hydrogen bonds between the chromophore's phenolate oxygen atom and the side chains of the highly conserved Tyr42 and Glu46 residues remain intact. Just as in water, these hydrogen bonds are essential to promote excited-state decay from the double-bond twisted minimum.

Upon returning to the ground state the chromophore either relaxes back to the original trans conformation (180°), or it

continues isomerizing to a cis conformation (0°). In the latter case, the relaxation also involves a flip of the thioester linkage, which causes the carbonyl group to rotate by 180°. During this rotation, the hydrogen bond between the carbonyl oxygen and the Cys69 backbone amino group is broken (Figure 4). The loss of this hydrogen bond has been observed in time-resolved infrared spectroscopy<sup>106,107</sup> as well as in time-resolved Laue crystallography studies<sup>108</sup>. In figure 10, we have fitted a snapshot, taken 5 ps after photoexcitation, onto the X-ray structure of the pRCW intermediate (PDB ID: 1TS7), using the C<sub>α</sub> atoms. The root-mean-square deviation of this fit was less than 0.09 nm, indicating the high similarity of these two structures.



**Fig. 10** Comparison between time-resolved Laue structure (left)<sup>108</sup> and simulated structure (right)<sup>19</sup> of the pRCW intermediate state. The structures were superimposed, using the C<sub>α</sub> atoms for the least-squares fit.

Although the number of trajectories is too small to rule out that single-bond isomerization could also occur in the wild-type protein, no single-bond isomerization was observed. Because in water single-bond isomerization is the main decay channel, the latter observation implies that the protein not only provides the hydrogen bonds required for ultrafast decay, but also controls which of the chromophore bonds isomerizes upon photoexcitation. This finding is in line with calculations by Yamada *et al.*, who used ONIOM CASSCF calculations to reveal the initial excited state relaxation of the chromophore in the protein.<sup>109</sup> We speculated that the positive guanidinium moiety of Arg52 located just above the chromophore ring, acts as the 'catalytic' residue that enforces double-bond isomerization. Initially, we tested this hypothesis by removing the charge from the Arg52 side chain in a snapshot taken at the hop, and recalculating the energy gap. We found that the gap increases by 40 kJ/mol<sup>19</sup>. This result led us to propose that a preferential electrostatic stabilization of the double-bond twisted S<sub>1</sub> minimum (Figure 6) by the positive Arg52 favors double-bond isomerization over single-bond isomerization in the wild-type protein. More recent work of Martínez and co-workers on an isolated pCK<sup>-</sup> chromophore in the presence of

a point charge confirm that the photoisomerization can indeed be controlled by nearby charges<sup>90</sup>.

To elucidate the role of this arginine in the activation process in more detail, we have also performed 40 excited-state dynamics simulations on the Arg52Gln mutant of PYP<sup>102</sup>, starting from its X-ray structure (PDB-ID: 2D02)<sup>110</sup>. This mutant has absorption and emission spectra very similar to the wild-type<sup>111</sup> and can still enter the photocycle, albeit with a lower rate and quantum yield<sup>112</sup>. Without the positive Arg52, the predominant excited-state reaction in the simulations involved isomerization of a single bond in the chromophore, rather than the double bond (Figure 11)<sup>102</sup>. Although the latter was also observed, it was a minor decay channel. Again, the number of trajectories is too low to make quantitative predictions. Nevertheless, the mutant simulations confirm that within our model the role of Arg52 is to steer the initial events after photon absorption to ensure rotation of the double bond rather than the single bond in the chromophore.

Although single-bond isomerization does not result in the formation of the cis chromophore, a 180° flip of the thioester group and a rupture of the hydrogen bond to Cys69 (Figure 4) was observed in 7 simulations. Together with the experimental observation that the mutant can still enter the photocycle<sup>111</sup>, this suggests that the key step to enter the photocycle is the carbonyl flip, rather than the double-bond isomerization. This suggestion is further supported by the observation that when the protein is reconstituted with chromophore analogues that cannot isomerize<sup>100,101</sup>, PYP can still enter the photocycle<sup>113</sup>. In these chromophores the isomerization is prevented by either a covalent bridge that blocks both single- and double-bond rotations, or by oxidation of the double bond to a triple bond.

However, since trans-to-cis isomerization of the double bond was also observed, we cannot rule out the possibility that only the latter process is responsible for photoactivation and that the single-bond isomerization pathway is a dead-end in the process, responsible for the lower activation yield in the mutant. Support for the involvement of the double-bond photoisomerization comes from time-resolved absorption spectroscopy measurements that show the appearance of the cis photoproduct on a sub-picosecond timescale in this mutant<sup>111</sup>. In these experiments, the lower quantum yield and slower deactivation dynamics as compared to the wild type were attributed to a lower structural integrity of the chromophore pocket. Based on our simulations, however, we would attribute the difference in efficiency to the involvement of the single-bond isomerization channel.

Also in the mutant the hydrogen-bonding interactions are essential for the decay process. During rotation of the single bond, the hydrogen bond between the carbonyl oxygen and Cys69 backbone amino group is broken. In contrast to the wild-type, the rupture of this bond occurs before the decay. Therefore this may seem at conflict with the observation that

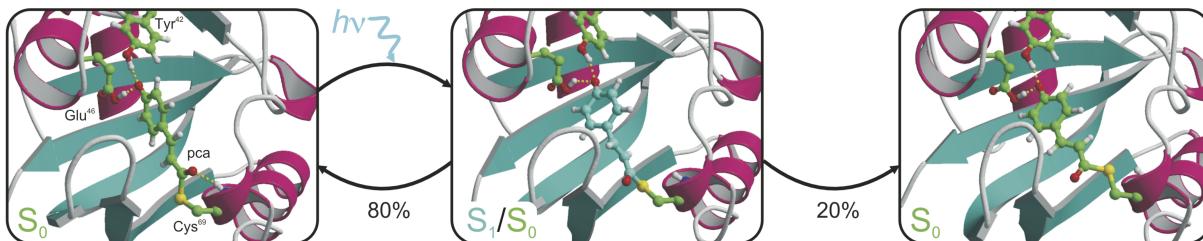
in water three hydrogen bonds at the carbonyl are required to make the S<sub>1</sub>/S<sub>0</sub> seam accessible from the single-bond twisted minimum. However, as shown in Figure 12, new hydrogen bonds are rapidly formed between the carbonyl oxygen atom and the backbone amino groups of Tyr98 and Asp97. A water molecule from outside enters the chromophore pocket to donate a third hydrogen bond. This is possible because the chromophore pocket has become more exposed to the solvent due to the replacement of the rather bulky arginine side chain by the much smaller glutamine side chain. With these three hydrogen bonds stabilizing the negative charge on the alkene moiety, the chromophore rapidly decays to S<sub>0</sub>. Thus, the decay mechanisms in the Arg52Gln mutant and in water are essentially the same in our simulations<sup>91,102</sup>.

To summarize, the simulations have provided detailed structural and dynamical information at a resolution well beyond that achievable by other means. From the simulations, we have identified key amino acids and the mechanism by which they control the primary events in the photocycle of PYP. These are (i) double-bond photoisomerization, and (ii) the break of a hydrogen bond between the chromophore and the protein backbone. These events trigger a proton transfer from the protein to the chromophore, which ultimately leads to the signaling state of PYP<sup>114–121</sup>.

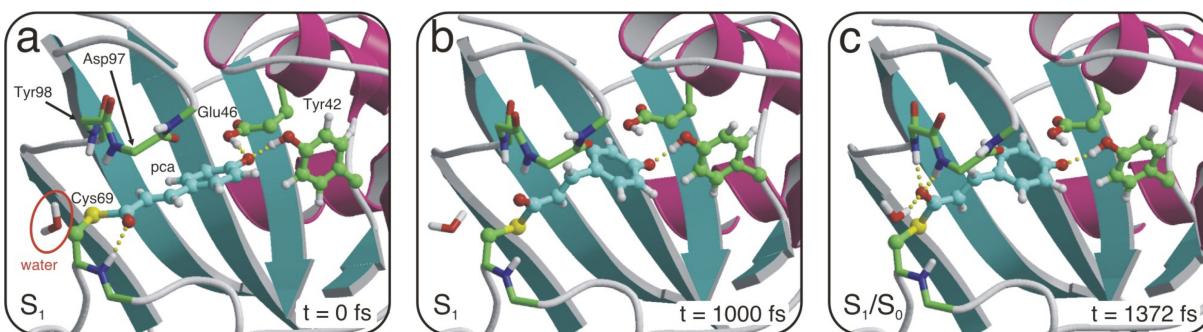
### Comparison to experimental data and challenges

Time-resolved spectroscopy in the visible and mid-infrared spectral regions has provided detailed information on the characteristics of PYP intermediates and their kinetics<sup>106,107,122</sup>. In addition, crystallographic studies have provided detailed information about the structural changes that follow photoexcitation<sup>108,123</sup>. Thus, there is plenty of experimental data available to compare the outcome of the simulations to. As we have already shown, there is good agreement between the pRCw intermediate found in the simulations and experiment. Furthermore, excited-state lifetimes, defined in our simulations as the time until the surface hop occurs, agree reasonably with experiment: 0.3 ps for pCK<sup>–</sup> in water (exp. ~1 ps<sup>93</sup>); 0.2 ps for the wild type (exp. ~0.4 ps<sup>124,125</sup>); 0.4 ps for the Arg52Gln mutant (exp. 0.6 ps<sup>111</sup>). Isomerization quantum yields are also in line with experiment: 0.06 for pCK<sup>–</sup> in water<sup>91</sup> (exp. ~0.93); and 0.3 for the wild type<sup>19</sup> (exp. 0.35<sup>76</sup>); 0.2 for the Arg52Gln mutant<sup>102</sup> (exp. 0.21<sup>112</sup>). However, the apparent agreement of the latter quantities with the spectroscopic data is no guarantee that the underlying dynamics is correct.

Both quantum yield and lifetime are highly sensitive to the sampling as well as the quality of the underlying Hamiltonian. Because sampling is restricted to a low number of trajectories, we can only explore a limited area of the phase space available to the true system. This problem is most severe in the protein simulations. Therefore, we cannot exclude that we



**Fig. 11** Snapshots from an excited-state trajectory of the Arg52Gln mutant of PYP, showing the chromophore (thio-pCA<sup>-</sup>) in the active site pocket. The first snapshot is at the excitation. The second shows the configuration at the radiationless transition from S<sub>1</sub> to S<sub>0</sub>. The third snapshot shows the photoproduct. In the mutant isomerization takes place around the single bond. Like in the wild-type protein, the carbonyl oxygen of the thioester linkage flips, causing the break of the hydrogen bond to the backbone of Cys69. Adapted from Groenhof *et al.*<sup>102</sup>



**Fig. 12** Snapshots from an excited-state trajectory of the Arg52Gln mutant of PYP, demonstrating that three hydrogen bonds to the carbonyl moiety are essential for S<sub>1</sub> decay near the single-bond twisted minimum. The first snapshot (a) is at the excitation to S<sub>1</sub>. The second snapshot (b) shows the twisted configuration without hydrogen bonds to the carbonyl. The gap between S<sub>1</sub> and S<sub>0</sub> is far too high for decay at this configuration. However, as the third snapshot (c) shows two backbone amino groups and a bulk water that has moved into the chromophore pocket during the excited-state dynamics, donate the three hydrogen bonds that are required for efficient decay from the S<sub>1</sub> minimum. Adapted from Boggio-Pasqua *et al.*<sup>91</sup>

have sampled only minor conformations that in reality do not contribute significantly to the observed photochemistry. Lack of sampling can only be overcome by running an impossibly large number of simulations. Thus, even if all accuracy issues of the Hamiltonian may be overcome in the future, the sampling problem will persist.

Since we use classical molecular dynamics, the outcome of a trajectory is determined by the QM/MM potential energy surface. The accuracy of the predicted lifetimes, and isomerization quantum yields is therefore limited by the accuracy of the level of theory. In our simulations, we use the CASSCF method in combination with default molecular mechanics force fields to approximate the potential energy surface of hydrated PYP. The S<sub>1</sub> state of the chromophore is a charge transfer state.<sup>126</sup> Because the CASSCF wave function lacks dynamic electron correlation, we systematically overestimate the S<sub>1</sub> energy. Wave function accuracy is further compromised because we have to truncate our active spaces, and use small split-valence basis sets (without polarization and diffuse functions) to overcome computational bottlenecks during the molecular dynamics simulations. Finally, the use of default force field parameters to describe the interaction with the rest of the systems also introduces errors. Unfortunately, these errors are more difficult to control than the ones in the wave function, as there is no systematic way of improving the MM description in QM/MM simulations. Because of all these approximations, a quantitative agreement of quantum yields and excited-state lifetimes needs to be taken with care.

The accuracy of the QM/MM potential energy surface may also affect the qualitative predictions, in particular, the involvement of single-bond isomerization. To illustrate the effect of the level of theory on the branching between the single- and double-bond isomerization channels, we have explored the S<sub>1</sub> potential energy surface of pCK<sup>-</sup> at different levels of theory (Table 2). We optimized the S<sub>1</sub> planar-like structure and the transition states for single- and double-bond isomerizations to evaluate the respective potential energy barriers. Although the single- and double-bond twisted minima exist on the S<sub>1</sub> surface at all levels, the barriers separating these minima from the S<sub>1</sub> planar-like structure strongly depend on the level of theory (Table 2).

The use of a small active space (6e,6o) and basis set (3-21G) gives a similar value ( $\sim 8$  kJ/mol) for the single- and double-bond torsion barriers. When state-averaged orbitals are used like in our MD simulations, the single-bond torsion barrier vanishes, while the double-bond torsion barrier is reduced to  $\sim 4$  kJ/mol, which changes the outcome of the dynamics of the system. In the first case, one can expect to observe both single- and double-bond isomerizations, while in the second case, single-bond isomerization will be favored because it is a barrierless process. Compared with the CASSCF results using the full  $\pi$  system (12e,11o), the absence of the single-bond

torsion barrier is confirmed, but the barrier for double-bond isomerization is too low at the state-averaged (6e,6o) level. The same trend is observed at higher levels of theory that include dynamic electron correlation, such as CASPT2 and EOM-CCSD. They all predict a very small single-bond torsion barrier (0-4 kJ/mol) and a more significant double-bond torsion barrier (10-30 kJ/mol).

These results could explain why we do not have a perfect quantitative agreement with the experimental data for the trans-to-cis isomerization quantum yield of pCK<sup>-</sup> in water. While no such isomerization was observed experimentally<sup>93</sup>, a small quantum yield of 7% was calculated from our simulations, which is probably due to the too low energy barrier for double-bond isomerization state-averaged CASSCF (6e,6o) level.

Since the branching ratio of single- and double-bond photoisomerization depends critically on the relative height of the barriers, a systematic deviation may artificially favor one channel over the other. Thus, the accuracy may not only affect the time scale, but also the qualitative outcome of the excited-state dynamic simulations. This is an additional matter of concern when the photoisomerization in the protein environment is considered, as excited-state potential energy barriers cannot be routinely computed in such complex environments. Moreover, the description of the chromophore-protein interactions, in particular the hydrogen bonds, can also influence these relative barriers (see discussion below).

To validate the small active space used in our simulations, and at the same time check the influence of the force field, we have repeated two simulations for both wild-type and Arg52Gln mutant in which we combined the CASSCF(10,9)/6-31G\* level of theory for the chromophore and the Amber03 force field<sup>58</sup> for the rest. In these simulations, the active space thus contained 10 electrons in 9  $\pi$  orbitals. Calculations on the isolated chromophore at this level demonstrate that leaving out the lowest-energy bonding and highest energy antibonding  $\pi$  orbitals does not significantly affect the properties of the chromophore. The computational cost of the simulations at the higher level of theory (almost 1 hour per step on four Intel Xeon CPUs running at 2.6 GHz) prevented us from generating more than four trajectories. The results of the simulations at the higher level confirm the simulations at the lower level. In the wild-type, two unsuccessful isomerizations around the double bond were observed, whereas in the mutant trajectories excited-state decay involved a rotation around the single bond. Despite the very low number of simulations at the higher level, the lower level of CASSCF theory used in our previous studies seems to be adequate to describe the dynamics of the chromophore.

The S<sub>1</sub> surface is also very sensitive to interactions of the chromophore with the environment. Hydrogen bonds not only alter the position of the S<sub>1</sub>/S<sub>0</sub> seam, but also influence the

**Table 2** Potential energy barriers (kJ/mol) for single-bond and double-bond torsions at different levels of theory.

System and level of theory <sup>a</sup>	SB-twist barrier <sup>b</sup>	DB-twist barrier	$\phi_{\text{SB}}^c$	$\phi_{\text{dB}}^d$
<b>pCK<sup>-</sup></b>				
CASSCF(6,6)/3-21G	8.5	8.2	17°	35°
SA2-CASSCF(6,6)/3-21G	n.f. (~ 0)	3.8	--	21°
CASSCF(8,8)/6-31G*	0.1	11.8	13°	38°
SA2-CASSCF(8,8)/6-31G*	n.f. (~ 0)	15.3	--	55°
CASSCF(12,11)/6-31G*	~ 0	9.4	7°	35°
CASSCF(12,11)/6-31+G*	0.9	7.0	22°	32°
CASPT2(12,11)/cc-pVDZ	n.c.	17.5	--	35°
CASPT2(12,11)/aug-cc-pVDZ	0.8	28.6	22°	32°
EOM-CCSD/cc-pVDZ	n.c.	10.0	--	35°
EOM-CCSD/aug-cc-pVDZ	3.9	14.7	22°	32°
<b>pCK<sup>-</sup> + 2 H<sub>2</sub>O near phenolate</b>				
CASSCF(12,11)/6-31G*	2.7	5.1	32°	28°
CASPT2(12,11)/6-31G*	1.1	11.6	32°	28°

<sup>a</sup> CASPT2 and EOM-CCSD calculations were performed at the corresponding CASSCF optimized geometries. <sup>b</sup> n.f.: transition state not found because the potential energy surface is nearly flat, so that the barrier is nearly zero; n.c.: not computed because CASSCF(12,11)/6-31G\* does not provide a well defined transition state. <sup>c</sup>  $\phi_{\text{SB}}$ : dihedral angle measuring the torsion around the single bond at the S<sub>1</sub> single-bond-twisted transition state optimized at the CASSCF level. <sup>d</sup>  $\phi_{\text{DB}}$ : dihedral angle measuring the torsion around the double bond at the S<sub>1</sub> double-bond-twisted transition state optimized at the CASSCF level.

height of the barriers between the Franck-Condon region and the twisted S<sub>1</sub> minima<sup>127</sup>. Since we model these interactions at the QM/MM level, the question arises if the employed Coulomb-only embedding method is sufficiently accurate. For pCK<sup>-</sup> we reproduced the stabilizing effect of hydrogen bonding on the seam, as observed in the QM/MM dynamics, at the CASSCF(12/11)/6-31G\* level for the microsolvated chromophore<sup>91</sup>. The heights of the barriers, however, could not be determined at the QM/MM level.

Because twisting the single bond is accompanied by a migration of negative charge away from the phenolate ring, hydrogen bonding at the phenolate oxygen (O<sub>4'</sub>, Figure 5) hampers this process and the barrier to single-bond rotation increases (Table 2). Nevertheless, the barrier for single-bond twisting (2.7 and 1.1 kJ/mol at CASSCF and CASPT2 levels, respectively) is still significantly lower than for double-bond twisting (5.1 and 11.6 kJ/mol at CASSCF and CASPT2 levels, respectively). The predominance of single-bond isomerization observed in the QM/MM trajectories is consistent with higher-level QM data. The QM/MM description of the hydrogen-bond interactions between the chromophore and water seems thus sufficiently realistic.

Whether the QM/MM model also provides a realistic description of the hydrogen bonds in the protein is more difficult to answer. Recent crystallography studies have shown

that the hydrogen bonds between the phenolate oxygen and the side chains of Tyr42 and Glu46 are shorter (O-O distance about 0.25 nm) than normal hydrogen bonds (0.27 nm)<sup>128–130</sup>. At closer distances, the hydrogen bonds may perturb the electronic structure of the chromophore stronger and thereby affect the excited state dynamics. Recent computations by Gromov *et al.* at the CC2 level of theory<sup>131</sup> confirm that there is indeed such effect<sup>132</sup>. However, neither the Gromos96, nor the Amber03 force fields can model such short hydrogen bonds due to the strong repulsion ( $C_{12}$ , equation 10) in the oxygen-oxygen Lennard-Jones potential. In classical simulations of PYP, using an X-ray structure with the short hydrogen bonds (PDB-ID: 2ZOH), the hydrogen bonds are immediately restored to longer hydrogen bonds (data not shown). Thus, we cannot reproduce these structural features in our QM/MM simulations of PYP.

To address the effect of the shorter hydrogen bonds in our simulations, a different QM/MM setup would be required. One option would be to enlarge the QM subsystem to include also the sidechains of Tyr42 and Glu46 and increase the level of theory at the same time to improve the hydrogen bonding energetics. This strategy, however, would dramatically increase the computational demands. An alternative would be to use the constraint (or frozen) DFT method<sup>64–66,133</sup> to describe the interactions between the chromophore and the hydrogen

bond donors. Because the SCF is performed on smaller fragments, while keeping the electron density of the remaining fragments frozen, this method is computationally more efficient than full *ab initio* or DFT and more accurate than normal QM/MM. However, the validity of the available DFT functionals to describe the excited state potential energy surface of the chromophore, as well as the surface crossing remains to be verified.

At the moment, however, we can only speculate about the role of the short hydrogen bonds on the photoisomerization process in PYP. Before we do so, we note that several authors have suggested that at least one of these hydrogen bonds (Glu46) can be classified as a low-barrier-hydrogen bond (LBHB).<sup>130,134</sup> LBHBs are characterized by a single, rather than double well potential for the hydrogen, so that it is equally shared by donor and acceptor. The hydrogen bond therefore is believed to have substantial covalent character.<sup>135</sup> Although in vacuum, where delocalization of charge is energetically favorable, LBHB may be rather common, it is not clear whether LBHB can exist in polar media, as the delocalization of charge would reduce the solvation by the medium's dipoles. It was shown for instance that in the catalytic triad of serine proteases, the solvation by the protein and solvent dipoles strongly favors an ionic hydrogen bond between Histidine and Aspartate, rather than a LHBH.<sup>136,137</sup> Whether the chromophore pocket of PYP is sufficiently apolar for a LHBH to form between the chromophore and Glu46 remains to be established. Recent computations by Saito and Ishikita suggest that this is not the case, and that this hydrogen bond is of ionic character.<sup>138</sup>

Perhaps the most appropriate way to resolve this issue is to construct a valence bond Hamiltonian for the chromophore and the hydrogen bond donors and perform the dynamics within the empirical valence bond (EVB) framework,<sup>139</sup> as was done by Warshel and co-workers to investigate the claims of LHBH in serine proteases.<sup>137</sup> As the EVB method provides also the computationally most efficient means to describe the dynamics of bond breaking processes, it is probably also the method of choice to investigate the proton transfer reaction which takes place after isomerization, as was done for instance in bacteriorhodopsin<sup>140</sup>.

One role of the short hydrogen bonds in photoisomerization process may be to bias the isomerization towards the double bond. By stabilizing the negative charge on the ring, these hydrogen bonds may increase the single-bond isomerization barrier, while simultaneously decrease the double-bond isomerization barrier. Recent computations by Gromov *et al.* confirm such picture<sup>127</sup>. Previously, we proposed that in the protein double-bond isomerization is favored over single-bond isomerization by the presence of the highly conserved Arg52. The electrostatic interaction between the excited-state charge distribution (Figure 6) on the chromophore and the positively

charged guanidinium group located near the ring selectively stabilizes the twisted double-bond S<sub>1</sub> minimum, as well as the barrier to reach that minimum. However, a recent high-resolution neutron crystallographic analysis combined with high-resolution X-ray crystallography has revealed that the side chain of Arg52 is in its neutral deprotonated state<sup>130</sup>. This finding obviously challenges our interpretation of the role of Arg52 in the photoactivation process. We therefore speculate that selection between the two isomerization channels may be due to the short hydrogen bonds alone. New calculations at a higher level of theory are required to confirm this hypothesis.

Another important experimental finding that is difficult to reconcile with our simulations, is the observation of a ground-state intermediate different from pR<sub>CW</sub> in time-resolved infrared measurements<sup>107,122</sup>. This intermediate forms in competition with pR<sub>CW</sub> but decays back into the trans ground state in few picoseconds. Such ground-state intermediate was also observed for chromophore analogues in solution.<sup>99,101</sup> Although it is tempting to assign this intermediate to the single bond isomerization product, more recent experiments, in which the single bond adjacent to the chromophore ring was locked, suggest that this cannot be the case.<sup>141</sup> Absorption difference IR spectra suggest that in this intermediate the chromophore is in the cis conformation, but in contrast to pR<sub>CW</sub> maintains its hydrogen bond to the amino group of Cys69. In our simulations, no such intermediate was observed. Instead, the unsuccessful trajectories directly decay back into the trans minimum after the surface hop and never reach a cis conformation. Thus, although the quantum yield of activation is in good agreement with experiment, the mechanism of the unsuccessful isomerization attempts seems not.

The experimental data seem to imply that upon photon absorption the chromophore always achieves a trans-to-cis isomerization, and that the success of activation depends on breaking the hydrogen bond between the chromophore and the backbone. In contrast, only 30 % of our trajectories reach the isomerized state, and breaking of the hydrogen bond is successful in all of these cases. At this stage it is difficult to say what is the cause for the discrepancy. Given the role of the hydrogen motion on the success rate of the photoisomerization in water, an additional source of error may be the classical description of such motions.

Finally, we briefly comment on our surface-hopping algorithm. Our hopping method is based on a local diabatic representation of the intersection region. This method can qualitatively capture the main features of the hopping process but is too simple to yield accurate quantitative predictions. For instance, for small molecules in isolation, the hopping probability is generally underestimated and the decay time of excited states is on average five times larger than that computed with the fewest-switches method<sup>89</sup>. The main weakness of the diabatic hopping method is that it does not incorporate any

contribution from regions with weak non-adiabatic coupling. Since the location of the hop determines lifetime and quantum yield, part of the discrepancies may be resolved by using the fewest-switches algorithm instead.

In this section we have reassessed the validity of QM/MM excited-state dynamics simulations. We have shown that our model yields results that are in agreement with experiment, but also shows some deviations. We made clear how the simulations could be improved, albeit at the expense of a higher demand on computational resources. Thanks to the ongoing improvements in the latter, as well as in algorithmic developments, we are confident however that in the near future systematic improvements in the description of the interactions are possible that can increase the predictive power of the method. Nevertheless, even at this early stage, important insights into photochemical processes can be obtained. Our purpose in modeling such processes is not to provide a completely correct description that would make experiment obsolete, but rather to generate ideas to design new experiments that will ultimately lead to a better understanding. Moreover, experiments, designed to validate specific aspects of the theory, such as the involvement of single-bond isomerization,<sup>141</sup> or the role of Arg52,<sup>111</sup> are also important to systematically improve the theory.

## Conclusion and Outlook

Understanding light-driven processes is a major goal of the bio- and nanosciences. The underlying molecular mechanisms are typically governed by sub-picosecond atomic motions. Mechanisms on such ultrafast timescales are very challenging to probe by experiment. Here, molecular dynamics simulations have become an invaluable tool to understand such processes in atomic detail. In this perspective, we have reviewed our approach to model excited-state processes in biological systems. In the application that we have selected here, the simulations could provide detailed structural and dynamical information of the photobiological processes in photoactive yellow protein at a resolution well beyond what is achievable experimentally. This application also demonstrates what is feasible today with on-the-fly molecular dynamics simulations, and where the limits are. These limits are predominantly imposed by the current state of computer technology, which restricts both system size and timescale of the processes under study. However, the expected increase of computer power, complemented by the development of more efficient electronic structure methods and new algorithms, will enable the study of larger systems and longer timescales in the future. Therefore, excited-state molecular dynamics simulation has the potential to ultimately lead to a better understanding of photobiological reactions. Furthermore, the simulations will enable the prediction of photochemical properties and thereby aid the rational

design of artificial light-driven systems<sup>142</sup>.

## Acknowledgements

We thank the University of Toulouse for providing an invited lecturer position for G. G., and the Volkswagen foundation for financial support.

## References

- 1 A. Warshel, *Nature*, 1976, **260**, 679–683.
- 2 A. Warshel and N. Barboy, *J. Am. Chem. Soc.*, 1982, **104**, 1469–1476.
- 3 A. Warshel, Z. T. Chu and J. K. Hwang, *Chem. Phys.*, 1991, **158**, 303–314.
- 4 A. Warshel and Z. Chu, *J. Phys. Chem. B*, 2001, **105**, 9857–9871.
- 5 A. Warshel and W. Parson, *Quart. Rev. Biophys.*, 2001, **34**, 563–679.
- 6 D. Polli, P. Altoe, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R. A. Mathies, M. Garavelli and G. Cerullo, *Nature*, 2010, **467**, 440–443.
- 7 L. M. Frutos, T. Andruniow, F. Santoro, N. Ferre and M. Olivucci, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 7764–7769.
- 8 S. Hayashi, E. Tajkhorshid and K. Schulten, *Biophys. J.*, 2009, **96**, 403–416.
- 9 S. Hayashi, E. Tajkhorshid and K. Schulten, *Biophys. J.*, 2003, **85**, 1440–1449.
- 10 X. Li, L. W. Chung, H. Mizuno, A. Miyawaki and K. Morokuma, *J. Chem. Phys. Lett.*, 2010, **1**, 3328–3333.
- 11 M. Andresen, A. C. Stiel, S. Trowitzsch, G. Weber, C. Eggeling, M. C. Wahl, S. W. Hell and S. Jakobs, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 13005–13009.
- 12 W. Parson and A. Warshel, *J. Phys. Chem. B*, 2004, **108**, 10474–10483.
- 13 A. Warshel, Z. Chu and W. Parson, *Science*, 1989, **246**, 112–116.
- 14 Z. Chu, A. Warshel and W. Parson, *Photosynth. Res.*, 1989, **22**, 39–46.
- 15 G. Groenhof, L. V. Schäfer, M. Boggiro-Pasqua, M. Goette, H. Grubmüller and M. A. Robb, *J. Am. Chem. Soc.*, 2007, **129**, 6812–6819.
- 16 Y. Lu, Z. Lan and W. Thiel, *Angew. Chem.-Int. Ed.*, 2011, **50**, 6864–6867.
- 17 D. Nachtigallova, T. Zeleny, M. Ruckenbauer, T. Mueller, M. Barbatti, P. Hobza and H. Lischka, *J. Am. Chem. Soc.*, 2010, **132**, 8261–8263.
- 18 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
- 19 G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. De Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark and M. A. Robb, *J. Am. Chem. Soc.*, 2004, **126**, 4228–4233.
- 20 I. Schapiro, F. Melaccio, E. N. Laricheva and M. Olivucci, *Photochem. Photobiol. Sci.*, 2011, **10**, 867–886.
- 21 W. Domcke, D. R. Yarkony and H. Köppel (eds.), *Conical Intersections, Advanced Series in Physical Chemistry Vol 15*, World Scientific, Singapore, 2004.
- 22 F. Bernardi, M. Olivucci and M. A. Robb, *Chem. Soc. Rev.*, 1996, **25**, 321.
- 23 M. M. J. Bearpark and M. A. Robb in, *Reviews of reactive intermediate chemistry*, Wiley, New Jersey, 2007, pp. 379–414.
- 24 J. C. Tully and R. Preston, *J. Chem. Phys.*, 1972, **55**, 562–572.
- 25 J. C. Tully, *J. Chem. Phys.*, 1990, **93**, 1061–1071.
- 26 J. C. Tully, *Int. J. Quant. Chem.*, 1991, **25**, 299–309.
- 27 J. C. Tully, *Faraday Discuss.*, 1998, **110**, 407–419.
- 28 T. J. Martinez, M. Ben-Nun and R. D. Levine, *J. Phys. Chem.*, 1996, **100**, 7884–7895.
- 29 W. Parson and A. Warshel, *Chem. Phys.*, 2004, **296**, 201–216.

- 30 O. V. Prezhdo and P. J. Rossky, *J. Chem. Phys.*, 1997, **107**, 825–834.
- 31 F. Webster, P. J. Rossky and R. A. Friesner, *Comp. Phys. Comm.*, 1991, **63**, 494–522.
- 32 F. Webster, E. Wang, P. Rossky and R. Friesner, *J. Chem. Phys.*, 1994, **100**, 4835–4847.
- 33 E. R. Bittner and P. J. Rossky, *J. Chem. Phys.*, 1995, **103**, 8130–8143.
- 34 S. Hammes-Schiffer and J. C. Tully, *J. Chem. Phys.*, 1994, **101**, 4657–4667.
- 35 P. E. M. Siegbahn, J. Almlöf, A. Heiberg and B. O. Roos, *J. Chem. Phys.*, 1981, **74**, 2384–2396.
- 36 A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612–&.
- 37 A. Toniolo, S. Olsen, L. Manohar and T. J. Martinez, *Faraday Discuss.*, 2004, **127**, 149–163.
- 38 T. W. Keal, A. Koslowski and W. Thiel, *Theor. Chem. Acc.*, 2007, **118**, 837–844.
- 39 K. K. Baek and T. J. Martinez, *Chem. Phys. Lett.*, 2003, **375**, 299–308.
- 40 I. Tavernelli, U. F. Rohrig and U. Rothlisberger, *Mol. Phys.*, 2005, **103**, 963–981.
- 41 A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009–4037.
- 42 B. Levine, C. Ko, J. Quenneville and T. Martinez, *Mol. Phys.*, 2006, **104**, 1039–1051.
- 43 F. Cordova, L. J. Doriol, A. Ipatov, M. E. Casida, C. Filippi and A. Vela, *J. Chem. Phys.*, 164111–164118.
- 44 J. D. Coe and T. J. Martinez, *J. Phys. Chem. A.*, 2006, **110**, 618–630.
- 45 J. D. Coe and T. J. Martinez, *J. Am. Chem. Soc.*, 2005, **127**, 4560–4561.
- 46 O. Weingart, A. Migani, M. Olivucci, M. A. Robb, V. Buss and P. Hunt, *J. Phys. Chem. A.*, 2004, **108**, 4685–4693.
- 47 M. Boggio-Pasqua, M. J. Bearpark, P. A. Hunt and M. A. Robb, *J. Am. Chem. Soc.*, 2002, **124**, 1456–1470.
- 48 A. Sanchez-Galvez, P. Hunt, M. A. Robb, M. Olivucci, T. Vreven and H. B. Schlegel, *J. Am. Chem. Soc.*, 2000, **122**, 2911–2924.
- 49 H. J. Werner and W. Meyer, *J. Chem. Phys.*, 1981, **74**, 5794–5801.
- 50 G. A. Worth and M. A. Robb, *Adv. Chem. Phys.*, 2002, **124**, 355–431.
- 51 B. O. Roos, *Acc. Chem. Res.*, 1999, **32**, 137–144.
- 52 H. J. C. Berendsen, *Science*, 2001, **271**, 954–955.
- 53 F. Jensen, *Introduction to Computational Chemistry*, Wiley, New York, 2001.
- 54 B. T. Thole, *Chem. Phys.*, 1981, **59**, 341–350.
- 55 G. Lamoureux and B. Roux, *J. Chem. Phys.*, 2003, **119**, 3025–3039.
- 56 A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.*, 1991, **95**, 3358–3363.
- 57 W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and user guide*, Vdf Hochschulverlag, Zürich, 1996.
- 58 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, *J. Comp. Chem.*, 2003, **24**, 1999–2012.
- 59 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.
- 60 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
- 61 B. H. Besler, K. M. Merz and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431–439.
- 62 M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700–733.
- 63 P. Amara and M. J. Field, *Theor. Chem. Acc.*, 2003, **109**, 43–52.
- 64 T. Wesolowski and A. Warshel, *J. Phys. Chem.*, 1993, **97**, 8050–8053.
- 65 T. Wesolowski and J. Weber, *Chem. Phys. Lett.*, 1996, **248**, 71–76.
- 66 T. Wesolowski, R. Muller and A. Warshel, *J. Phys. Chem.*, 1996, **100**, 15444–15449.
- 67 X. Assfeld and J. L. Rivail, *Chem. Phys. Lett.*, 1996, **263**, 100–106.
- 68 D. M. Philipp and R. A. Friesner, *J. Comput. Chem.*, 1999, **20**, 1468–1494.
- 69 J. L. Gao, P. Amara, C. Alhambra and M. Field, *J. Phys. Chem. A*, 1998, **102**, 4714–4721.
- 70 R. M. Nicoll, S. A. Hindle, G. MacKenzie, I. H. Hillier and N. A. Burton, *Theor. Chem. Acc.*, 2001, **106**, 105–112.
- 71 A. Rodriguez, C. Oliva, M. Gonzalez, M. van der Kamp and A. J. Mulholland, *J. Phys. Chem. B*, 2007, **111**, 12909–12915.
- 72 J. M. Olsen, K. Aidan and J. Kongsted, *J. Chem. Theor. Comp.*, 2010, **6**, 3721–3734.
- 73 T. Schwabe, J. M. H. Olsen, K. Sneskov, J. Kongsted and O. Christiansen, *J. Chem. Theor. Comp.*, 2011, **7**, 2209–2217.
- 74 K. Sneskov, T. Schwabe, O. Christiansen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2011, **13**, 18551–18560.
- 75 T. Rocha-Rinza, K. Sneskov, O. Christiansen, U. Ryde and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1585–1589.
- 76 K. J. Hellingwerf, J. Hendriks and T. Gensch, *J. Phys. Chem. A.*, 2003, **107**, 1082–1094.
- 77 W. Ryan, D. Gordon and D. Levy, *J. Am. Chem. Soc.*, 2002, **124**, 6194–6201.
- 78 M. de Groot and W. Buma, *J. Phys. Chem. A*, 2005, **109**, 6135–6136.
- 79 S. Smolarek, A. Vdovin, D. L. Perrier, J. P. Smit, M. Drabbel and W. J. Buma, *J. Am. Chem. Soc.*, 2010, **132**, 6315–6317.
- 80 C. Ko, B. Levine, A. Toniolo, L. Manohar, S. Olsen, H. Werner and T. Martinez, *J. Am. Chem. Soc.*, 2003, **125**, 12710–12711.
- 81 Q. Li and W. Fang, *Chem. Phys.*, 2005, **313**, 71–75.
- 82 E. V. Gromov, I. Burghardt, J. T. Hynes, H. Köppel and L. S. Cederbaum, *J. Photochem. Photobiol. A*, 2007, **190**, 241–257.
- 83 T. Rocha-Rinza, O. Christiansen, J. Rajput, A. Gopalan, D. B. Rahbek, L. H. Andersen, A. V. Bochenkova, A. A. Granovsky, K. B. Bravaya, A. V. Nemukhin, K. L. Christiansen and M. B. Nielsen, *J. Phys. Chem. A*, 2009, **113**, 9442–9449.
- 84 J. Rajput, D. B. Rahbek, G. Aravind and L. H. Andersen, *Biophys. J.*, 2010, **98**, 488–492.
- 85 I. Nielsen, S. Boye-Peronne, M. El Ghazaly, M. Kristensen, S. Nielsen and L. Andersen, *Biophys. J.*, 2005, **89**, 2597–2604.
- 86 L. Lammich, J. Rajput and L. H. Andersen, *PHYSICAL REVIEW E*, 2008, **78**, year.
- 87 D. Zuev, K. B. Bravaya, T. D. Crawford, R. Lindh and A. I. Krylov, *J. Chem. Phys.*, 2011, **134**, year.
- 88 A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko and T. J. Martinez, *J. Chem. Phys. B*, 2009, **113**, 3280–3291.
- 89 E. Fabiano, G. Groenhof and W. Thiel, *Chem. Phys.*, 2008, **351**, 111–116.
- 90 C. Ko, A. M. Virshup and T. J. Martinez, *Chem. Phys. Lett.*, 2008, **460**, 272–277.
- 91 M. Boggio-Pasqua, M. A. Robb and G. Groenhof, *J. Am. Chem. Soc.*, 2009, **131**, 13580–13581.
- 92 I. Lee, W. Lee and A. Zewail, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 258–262.
- 93 A. Espagne, D. H. Paik, P. Chagenet-Barret, P. Plaza, M. M. Martin and A. H. Zewail, *Photochem. Photobiol. Sci.*, 2007, **6**, 780–787.
- 94 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
- 95 O. Weingart, *Chem. Phys.*, 2008, **349**, 348–355.
- 96 O. Weingart, P. Altoe, M. Stenta, A. Bottone, G. Orlandi and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2011, **13**, 3645–3648.

- 97 I. Schapiro, M. N. Ryazantsev, L. Manuel Frutos, N. Ferre, R. Lindh and M. Olivucci, *J. Am. Chem. Soc.*, 2011, **133**, 3354–3364.
- 98 R. C. Haddon and L. T. Scott, *Pure Appl. Chem.*, 1986, **58**, 137–142.
- 99 D. Larsen, M. Vengris, I. van Stokkum, M. van der Horst, F. de Weerd, K. Hellingwerf and R. van Grondelle, *Biophys. J.*, 2004, **86**, 2538–2550.
- 100 M. Vengris, M. van der Horst, G. Zgrablje, I. van Stokkum, S. Haacke, M. Chergui, K. Hellingwerf, R. van Grondelle and D. Larsen, *Biophys. J.*, 2004, **87**, 1848–1857.
- 101 M. Vengris, D. Larsen, M. van der Horst, O. Larsen, K. Hellingwerf and R. van Grondelle, *J. Phys. Chem. B*, 2005, **109**, 4197–4208.
- 102 G. Groenhof, L. V. Schaefer, M. Boggio-Pasqua, H. Grubmueller and M. A. Robb, *J. Am. Chem. Soc.*, 2008, **130**, 3250–3251.
- 103 M. Baca, G. E. O. Borgstahll, M. Boissinot, P. M. Burke, D. R. Williams, K. A. Slater and E. D. Getzoff, *Biochemistry*, 1994, **33**, 14369–14377.
- 104 E. V. Gromov, I. Burghardt, H. KoeppeI and L. S. Cederbaum, *J. Am. Chem. Soc.*, 2007, **129**, 6798–6806.
- 105 P. B. Coto, D. Roca-Sanjuan, L. Serrano-Andres, A. Martin-Pendas, S. Marti and J. Andres, *J. Chem. Theor. Comp.*, 2009, **5**, 3032–3038.
- 106 K. Heyne, O. F. Mohammed, A. Usman, J. Dreyer, E. T. J. Nibbering and M. A. Cusanovich, *J. Am. Chem. Soc.*, 2005, **127**, 18100–18106.
- 107 L. J. G. W. van Wilderen, M. A. van der Horst, I. H. M. van Stokkum, K. J. Hellingwerf, R. van Grondelle and M. L. Groot, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 15050–15055.
- 108 H. Ihee, S. Rajagopal, V. Srainer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff and K. Moffat, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 7145–7150.
- 109 A. Yamada, T. Ishikura and T. Yamato, *Proteins*, 2004, **55**, 1063–1069.
- 110 N. Shimizu, H. Kamikubo, Y. Yamazaki, Y. Imamoto and M. Kataoka, *Biochemistry*, 2006, **45**, 3542–3547.
- 111 P. Changenet-Barret, P. Plaza, M. M. Martin, H. Chosrowjan, S. Taniguchi, N. Mataga, Y. Imamoto and M. Kataoka, *Chem. Phys. Lett.*, 2007, **434**, 320–325.
- 112 K. Takeshita, Y. Imamoto, M. Kataoka, K. Mihara, F. Tokunaga and M. Terazima, *Biophys. J.*, 2002, **83**, 1567–1577.
- 113 R. Cordfunke, R. Kort, A. Pierik, B. Gobets, G. J. Koomen, J. W. Verhoeven and K. J. Hellingwerf, *Proc. Natl. Acad. Sci. USA*.
- 114 A. H. Xie, W. D. Hoff, A. R. Kroon and K. J. Hellingwerf, *Biochemistry*, 1996, **35**, 14671–14678.
- 115 J. Hendriks, W. D. Hoff, W. Crielaard and K. J. Hellingwerf, *J. Biol. Chem.*, 1999, **274**, 17655–17660.
- 116 M. Yoda, Y. Inoue and M. Sakurai, *J. Phys. Chem. B*, 2003, **107**, 14569–14575.
- 117 M. Kamiya, S. Saito and I. Ohmine, *J. Phys. Chem. B*, 2007, **111**, 2948–2956.
- 118 J. Vreede, W. Crielaard, K. Hellingwerf and P. Bolhuis, *Biophys. J.*, 2005, **88**, 3525–3535.
- 119 J. Vreede, J. Juraszek and P. G. Bolhuis, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 2397–2402.
- 120 E. J. M. Leenders, L. Guidoni, U. Roethlisberger, J. Vreede, P. G. Bolhuis and E. J. Meijer, *J. Phys. Chem. B*, 2007, **111**, 3765–3773.
- 121 P. L. Ramachandran, J. E. Lovett, P. J. Carl, M. Cammarata, J. H. Lee, Y. O. Jung, H. Ihee, C. R. Timmel and J. J. van Thor, *J. Am. Chem. Soc.*, 2011, **133**, 9395–9404.
- 122 D. S. Larsen, I. H. M. van Stokkum, M. Vengris, M. A. van der Horst, F. L. de Weerd, K. J. Hellingwerf and R. van Grondelle, *Biophys. J.*, 2004, **87**, 1858–1872.
- 123 R. Kort, K. J. Hellingwerf and R. B. G. Ravelli, *J. Biol. Chem.*, 2004, **279**, 26417–26424.
- 124 N. Mataga, H. Chosrowjan, Y. Shibata, Y. Imamoto and F. Tokunaga, *J. Phys. Chem. B*, 2000, **104**, 5191–5199.
- 125 N. Mataga, H. Chosrowjan and S. Taniguchi, *j. Photochem. Photobiol.* C, 2004, **5**, 155–168.
- 126 V. Molina and M. Merchan, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 4299–4304.
- 127 E. V. Gromov, I. Burghardt, H. KoeppeI and L. S. Cederbaum, *J. Phys. Chem. A*, 2011, **115**, 9237–9248.
- 128 S. Anderson, S. Crosson and K. Moffat, *Act. Cryst. D*, 2004, **60**, 1008–1016.
- 129 S. Z. Fisher, S. Anderson, R. Henning, K. Moffat, P. Langan, P. Thiagarajan and A. J. Schultz, *Act. Cryst. D*, 2007, **63**, 1178–1184.
- 130 S. Yamaguchi, H. Kamikubo, K. Kurihara, R. Kuroki, N. Niimura, N. Shimizu, Y. Yamazaki and M. Kataoka, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 440–444.
- 131 C. Hattig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154–5161.
- 132 Gromov, Evgeniy V. and Burghardt, Irene and KoeppeI, Horst and Cederbaum, Lorenz S., *Photochem. Photobiol. A: Chem.*, 2012.
- 133 T. Wesolowski, *J. Chem. Phys.*, 1997, **106**, 8516–8526.
- 134 P. A. Sigala, M. A. Tsuchida and D. Herschlag, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 9232–9237.
- 135 W. W. Cleland and M. M. Kreevoy, *Science*, 1994, **264**, 1887–1890.
- 136 A. Warshel and A. Papazyan, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 13665–13670.
- 137 C. Schutz and A. Warshel, *Proteins*, 2004, **55**, 711–723.
- 138 Saito K. and Ishikita H., *Proc. Natl. Acad. Sci. USA*, 2012.
- 139 A. Warshel and R. Weiss, *J. Am. Chem. Soc.*, 1980, **102**, 6218–6226.
- 140 S. Braun-Sand, P. K. Sharma, Z. T. Chu, A. V. Pisliakov and A. Warshel, *Biochim. Biophys. Acta-Bioenergetics*, 2008, **1777**, 441–452.
- 141 A. D. Stahl, M. Hospes, K. Singhal, I. van Stokkum, R. van Grondelle, M. L. Groot and K. J. Hellingwerf, *Biophys. J.*, 2011, **101**, 1184–1192.
- 142 W. R. Browne and B. L. Feringa, *Nature Nanotech.*, 2006, **1**, 25–35.