6

Computer Simulations of Photobiological Processes: The Effect of the Protein Environment

Gerrit Groenhof[a], Martial Boggio-Pasqua[b], Lars V. Schäfer[c] and Michael A. Robb[d]

Contents

[a] Computational Biomolecular Chemistry, Department of Computational and Theoretical Biophysics, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37077, Göttingen, Germany
[b] Laboratoire de Chimie et Physique Quantiques, IRSAMC, CNRS et Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse, France
[c] Department of Biophysical Chemistry, University of Groningen, Nijenborg 4, 9747 AG Groningen, The Netherlands
[d] Department of Chemistry, Imperial College London, London SW7 2AZ, UK

1. INTRODUCTION

Interaction between biology, physics, and chemistry is presently providing a window into an exciting new era of bioinspired nanotechnology. In particular, photobiological processes, like vision or photosynthesis, in which sunlight is used as the source of energy to bring about a chemical reaction, provide valuable templates to create tools for biomolecular imaging, information technology, and renewable energy. Mimicking photobiological processes requires a complete understanding of the underlying molecular dynamics (MD). As the relevant time and spatial resolution are notoriously hard to access experimentally, computer simulations are the methods of choice to deepen our understanding of how proteins have evolved to mediate photochemical reactions and to use these insights to create devices that mimic biological functions. In this chapter, we present the approach we use to perform excited state MD simulations of photo-induced processes in biological systems. We start by reviewing the theoretical concepts of photochemical reactions. We then discuss how we have used these concepts to create a practical simulation methodology. We conclude this chapter with a short review of selected applications on photobiological systems. These simulations not only reveal the detailed sequence of events that follow photon absorption, but also demonstrate how the biological environment controls the excited state dynamics.

2. THEORY

The size and complexity of a typical photobiological system, together with the timescales that must be reached, necessitate the use of classical 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 (QM) description is required for those parts of the system that are involved in photon absorption. For the remainder, a simple molecular mechanics (MM) forcefield model suffices. The interactions in the systems are thus computed within a hybrid QM/MM framework.

To model the dynamics of a photoactivated process, the ground and excited state potential energy surfaces must be described accurately. As we show schematically in Figure 6.1, a photochemical reaction starts in the excited state (S_1) but ends in the ground state (S_0) after radiationless decay via the conical intersection seam. To model the deactivation process, we use a diabatic surface-hopping algorithm in our MD simulations that allows the trajectory to hop between the surfaces when the intersection seam is reached.
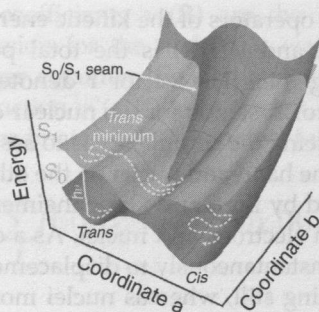
Figure 6.1 Schematic overview of a photochemical reaction pathway (dashed line). After photon absorption, evolution takes place on the excited state potential energy surface (red) until the system hits the S_1/S_0 intersection seam. At the intersection, a radiationless transition to the ground state occurs (blue). After the decay, the system continues evolving in the ground state. Please refer online version for color image.

2.1. Born–Oppenheimer approximation

The aim in computational chemistry is to find and interpret the solution for the many-body Schrödinger equation of chemical systems:

$$H\Psi = E\Psi, \tag{1}$$

with H the system's Hamilton operator, or Hamiltonian, that returns the total energy E of the system when operating on the many-body wavefunction Ψ. As in classical mechanics, the Hamiltonian is defined as the sum of the kinetic T and potential energy V:

$$H = T + V. \tag{2}$$

From the wavefunction Ψ, all static properties of the system can be derived.

Dynamic information is obtained by integrating the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t}\Psi = H\Psi, \tag{3}$$

with $\hbar$ the Planck constant divided by 2π. Unfortunately, an exact solution exists only if there are at most two interacting particles in the system. Solving the equations for any larger system requires approximations.

Since biological molecules are mainly composed of first and second row elements, the electronic velocities are sufficiently low for relativistic effects to be ignored. Within this approximation, the nonrelativistic Hamiltonian is given by

$$H = T_N + T_e + U(\mathbf{r}, \mathbf{R}), \tag{4}$$

where T_e and T_N are the operators of the kinetic energy of the electrons and the nuclei, respectively, and $U(\mathbf{r}, \mathbf{R})$ is the total potential energy of the electrons and nuclei together. The vector $\mathbf{r}$ denotes the set of electronic coordinates and the vector $\mathbf{R}$ stands for the nuclear coordinates.

The next step in reducing the complexity is to assume that the dynamics of the electrons on the one hand and nuclei on the other are decoupled. This approximation, proposed by Born and Oppenheimer, is based on the large mass difference between electrons and nuclei. As a consequence, the much lighter electrons adapt instantaneously to displacements of the nuclei. Electrons "see" nuclei standing still, whereas nuclei move on potential energy landscapes created instantly by the faster electrons. Within the Born–Oppenheimer approximation, the electronic and nuclear degrees of freedom can thus be treated independently.

First, the Schrödinger equation is solved for electrons moving in a framework of fixed nuclei. Thus, the nuclear kinetic energy operator (T_N) is omitted from the full Hamiltonian [Eq. (4)] to yield

$$H^e = T_e + U(\mathbf{r}, \mathbf{R}), \tag{5}$$

where the superscript e indicates the *electronic* Hamiltonian. The electronic wavefunctions are the eigenfunctions of this Hamiltonian

$$H^e \psi_i(\mathbf{r}; \mathbf{R}) = V_i(\mathbf{R}) \psi_i(\mathbf{r}; \mathbf{R}), \tag{6}$$

with $\psi_i(\mathbf{r};\mathbf{R})$ and $V_i(\mathbf{R})$ as the electronic wavefunctions and their electronic energies, respectively, that both depend *parametrically* on the nuclear coordinates. The wavefunctions are the adiabatic electronic states, representing the electronic ground state ($i = 0$), excited state ($i = 1$), second excited state ($i = 2$), and so on. How these wavefunctions are obtained in practical computations is the subject of modern quantum chemistry [1] and is briefly discussed below in Section 2.5.

The adiabatic wavefunctions can be made *orthonormal*, that is,

$$\langle \psi_i | \psi_j \rangle = \int_{-\infty}^{\infty} \psi_i(\mathbf{r}, \mathbf{R})^* \psi_j(\mathbf{r}, \mathbf{R}) d\mathbf{r} = \delta_{ij}, \tag{7}$$

where the Dirac bracket notation ($\langle | \rangle$) has been introduced that will be used throughout the text and where δ_{ij} is the Kronecker delta function, which is 1 if $i = j$ and 0 otherwise.

Within the Born–Oppenheimer approximation, the adiabatic electronic states provide a complete basis to expand the molecular wavefunction in

$$\Psi(\mathbf{r}, \mathbf{R}) = \sum_n \chi_n(\mathbf{R}) \psi_n(\mathbf{r}; \mathbf{R}), \tag{8}$$

where the expansion coefficients $\chi_n(\mathbf{R})$ are the nuclear wavefunctions. Substituting the expression for the molecular wavefunction in the time-independent Schrödinger equation [Eq. (1)] and multiplying by the adiabatic electronic wavefunction $\psi_i^*(\mathbf{r}; \mathbf{R})$ from the left, followed by integration over the electronic coordinates $\mathbf{r}$ leads to the following set of coupled equations:

$$\sum_j H_{ij}(\mathbf{R})\chi_j(\mathbf{R}) = E(\mathbf{R})\chi_i(\mathbf{R}), \tag{9}$$

with

$$H_{ij}(\mathbf{R}) = \langle \psi_i(\mathbf{r}; \mathbf{R})|H|\psi_j(\mathbf{r}; \mathbf{R})\rangle = \langle \psi_i(\mathbf{r}; \mathbf{R})|T_N|\psi_j(\mathbf{r}; \mathbf{R})\rangle + V_i(\mathbf{R})\delta_{ij}. \tag{10}$$

The nuclear kinetic energy operator is defined as

$$T_N = -\sum_k \frac{\hbar^2}{2M_k}\nabla_{\mathbf{R}_k}^2, \tag{11}$$

in which M_k is the mass of nucleus k and the sum runs over all nuclei. Using this relation, Eq. (10) can be rewritten as follows [2]:

$$H_{ij}(\mathbf{R}) = [T_N + V_i(\mathbf{R})]\delta_{ij} - \Lambda_{ij}(\mathbf{R}). \tag{12}$$

The *nonadiabatic* operator elements $\Lambda_{ij}(\mathbf{R})$ are defined as

$$\Lambda_{ij}(\mathbf{R}) = \sum_k F_{ij}^k(\mathbf{R})\nabla_{\mathbf{R}_k} + G_{ij}(\mathbf{R}), \tag{13}$$

$$F_{ij}^k(\mathbf{R}) = \frac{\hbar^2}{M_k}\langle \psi_i(\mathbf{r}; \mathbf{R})|\nabla_{\mathbf{R}_k}|\psi_j(\mathbf{r}; \mathbf{R})\rangle, \tag{14}$$

$$G_{ij}(\mathbf{R}) = \sum_k \frac{\hbar^2}{2M_k}\langle \psi_i(\mathbf{r}; \mathbf{R})|\nabla_{\mathbf{R}_k}^2|\psi_j(\mathbf{r}; \mathbf{R})\rangle. \tag{15}$$

In contrast to H^e [Eq. (5)], in which T_N was omitted, H_{ij} is *not* diagonal on the basis of the adiabatic electronic wavefunctions (ψ_i). The individual electronic states are thus coupled *via* nuclear motions. Nonadiabatic coupling is the key player in a photochemical reaction, as we will show below.

In the limit that the electronic wavefunctions vary very slowly with the nuclear dynamics, the nonadiabatic operators $F_{ij}^k(\mathbf{R})$ and $G_{ij}(\mathbf{R})$ are vanishingly small and can be safely neglected. Thus, the crux of the Born–Oppenheimer approximation is that H_{ij} is assumed to be diagonal:

$$H_{ij}(\mathbf{R}) = [T_N + V_i(\mathbf{R})]\delta_{ij}. \tag{16}$$

Under this assumption, the total molecular wavefunction becomes a product of a *single* nuclear and a *single* electronic wavefunction:

$$\Psi_i^{tot}(\mathbf{R}, \mathbf{r}) = \chi_i(\mathbf{R})\psi_i(\mathbf{r}; \mathbf{R}), \tag{17}$$

which implies that the nuclei move on a *single* electronic potential energy surface $V_i(\mathbf{R})$ of a given electronic state i, and the electronic wavefunction remains in that state. The Born–Oppenheimer approximation not only reduces the computational complexity of the equations that have to be solved, but also provides a conceptually intuitive picture of molecular structure and dynamics.

2.2. Conical intersections

The Born–Oppenheimer is valid as long as the separation between the electronic energy levels is large compared to the separation between the vibrational energy levels. Since this is true for almost all ground state chemical processes, the Born–Oppenheimer provides the basis for modern quantum chemistry. For photochemical processes, however, this is usually not true.

During a photochemical reaction, the system samples regions of 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 will occur between nuclear vibrations and electronic transitions. The populations of the adiabatic wavefunctions become strongly dependent on the nuclear dynamics and the nonadiabatic coupling operator [Λ, Eq. (13)] can no longer be ignored. Thus, in regions of strong non-adiabatic coupling the Born–Oppenheimer approximation breaks down. Nuclear dynamics induces population transfer between different electronic states. Furthermore, 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.

In theory, *all* electronic states $\psi_i(\mathbf{r}; \mathbf{R})$ are involved in the nonadiabatic coupling. In practice, however, there is only significant coupling between electronic states that have comparable energies. Therefore, only a small number of states needs to be included in the nonadiabatic coupling matrix Λ, which considerably reduces its dimensionality.

The nonadiabatic coupling operators Λ_{ij} are nonlocal derivative operators that depend inversely on the energy gap between the coupled adiabatic states [3]:

$$F_{ij}^k(\mathbf{R}) = \frac{\hbar^2}{M_k}\langle\psi_i(\mathbf{r}; \mathbf{R})|\nabla_{\mathbf{R}_k}|\psi_j(\mathbf{r}; \mathbf{R})\rangle = \frac{\hbar^2}{M_k}\frac{\langle\psi_i(\mathbf{r}; \mathbf{R})|\nabla_{\mathbf{R}_k}H^e|\psi_j(\mathbf{r}; \mathbf{R})\rangle}{V_j - V_i}. \tag{18}$$

When the gap $(V_j - V_i)$ gets smaller, the coupling increases, and the nuclear wavefunction that is initially on one surface will spread to the other surface without losing energy. The coupling thus induces a radiationless transition between the electronic states.

If the gap vanishes, that is, $V_j - V_i = 0$, the coupling becomes infinite. Due to the nonlocal nature of the coupling matrix elements in the adiabatic representation, it is more convenient to switch to a diabatic representation, in which the nonadiabatic coupling operator is a local, potential-like operator. The switch to the diabatic wavefunctions φ is achieved by a unitary transformation of the adiabatic wavefunctions ψ at each point in space [3]:

$$\varphi = \mathbf{S}(\mathbf{R})\psi. \tag{19}$$

In the diabatic representation the complete Hamiltonian [Eq. (12)] becomes

$$H_{ij}(\mathbf{R}) = T_N\delta_{ij} + W_{ij}(\mathbf{R}), \tag{20}$$

and the molecular Schrödinger equation [Eq. (12)] can be written in matrix notation as

$$\mathbf{H}\chi = [T_N\mathbf{1} + \mathbf{W}(\mathbf{R})]\chi = \mathbf{E}\chi, \tag{21}$$

in which $\mathbf{1}$ is the identity matrix and $\mathbf{W}(\mathbf{R})$ is the diabatic potential energy matrix, which, in contrast to the adiabatic potential matrix $\mathbf{V}(\mathbf{R})$, contains only *local* terms.

To illustrate the concept of the surface crossing, we consider a molecule in which there is coupling between two diabatic states A and B, but not to any other state. We can expand the potential matrix elements in a Taylor series around an arbitrary point $\mathbf{R}_0$:

$$\mathbf{W}(\mathbf{R} - \mathbf{R}_0) = \mathbf{W}^{(0)} + \mathbf{W}^{(1)} + \mathbf{W}^{(2)} + \cdots, \tag{22}$$

At $\mathbf{R}_0$, we can choose the diabatic and adiabatic states to be equal. Then, the zeroth-order matrix, $\mathbf{W}^{(0)}$, is a diagonal matrix in which the elements correspond to the energies E_A and E_B of the diabatic states ψ_A and ψ_B at $\mathbf{R}_0$, which, by our choice of origin, are identical to the *adiabatic* energies V_1 and V_2:

$$\mathbf{W}^{(0)} = \frac{E_A + E_B}{2}\mathbf{1} + \begin{pmatrix} -\dfrac{E_B - E_A}{2} & 0 \\ 0 & \dfrac{E_B - E_A}{2} \end{pmatrix} = \mathbf{V}(\mathbf{R}_0). \tag{23}$$

For very small displacements ΔR around R_0, the Taylor expansion in Eq. (22) can be truncated after the first-order term

$$W^{(1)} = \frac{\lambda \cdot \Delta R}{2} \mathbf{1} + \begin{pmatrix} -\frac{1}{2} \delta\kappa \cdot \Delta R & \kappa^{AB} \cdot \Delta R \\ \kappa^{AB} \cdot \Delta R & \frac{1}{2} \delta\kappa \cdot \Delta R \end{pmatrix}, \tag{24}$$

in which we have introduced linear potential constants that are defined as

$$\begin{aligned} \delta\kappa &\equiv \nabla_R (E_B - E_A)|_{R_0}, \\ \kappa^{AB} &\equiv \nabla_R \langle \varphi_A | H^e | \varphi_B \rangle|_{R_0}, \\ \lambda &\equiv \nabla_R (E_A + E_B)|_{R_0}. \end{aligned} \tag{25}$$

We can choose R_0 to be the point of degeneracy, so that E_A and E_B are equal at R_0 and $W^{(0)}$ is 0. The adiabatic potential energy surfaces V_1 and V_2 are obtained by diagonalizing the diabatic potential matrix W. Thus, for the two-state system considered here,

$$V_{1,2} = \frac{1}{2} \lambda \cdot \Delta R \pm \frac{1}{2} \sqrt{[\delta\kappa \cdot \Delta R]^2 + 4[\kappa^{AB} \cdot \Delta R]}. \tag{26}$$

The necessary condition for a crossing between the two potential energy surfaces at R_0 is that the two electronic energies are identical. Therefore, two conditions need to be fulfilled simultaneously:

$$\begin{aligned} \delta\kappa \cdot \Delta R &= 0, \\ \kappa^{AB} \cdot \Delta R &= 0. \end{aligned} \tag{27}$$

Thus, to first order, the degeneracy is lifted in the two-dimensional space spanned by the vectors $\delta\kappa$ and κ^{AB}, which are the gradient difference vector and derivative coupling vector, respectively. The space spanned by these two vectors is often referred to as the branching space, or g–h plane. Furthermore, as is evident from Eq. (26) and illustrated in Figure 6.2, the topology of the surfaces is that of a double cone, with the point of degeneracy at the apex. 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 N internal degrees of freedom, the intersection space thus forms an $N - 2$-dimensional seam, each point of which is a conical intersection. If a molecule has less than two degrees of freedom, the two conditions for degeneracy [Eq. (27)] cannot be simultaneously fulfilled. In diatomic molecules, for example, electronic states of same symmetry cannot cross, which led Von Neumann and Wigner to propose their famous noncrossing rule [4].
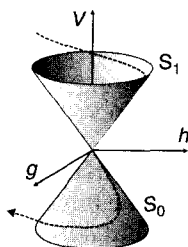
Figure 6.2 A conical intersection plotted in the two-dimensional branching space that is spanned by the gradient difference vector (g) and the derivative coupling vector (h). Motion away from the intersection in the g–h plane lifts the degeneracy between the two electronic states S_1 and S_0. The dotted line shows a path of a nuclear trajectory passing from one electronic state (S_1) to another (S_0) through the intersection funnel.

The concept of the intersection seam is illustrated in Figure 6.3 for a hypothetical triatomic molecule. In this molecule there are three internal degrees of freedom: two bond lengths (x_1 and x_2) and one angle (α). Since *two* degrees of freedom are required to span the branching space, there is only *one* degree of freedom available for the intersection seam. For the sake of simplicity, we let the branching space coordinates (i.e., the gradient difference and derivative coupling vectors) coincide with the two bond length variables x_1 and x_2 in our molecule. Note that in real triatomic molecules, the branching space coordinates can be linear combinations of the three internal degrees of freedom.
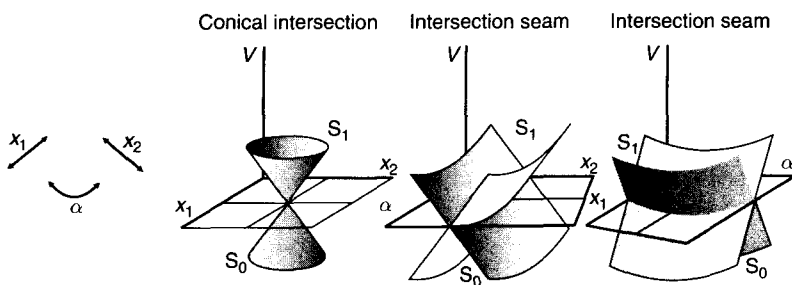

Figure 6.3 Surface crossing between two potential energy surfaces S_1 and S_0 in a hypothetical triatomic molecule (left). We let x_1 and x_2 be parallel to the gradient difference vector and the derivative coupling vector, respectively, and α be the remaining degree of freedom. When projected onto the branching space spanned by x_1 and x_2, the surfaces of S_0 and S_1 intersect in a single point, the conical intersection. In the two other subspaces, spanned by either x_2 and α, or by x_1 and α, there is an intersection line between the surfaces. In the second plot x_1 and in the third x_2 are assumed to be at their conical intersection coordinates.

In the first plot of Figure 6.3 the two adiabatic potential energy surfaces (denoted S_0 and S_1) are projected onto the branching space. The surfaces adopt a double cone shape with a single intersection point at the apex. Any displacement away from the apex in either the x_1 or x_2 coordinate lifts the degeneracy.

The second plot shows the two surfaces in the subspace spanned by the derivative coupling vector x_2 and the third independent degree of freedom, the bond angle α. The gradient difference vector x_1 is kept fixed at the conical intersection. Therefore, the only coordinate left to lift the degeneracy is x_2. If the latter is at the conical intersection as well, the surfaces intersect irrespective of the value of the angle variable (α). The intersection seam is thus a one-dimensional line that is parallel to α.

The third plot shows the projection of the two surfaces onto the subspace spanned by the gradient difference vector x_1 and the bond angle variable α. Now the derivative coupling vector x_2 is constrained to be at the conical intersection. Under this condition, only x_1 can lift the degeneracy and the surfaces intersect along a seam parallel to α.

The conical intersection has a number of peculiar characteristics. For instance, if we consider a small displacement $\mathbf{q}$ away from the cone tip in the branching plane, the diabatic energy matrix $U_{A,B}$ becomes

$$U_{A,B}(\mathbf{q}) = \begin{pmatrix} H_{AA}(\mathbf{q}) & H_{AB}(\mathbf{q}) \\ H_{AB}(\mathbf{q}) & H_{BB}(\mathbf{q}) \end{pmatrix} = S(\mathbf{q})\mathbf{1} + \begin{pmatrix} -\Delta H(\mathbf{q}) & H_{AB}(\mathbf{q}) \\ H_{AB}(\mathbf{q}) & \Delta H(\mathbf{q}) \end{pmatrix}, \quad (28)$$

with [see Eq. (24)]

$$\Delta H(\mathbf{q}) \quad = \quad \frac{H_{BB}(\mathbf{q}) - H_{AA}(\mathbf{q})}{2} \quad = \quad \delta\boldsymbol{\kappa} \cdot \mathbf{q},$$

$$H_{AB}(\mathbf{q}) \quad = \quad \boldsymbol{\kappa}^{AB} \cdot \mathbf{q}, \quad (29)$$

$$S(\mathbf{q}) \quad = \quad \frac{H_{BB}(\mathbf{q}) + H_{AA}(\mathbf{q}).}{2}$$

The matrix $U_{A,B}(\mathbf{q})$ is the two-state Hamiltonian matrix defined on the basis of the eigenvectors at the reference point $\mathbf{R}_0$, at which the diabatic and adiabatic bases are identical. The diabatic potential matrix $U_{A,B}(\mathbf{q})$ can be diagonalized by the rotation matrix $\mathbf{T}(\mathbf{q})$ [5], to obtain the adiabatic energies $V_{1,2}$

$$\mathbf{T(q)} = \begin{pmatrix} \cos[\theta(\mathbf{q})] & -\sin[\theta(\mathbf{q})] \\ \sin[\theta(\mathbf{q})] & \cos[\theta(\mathbf{q})] \end{pmatrix}. \tag{30}$$

From Eq. (29), the rotation angle $\theta(\mathbf{q})$ is defined as [5–10]:

$$\theta(\mathbf{q}) = \frac{1}{2}\arctan\left[\frac{2H_{AB}(\mathbf{q})}{\Delta H(\mathbf{q})}\right]$$

$$= \frac{1}{2}\arctan\left[\frac{2\boldsymbol{\kappa}^{AB}\cdot\mathbf{q}}{\delta\boldsymbol{\kappa}\cdot\mathbf{q}}\right]$$

$$= \frac{1}{2}\arctan\left[\frac{y}{x}\right], \tag{31}$$

where we have introduced the *scaled* coordinates x and y. Replacing these scaled coordinates by polar coordinates r and ϕ

$$r = \sqrt{x^2 + y^2},$$

$$\phi = \arctan\left[\frac{2y}{x}\right], \tag{32}$$

yields the following expression for the rotation angle:

$$\theta(H_{AB}, \ \Delta H) = \frac{\phi}{2}. \tag{33}$$

Thus, a simple relationship exists between the polar angle (ϕ), which defines a rotation around the apex of the cone in the branching space, and the *mixing* angle (θ) for the diabatic states (φ_A, φ_B). Since the extent of mixing depends *only* on the polar angle ϕ and *not* on the radius r, it is constant along any straight line that starts from the apex of the double cone.

Because of this relationship, the adiabatic wavefunction changes sign upon a complete rotation around the apex of the cone in the branching space. This is demonstrated by comparing the wavefunctions at $\phi = \phi_0$ and $\phi_0 + 2\pi$. Substituting Eq. (33) into the rotation matrix [Eq. (30)] gives the following expression for the adiabatic states:

$$\psi_1 = \cos\left[\frac{\phi_0}{2}\right]\varphi_A - \sin\left[\frac{\phi_0}{2}\right]\varphi_B \tag{34}$$

and

$$\psi_2 = \sin\left[\frac{\phi_0}{2}\right]\varphi_A + \cos\left[\frac{\phi_0}{2}\right]\varphi_B. \tag{35}$$

If we substitute $\phi = \phi_0 + 2\pi$, we see that the adiabatic wavefunction has indeed changed sign after the 360° rotation:

$$\begin{aligned}\psi_1 &= \cos\left[\frac{\phi_0 + 2\pi}{2}\right]\varphi_A - \sin\left[\frac{\phi_0 + 2\pi}{2}\right]\varphi_B \\ &= \sin\left[\frac{\phi_0}{2}\right]\varphi_A - \cos\left[\frac{\phi_0}{2}\right]\varphi_B \\ &= -\psi_1.\end{aligned} \tag{36}$$

Because single valuedness of the wavefunction is one of the basic postulates of quantum mechanics, this result implies that the conical intersection is a singularity. This singularity is a consequence of the separation between electronic and nuclear degrees of freedom, which, as we have seen, is not valid near a surface crossing. Thus, the singularity only exists for the adiabatic electronic wavefunction and must be remedied by the nuclear wavefunction, so that the total molecular wavefunction is a single-valued function. Furthermore, as illustrated schematically in Figure 6.4, rotation
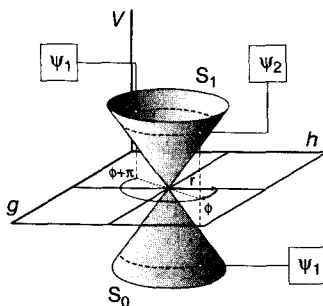

Figure 6.4 In the branching space (x,y), points lying on a circle centered at the apex of the double cone and with a small radius r are *diabatically* related. If we select a point on one of the surfaces and move it to the opposite side of the circle $(\phi \rightarrow \phi + \pi)$, its wavefunction becomes equal to the wavefunction of the other surface at the original position. If we complete the circle, the point is back at its original position, but its wavefunction has changed sign (not shown, see text).

by 90° permutes the order of the electronic states, as is demonstrated by substituting $\theta + \pi$ into Eq. (33):

$$
\begin{aligned}
\psi_1 &= \cos\left[\frac{\phi_0 + \pi}{2}\right]\varphi_A - \sin\left[\frac{\phi_0 + \pi}{2}\right]\varphi_B \\
&= \sin\left[\frac{\phi_0}{2}\right]\varphi_A - \cos\left[\frac{\phi_0}{2}\right]\varphi_B \\
&= -\psi_2.
\end{aligned}
\tag{37}
$$

As we will show in Section 2.3, this characteristic of the conical intersection can be used to detect the time step at which the seam is passed in a classical MD simulation of a photochemical reaction.

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 6.2). To illustrate the relationship between the surface crossing and photochemical reactivity, we draw a parallel with the transition state in ground state chemistry. The transition state forms the 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 also forms a bottleneck that separates the excited state branch of the reaction path from the ground state branch. The crucial difference between conical intersections and transition states 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 spike on the ground state energy surface and thus connects the excited state reactant to two or more products on the ground state *via* several reaction paths.

2.3. Excited state molecular dynamics

Photochemical reactions start in the excited state (S_1) but end in the ground state (S_0, Figure 6.1). To model the dynamics of such processes, we need a method to accurately compute the ground and excited state potential energy surfaces. In addition, we need an algorithm that models the radiationless transitions between the surfaces in a manner that is consistent with quantum mechanics.

If a sufficiently accurate description of the adiabatic Born–Oppenheimer potential energy surfaces is available, nuclear dynamics can be computed by numerically integrating either the time-dependent Schrödinger equation [Eq. (3)] or Newton's equations of motion. In the first case, quantum mechanics is used to follow the dynamics of nuclear wavepackets [$\chi_n(\mathbf{R})$,

Eq. (9)], evolving on the electronic potential energy surfaces. With this approach transitions between surfaces near or at conical intersections can be described correctly. A wavepacket traveling through or near an intersection spawns new packets on both surfaces. The transfer probability is controlled by the nonadiabatic coupling elements. At the conical intersection, the coupling has its maximal strength, resulting in an efficient transfer of the complete wavepacket to the lower surface.

A requirement for wavepacket dynamics is that the relevant potential energy surfaces have been computed beforehand. As the computational costs associated with precomputing these surfaces increases rapidly with the number of coordinates in the system, wavepacket dynamics is restricted to small isolated molecules. Alternatively, the multidimensional configuration space can be reduced to a lower dimensional subspace in which quantum dynamics is possible [11]. However, choosing suitable coordinates is difficult and requires averaging over the remaining degrees of freedom. Furthermore, the choice of the coordinates can strongly bias the outcome of a simulation. Therefore, for photobiological problems, the size and complexity of the systems involved severely limit the applicability of wavepacket simulations.

Alternatively, we can decide to ignore the QM character of the nuclei altogether and use Newton's equations of motion to compute MD trajectories. In terms of computational effort, classical MD is orders of magnitude more efficient than wavepacket dynamics and is therefore routinely used for computing the time evolution of large biomolecular systems. Classical mechanics has the additional advantage that the potential energy surface can be computed on the fly. Forces are evaluated for the geometry at time step t and used to compute the geometry at the next time step $t + \Delta t$. Thus, only at the configurations sampled by the classical trajectory, electronic structure calculations are required. For systems with many degrees of freedom, for which computing potential energy surfaces beforehand is not possible, the on-the-fly strategy is the only option for performing MD simulations.

Classical mechanics provides a computationally cheap alternative to wavepacket dynamics. However, because quantum effects are ignored, population transfer cannot occur and classical trajectories are restricted to a single potential energy surface. Therefore, in contrast to wavepacket dynamics, radiationless transitions do not take place spontaneously. Instead, a binary decision to jump to a different electronic surface must be made at every time step in a single trajectory. The criterion for switching between electronic states must result in a distribution of state populations over very many trajectories that reflects the populations of a full QM treatment. Furthermore, in contrast to wavepacket dynamics, classical trajectories do not capture coherence effects. When a nuclear wavepacket originally traveling on a single adiabatic potential energy surface

encounters a crossing, it typically produces two subpackets, or offsprings, one on each surface [12]. After such spawning event, the new wavepackets will still interfere with each other. However, as these subpackets travel on different surfaces and often have different initial velocities on their respective surfaces, they are well separated in phase space and it is highly unlikely that they will encounter each other again, especially in high-dimensional systems, such as solvated biomolecules. Therefore, the spawned wavepackets are effectively incoherent [12,13], and the classical trajectory approximation is justified in most situations.

2.4. Diabatic surface hopping

The interest in understanding photochemical processes has prompted the development of methods for the treatment of nonadiabatic effects in classical MD. Most, if not all of these methods are based on surface hopping: nuclei move on a single potential energy surface and nonadiabatic transitions are included by allowing the trajectory to hop from one surface to another. The size and complexity of biomolecular systems necessitate the use of computationally cheap surface-hopping algorithms. In this section we present the hopping procedure that we use in our simulations of photochemical processes in biological systems.

Our so-called diabatic hopping algorithm is based on the one-dimensional Landau and Zener equation, which relates the probability of a transition between two electronic states ψ_2 and ψ_1, to the nonadiabatic coupling, via:

$$P_{2 \to 1} = \exp\left[-\frac{1}{4}\pi\xi\right]. \qquad (38)$$

In this equation ξ is the Massey parameter defined as [14]

$$\xi = \frac{\Delta E}{\hbar \frac{\partial \mathbf{Q}}{\partial t} \cdot \mathbf{g}(\mathbf{Q})}, \qquad (39)$$

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

$$\mathbf{g}(\mathbf{Q}) = \langle \psi_1 | \nabla_{\mathbf{Q}} \psi_2 \rangle. \qquad (40)$$

If we differentiate ψ_2 with respect to t via $\frac{\partial}{\partial t}(\mathbf{Q})$, we can rewrite ξ as

$$\xi = \frac{\Delta E}{\hbar \langle \psi_1 | \frac{\partial \psi_2}{\partial t} \rangle}. \qquad (41)$$

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 of the simulation. In practice, however, it is possible to approximate $\langle\psi_1|\frac{\partial}{\partial t}\psi_2\rangle$ as follows.

If we integrate the classical equations of motion for the nuclei with a small time step Δt, we have at t:

$$
\begin{aligned}
\psi_1(t) &= \varphi_A, \\
\psi_2(t) &= \varphi_B,
\end{aligned}
\tag{42}
$$

where the $\varphi_{A,B}$ are the diabatic electronic states. At $t+\Delta t$, the states will have become mixed due to the nonadiabatic coupling. In the limit of an infinitesimally small time step, the change in the wavefunctions can be approximated by

$$
\begin{aligned}
\psi_1(t+\Delta t) &= \varphi_A + \beta\varphi_B, \\
\psi_2(t+\Delta t) &= -\beta\varphi_A + \varphi_B,
\end{aligned}
\tag{43}
$$

where β is a mixing coefficient. Numerical differentiation (finite differences) of the wavefunction gives

$$
\frac{\partial}{\partial t}\psi_2 = -\beta\frac{\varphi_A}{\Delta t},
\tag{44}
$$

and yields the following expression:

$$
\left\langle \psi_1 \left| \frac{\partial}{\partial t}\psi_2 \right. \right\rangle \approx -\frac{\beta}{\Delta t}.
\tag{45}
$$

Since

$$
\langle\psi_1(t)|\psi_2(t+\Delta t)\rangle = -\beta,
\tag{46}
$$

we can compute $\langle\psi_1(t)|\psi_2(t+\Delta t)\rangle$ as a numerical approximation for $\langle\psi_1|\frac{\partial}{\partial t}\psi_2\rangle$ in the Massey parameter [Eq. (41)].

Calculating the energy gap ΔE and $\langle\psi_1(t)|\psi_2(t+\Delta t)\rangle$ at every time step is straightforward, and we can use the Landau–Zener formula [Eq. (38)] 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 $\Delta E \approx 0$ and $\|\langle\psi_1(t)|\psi_2(t+\Delta t)\rangle\| \approx 1$. The former is trivially true by the definition of an intersection. The latter follows from Eq. (37) and is illustrated in the branching space projection of Figure 6.4: passing through the intersection during a single time step (Δt), leads to geometries at t and $t+\Delta t$ that are opposite to each other with respect to the apex. Thus, crossing the

seam is equivalent to rotating the wavefunction by a half of a circle in the branching space (Figure 6.4):

$$\langle \psi_1(t)|\psi_2(t+\Delta t)\rangle = \langle \psi_1(\phi)|\psi_2(\phi+\pi)\rangle$$

$$= -\langle \psi_1(\phi)|\psi_1(\phi)\rangle$$

$$= -\langle \psi_1(t)|\psi_1(t)\rangle$$

$$= -1. \tag{47}$$

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 nonadiabatic 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. The Landau–Zener model is clearly an approximation but helps one to keep a proper physical insight, which is crucial in understanding complex systems.

2.5. Excited state quantum chemistry

Although highly accurate methods for computing excited state electronic wavefunctions 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 MD simulations of large biomolecular systems. Alternatively, simple forcefields or existing semi empirical methods that are computationally efficient may be used, but their applicability is limited, unless properly reparameterized [15,16]. Therefore, for on-the-fly MD, a compromise between cost and accuracy has to be made.

A computationally feasible approach to describe excited state electronic structure is the equation of motion coupled cluster (EOM-CCSD) method and ab initio dynamics simulations of small isolated molecules have been performed at the EOM-CCSD level [17]. However, EOM-CCSD can only work well if the underlying CCSD method provides a good description of the ground electronic state. This can cause problems when bonds are being broken or formed and the ground electronic state has a significant multi-configurational character. The description of electronic states with a strong double excitation character also causes problems [18].

Time-dependent density functional theory (TD-DFT) also offers a computationally very efficient approach to describe excited states and consequently has been used in excited state MD simulations [19]. Similar to

EOM-CCSD, TD-DFT suffers from the deficiencies of the underlying mono-configurational DFT 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 π systems, doubly excited states, and charge transfer excited states [20].

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, multiconfigurational methods are required to provide wavefunctions that are sufficiently flexible to describe bond rearrangements, electronic state mixing, and electronic reorganizations. In addition, to calculate MD trajectories, analytical energy gradients are necessary. Since the complete active space self-consistent field (CASSCF) method fulfills these requirements, it has often been used in the framework of excited state dynamics simulations [21–25]. In CASSCF, a judicious set of occupied and virtual orbitals is chosen, the so-called active space orbitals. 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 active 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-averaged approach), if state bias or root-flipping problems [26] have to be avoided that occur near surface crossing regions [27].

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. Because in most cases CASSCF describes the topology of the potential energy surfaces of the involved states sufficiently accurate, it is widely used for mechanistic studies of photochemical reactions [28].

Dynamic correlation is accounted for in multireference perturbation theory approaches, such as CASPT2 [29]. CASPT2 provides a means of including dynamic correlation, while simultaneously describing static correlation. Recently, analytical CASPT2 energy gradients [30] have become available, which has opened the way for MD simulations. Coe et al. have already used these gradients to perform an ab initio MD simulation on the excited state proton transfer reaction in methyl salicylate [31]. However, the computational cost of this method still prevents its use for larger biomolecular systems, in which the number of correlated electrons is too large.

Alternatively, dynamic electron correlation is included to some extent in the restricted active space self-consistent field (RASSCF) method [29]. RASSCF allows larger active spaces and thus provides more flexibility in the choice of the active orbitals than CASSCF [32]. The larger active space is subdivided into three classes: orbitals with a limited number of valences, a fully active orbital set, and orbitals with a limited number of electrons. By eliminating the redundant configurations, the size of the configuration interaction problem can be greatly reduced in RASSCF compared to CASSCF without compromising accuracy. Because analytical gradients are available, the RASSCF method can be expected to be used in excited state dynamics simulations in the future. Unfortunately, however, it remains a difficult task even in RASSCF to select the correct orbital set for a given photochemical problem.

A promising alternative that circumvents the choice of active orbitals lies in the use of semi empirical configuration interaction methods. These methods offer a lower cost alternative, while still taking into account the correlation effects necessary to describe excited states. For example, the recently developed semi empirical OM2 method has been shown to describe accurately the well-known conical intersections of small molecules [33]. Thus, the use of new semiempirical methods, such as OM2, could hold great promise for nonadiabatic MD simulations of very large molecular systems in the near future.

Due to inaccuracy of approaches that are based on a single reference configuration, lack of validation of the semi empirical alternatives, and prohibitive computational demands of multireference perturbation methods, CASSCF is at present the most attractive option for excited state MD simulations. However, 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 minimal active spaces. These minimal spaces 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 energies at these geometries.

2.6. Mixed quantum classical molecular dynamics

MD computer simulations of biological systems have come of age. Since the first application of MD on a small protein in vacuum more than three decades ago [34], 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 [35]. In the interaction functions, the so-called

MM forcefield, simple chemical concepts are used to describe the potential energy of the system [36]:

$$V_{MM} = \sum_{i}^{N_{bonds}} V_i^{bond} + \sum_{j}^{N_{angles}} V_j^{angle} + \sum_{l}^{N_{torsions}} V_l^{torsion}$$

$$\sum_{i}^{N_{MM}} \sum_{j>i}^{N_{MM}} V_{ij}^{Coul} + \sum_{i}^{N_{MM}} \sum_{j>i}^{N_{MM}} V_{ij}^{LJ}, \tag{48}$$

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

$$V_{ij}^{Coul} = \frac{e^2 Q_i Q_j}{4\pi\epsilon_0 R_{ij}}, \tag{49}$$

in which R_{ij} denotes the interatomic distance, e the unit charge, and ϵ_0 the electric constant. Short-range Pauli repulsion and long-range dispersion attraction are most often described by a single Lennard-Jones potential:

$$V_{ij}^{LJ} = \left(\frac{C_{12}^{ij}}{R_{ij}}\right)^{12} - \left(\frac{C_6^{ij}}{R_{ij}}\right)^6, \tag{50}$$

with C_{12}^{ij} and C_6^{ij} a repulsion and attraction parameter, respectively, that depend on the atom types involved.

Electrons are thus ignored in molecular mechanics forcefields. 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 mechanics 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 QM level, while retaining the computationally cheaper forcefield (MM) for the remainder. This hybrid QM/MM strategy was originally introduced by Warshel and Levitt [37] and is illustrated in Figure 6.5.

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
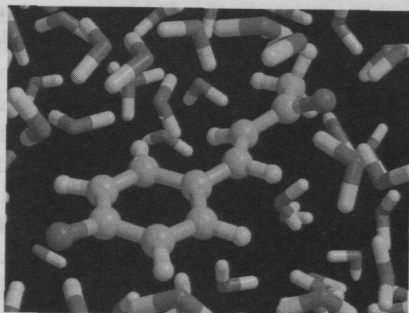
Figure 6.5   The QM/MM partitioning scheme used in recent simulations of a photoactive yellow protein chromophore analog in water. The atoms of the QM subsystem are shown in ball-and-stick representation and MM atoms are shown as thick sticks. The CASSCF method was employed to model the electronic structure of the chromophore, while the SPCE model [38] was used for the water molecules.

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 may or may not facilitate the reaction.

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

$$H^{hybrid} = H_{QM} + H_{MM} + H_{QM/MM}. \tag{51}$$

The interactions within the QM and MM regions are relatively straightforward to describe, that is, 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 forcefield interactions, that is, bonds, angles, torsions, and Lennard-Jones. 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 forcefield calculation is performed on the MM region. An improvement of the model is to use the isolated electronic wavefunction to derive partial atomic charges for the QM atoms and use these charges to compute the electrostatic QM/MM interactions with the MM atoms.

In the more popular electronic embedding scheme of Field and coworkers [39], the MM atoms enter the electronic Hamiltonian, as if they were QM nuclei:

$$H^e_{QM/MM} = H^e + \sum_i^{n_e} \sum_K^{N_{MM}} \frac{Q_i}{4\pi\epsilon_0 r_{iK}}$$

$$= -\frac{\hbar^2}{2m_e} \sum_i^{n_e} \nabla_i^2 + \sum_i^{n_e} \sum_{j>i}^{n_e} \frac{e^2}{4\pi\epsilon_0 r_{ij}}$$

$$-\sum_i^{n_e} \sum_A^{N_{QM}} \frac{e^2 Z_A}{4\pi\epsilon_0 r_{iA}} + \sum_K^{N_{MM}} \frac{e^2 Q_K}{4\pi\epsilon_0 r_{iK}}, \tag{52}$$

where H^e is the original electronic Hamiltonian for the isolated QM system, defined in Eq. (5); n_e is the number of electrons, N_{QM} the number of QM nuclei and N_{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 wavefunction can become overpolarized. Penetration of electron density into the MM region is an artefact 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 [40].

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

$$H^{nuc}_{QM/MM} = \sum_A^{N_{QM}} \sum_{B>A}^{N_{QM}} \frac{Z_A Z_B}{4\pi\epsilon_0 R_{AB}} + \sum_A^{N_{QM}} \sum_K^{N_{MM}} \frac{Z_A Q_K}{4\pi\epsilon_0 R_{AK}}. \tag{53}$$

In addition to electrostatics, there are also van der Waals interactions between the subsystems that are handled at the forcefield level [Eq. (50)], 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 forcefield functions.

If the QM and MM subsystems are connected by chemical bonds, care has to be taken when evaluating the QM wavefunction. 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 most easy solution is to introduce a monovalent link atom at an appropriate position along the bond vector between the QM and MM atoms. 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. Alternative approaches for capping the QM region are the frozen orbital approach [41] and the generalized hybrid orbital method [42]. In all of our simulation studies to date, we have used hydrogen link atoms for capping the QM subsystem.

The QM/MM scheme of Field and coworkers provides a conceptually intuitive way of including the effect of an environment on a chemical reaction. However, unless polarization is treated explicitly at the forcefield level, this QM/MM model is not internally consistent. In most forcefields polarization is not explicitly included, but is implicitly accounted for *via* the parameters in the other terms, most notably, in the nonbonded 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 such possible artifact, the nonbonded parameters in principle should be reparameterized 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.

3. APPLICATIONS

In this section we discuss some of our recent applications of excited state dynamics simulations on photobiological systems. We will show that in these systems, the protein environment controls the photochemical properties of the chromophore and steers the excited state dynamics.

3.1. Photoactive yellow protein

Photoactive yellow protein (PYP) is believed to be the primary photoreceptor for the photoavoidance response of the salt-tolerant bacterium *Halorhodospira halophila* [43]. PYP contains a deprotonated 4-hydroxy-cinnamic acid (or *p*-coumaric acid, pca) chromophore linked covalently to the γ-sulfur of Cys69 via a thioester bond (Figure 6.6). Upon absorbing a blue-light photon,
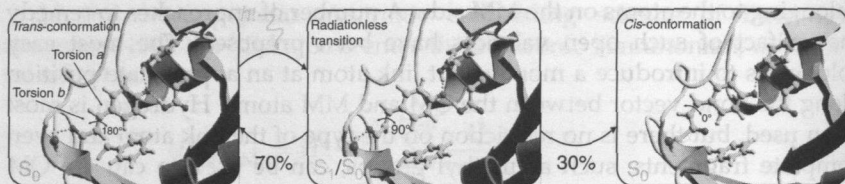
Figure 6.6 Snapshots from excited state trajectories of wild-type PYP showing the chromophore (pca) in the active site pocket. The first snapshot is at the excitation. The second shows the configuration at the radiationless transition from S_1 to S_0. The third snapshot shows the photoproduct, in which the carbonyl oxygen of the thioester linkage has flipped and is no longer hydrogen bonded to the backbone of Cys69.

PYP enters a fully reversible photocycle involving several intermediates on timescales ranging from a few hundred femtoseconds to seconds [43].

To understand the intrinsic photochemical properties of the PYP chromophore, we have performed geometry optimizations of an isolated chromophore analog [44]. In these optimizations, the complete π system of the chromophore was included in the active space, which thus consisted of 12 electrons in 11 π orbitals. In addition to optimizing the local minima on the S_1 potential energy surface and the barriers that separate them, we also located conical intersections in the vicinity of these mimima. The optimizations revealed that there are two minima on S_1: 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 at $90°$ (Figure 6.7). In the isolated chromophore, there is almost no barrier for reaching the single bond twisted S_1 minimum from the Franck–Condon region, whereas there is a significant barrier to double bond
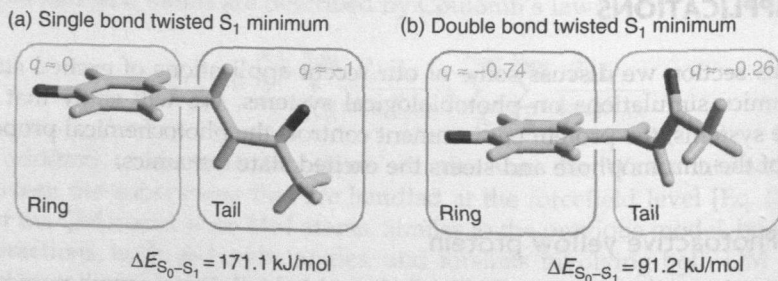

Figure 6.7 Excited state minimum energy configurations of a chromophore analog. In both the single bond twisted S_1 minimum (a) and the double bond twisted S_1 minimum (b), there is a substantial energy gap between the ground and excited states. The distribution of the negative charge in these minima is opposite.

rotation. Thus, after photon absorption in vacuum, the main relaxation channel on S_1 involves rotation of the single bond to 90°. We furthermore found that the S_1/S_0 intersection seam lies very far away from this minimum. As a consequence, radiationless decay is not very likely to occur in vacuum. In subsequent QM/MM simulations, we have probed the effect of different environments on the photochemistry of the chromophore.

To examine the effect of an aqueous environment, we have performed 91 QM/MM excited state dynamics simulations of a chromophore analog in water [44]. The chromophore was described at the CASSCF(6,6)/3-21G level of theory, while the water molecules were modeled by the SPCE forcefield [38]. The results of these simulations demonstrate that in water radiationless decay is very efficient [44]. 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 twisted S_1 minima, which reflects the difference in charge distribution between these minima [45]. In the single bond twisted S_1 minimum, the negative charge resides on the alkene moiety of the chromophore (Figure 6.7). Three strong hydrogen bonds to the carbonyl oxygen stabilize this charge distribution to such an extent that the seam almost coincides with the single bond twisted S_1 minimum (Figure 6.8). In the double bond twisted S_1 minimum, the negative charge is localized on the phenolate ring (Figure 6.7). Transient stabilization of this charge distribution by two or more strong hydrogen bonds to the phenolate oxygen brings the seam very close to this S_1 minimum (Figure 6.8). Thus, in water the ultrafast excited state decay is mediated by hydrogen bonds.
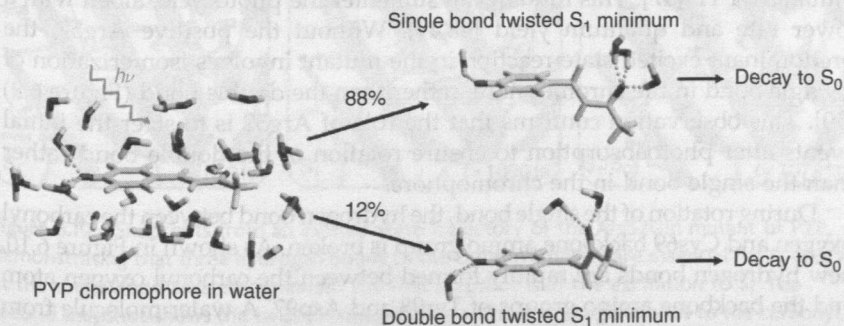

Figure 6.8 In water the chromophore undergoes both single and double bond isomerization. Excited state decay from these minima is very efficient due to stabilization of the chromophore's S_1 charge distribution by specific hydrogen bond interactions.

To find out how the protein mediates the photochemical process, we also carried out a series of QM/MM simulations of wild-type PYP. Figure 6.6 shows the primary events after photoexcitation in the simulation. The chromophore rapidly decays to the ground state *via* a 90° rotation of the double bond (Figure 6.6), rather than the single bond. During this photo-isomerization 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 enhance 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 180°. During this rotation, the hydrogen bond between the carbonyl oxygen and the Cys69 backbone amino group is broken (Figure 6.6). In total, 14 MD simulations were carried out, initiated from different snapshots from a classical ground state trajectory. The fluorescence lifetime (200 fs) and isomerization quantum yield (30%) in the simulations agree well with experiments (400 fs [46] and 35% [43], respectively).

In the wild-type protein no single bond isomerization was observed. Thus, the protein not only provides the hydrogen bonds required for ultra-fast decay, but also controls which of the chromophore's bond isomerizes upon photoexcitation. We identified the positive guanidinium moiety of Arg52 located just above the chromophore ring, as the "catalytic" residue that enforces double bond isomerization. The preferential electrostatic stabilization of the double bond twisted S_1 minimum by the positive Arg52 strongly favors double bond isomerization over single bond isomerization.

To elucidate the role of this arginine in the activation process in more detail, we performed excited state dynamics simulations on the Arg52Gln mutant of PYP [47]. This mutant can still enter the photocycle, albeit with a lower rate and quantum yield [48,49]. Without the positive Arg52, the predominant excited state reaction in the mutant involves isomerization of a single bond in the chromophore, rather than the double bond (Figure 6.9) [50]. This observation confirms that the role of Arg52 is to steer the initial events after photoabsorption to ensure rotation of the double bond rather than the single bond in the chromophore.

During rotation of the single bond, the hydrogen bond between the carbonyl oxygen and Cys69 backbone amino group is broken. As shown in Figure 6.10, 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. With these three hydrogen bonds stabilizing the negative charge on the alkene moiety, the chromophore rapidly decays to S_0. Thus, the decay mechanism in the Arg52Gln mutant and in water are essentially the same.
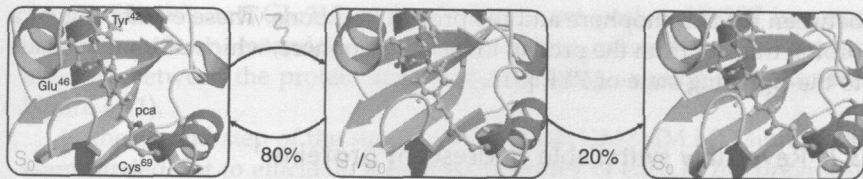
Figure 6.9 Snapshots from an excited state trajectory of the Arg52Gln mutant of PYP, showing the chromophore (pca) in the active site pocket. The first snapshot is at the excitation. The second shows the configuration at the radiationless transition from S_1 to S_0. The third snapshot shows the photoproduct. In 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.

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 was observed with a 20% quantum yield (Figure 6.9). Together with the experimental observation that the mutant has a photoactivation quantum yield of about 21% [49], this suggests that the key step to enter the photocycle is the oxygen flip rather than the double bond isomerization.

To summarize, the simulations are consistent with experimental observations and have provided detailed structural and dynamic information at a resolution well beyond that achievable by other means. From the simulations, key amino acids have been identified as well as the mechanism by which they control the primary events in the photocycle of PYP. These are (1) double bond photoisomerization and (2) the break of a hydrogen bond
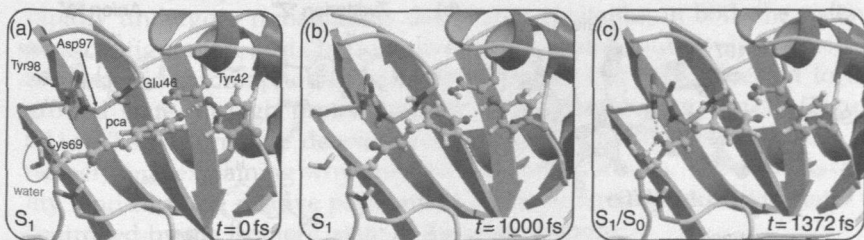

Figure 6.10 Snapshots from an excited state trajectory of the Arg52Gln mutant of PYP, demonstrating that three hydrogen bonds to the carbonyl moiety are essential for S_1 decay at the single bond twisted minimum. The first snapshot is at the excitation to S_1. The second snapshot shows the twisted configuration without hydrogen bonds to the carbonyl. The gap between S_1 and S_0 is far too high for decay at this configuration. However, as the third snapshot 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_1 minimum.

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 [51].

3.2. Reversibly switchable fluorescent proteins

Photochromic, or reversibly switchable fluorescent proteins (RSFPs) that can be photoswitched between a fluorescent and a nonfluorescent state have proven to be crucial for new innovative microscopy schemes. However, despite the availability of X-ray structures of fluorescent and nonfluorescent states of several RSFPs, there is as yet no consensus about how these protein achieve the switching. To reveal the molecular basis of the switching process we have carried out a two-step QM/MM study of asFP595, a natural occurring RSFP from the sea anemone *Anemonia sulcata* (Figure 6.11).

The first step was to locate the protons in the chromophore binding pocket, because these protons were not resolved in the available X-ray structures. For this purpose, simulated UV/Vis spectra were compared to the available experimental data. This comparison was backed up by continuum electrostatics calculations, and enabled to unambiguously determine the protonation of the *on* and *off* states of asFP595 [52]. These calculations ascertained that the neutral chromophore form is dominant in the *cis* conformation (*on*-state), whereas the zwitterionic and the anionic chromophores predominate in the *trans* conformation (*off*-state) (Figure 6.11b). These results reveal that the photoinduced *trans–cis* isomerization of the chromophore is accompanied by proton transfer events. As shown in Figure 6.11a, these proton transfers are mediated by the ionizable residues
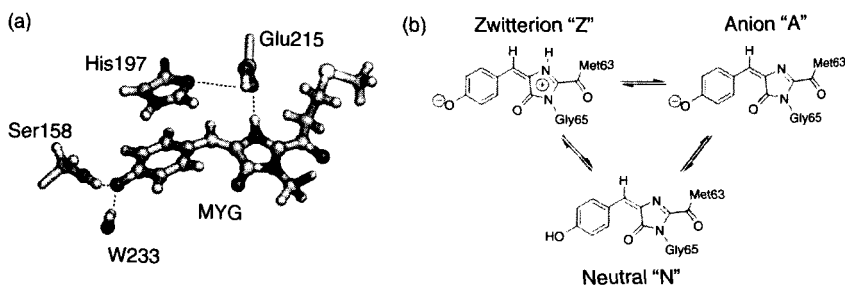

Figure 6.11 (a) Active site of the reversibly switchable fluorescent protein asFP595, showing the asFP595 chromophore (MYG) in the *trans* conformation and adjacent amino acid side chains. Dashed lines represent hydrogen bonds between MYG and Ser158, Glu215, and a crystallographic water molecule (W233), respectively. His197 is π stacked to the MYG phenoxy moiety and forms a hydrogen bond to Glu215. The carbon skeleton of the QM subsystem is shown in cyan, and the carbon atoms modeled by MM are shown in orange. (b) Schematic drawings of the important MYG protonation states.

Ser158, His197, and Glu215 in the chromophore pocket. His197 is conne-
cted through an extended hydrogen bonding network to the exterior
interface between the protein and the surrounding solution (not shown in
Figure 6.11).

In the second step of the study, nonadiabatic QM/MM MD simulations
were carried out to elucidate the photochemistry of each of the previously
identified protonation states of the chromophore [53]. In these simulations,
energies and gradients on the ground and excited states were calculated on
the fly at the CASSCF(6,6)/3-21G level of theory. Prior to the dynamics
simulations, minima and conical intersections were characterized in vacuo
at the RASSCF(18,7+4+5)[2,2]/6-31G* level of theory, that is, including all π
orbitals of the chromophore in the active space. The final six-electron, six-
orbital active space used in the QM/MM dynamics simulations was
selected from these RASSCF calculations such as to enable the simultaneous
description of the ground state and the first excited state.

The quantum yields and excited state lifetimes from the QM/MM
dynamics simulations agree well with experimental measurements [54]. In
addition, the simulations enable to predict the structures of the hitherto
unknown photochemical intermediates and the irreversibly fluorescent
state. Furthermore, the simulations revealed how the proton distribution
in the active site of the asFP595 controls the photochemical conversion
pathways of the chromophore in the protein matrix. The suggested mechan-
ism in Figure 6.12 shows that changes in the protonation state of the
chromophore and some proximal amino acids lead to different photoche-
mical states, which all turn out to be essential for the photoswitching
mechanism. These photochemical states are (1) a neutral chromophore,
which can photoisomerize back and forth between the *trans* and *cis* config-
uration on a subpicosecond timescale, (2) an anionic chromophore, which
rapidly undergoes radiationless decay after excitation in both the *cis* and
trans configuration, and (3) a long-lived and therefore putatively fluorescent
zwitterionic chromophore. The *trans* zwitterion can rapidly return to the
ground state through proton transfer to the neighboring Glu215 (Figure
6.11a). This alternative de-excitation pathway is not accessible for the *cis*
isomer, thus explaining why only the *cis* form fluoresces. The overall stabi-
lity (and thus the relative population) of the different protonation states is
controlled by the isomeric state of the chromophore.

On the basis of the simulations, it was proposed that radiation-induced
decarboxylation of the glutamic acid Glu215 could block the proton transfer
pathways that enable the deactivation of the zwitterions and thus leads to
irreversible fluorescence (Figure 6.12). Recent experiments on the structu-
rally similar protein Dronpa [55,56] also provide strong support for the
proposed protonation/deprotonation mechanism. The similarity between
the chromophores in a variety of fluoroproteins suggests that during mole-
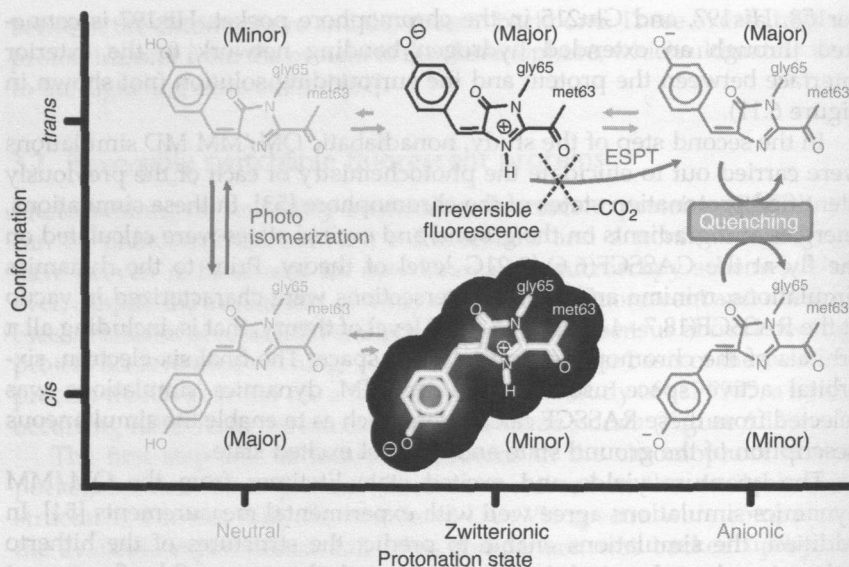cular evolution, the (*p*-hydroxybenzylidene)imidazolinone chromophoric

Figure 6.12 Scheme of the reversible photoswitching mechanism of asFP595. The fluorescent state Z_{cis} is highlighted. The green arrows indicate ground state equilibria, whereas the red arrows indicate excited state processes. The major protonation states are the zwitterionic and the anionic chromophores in the *trans* conformation, and the neutral chromophore in the *cis* conformation, as indicated in the square brackets.

moiety served as a template and that the photochromic properties—and thus the function—was fine-tuned by the protein environment.

4. FINAL REMARKS AND CONCLUSIONS

Understanding light-driven processes is one of the major goals of the bio- and nanosciences. The underlying molecular mechanisms are typically governed by subpicosecond atomic motions. Mechanisms on such ultrafast timescales are very challenging to probe by experiment. Here, MD simulations have become an invaluable tool to understand such processes in atomic detail.

In this contribution, we have reviewed our approach for excited state MD simulations. In the applications that we have selected here, the simulations have provided detailed structural and dynamical information of the photobiological process at a resolution well beyond what is achievable experimentally. The applications also demonstrate what is feasible today with on-the-fly MD 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 we can study today. 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 near future. Therefore, excited state MD simulation has the potential to lead to a better understanding of photobiological reactions. Ultimately, we expect these simulations to enable accurate predictions of photochemical properties and to become a standard tool for rational design of new photoactive systems.

REFERENCES

[1] A. Szabo, N.S. Ostlund, Modern Quantum Chemistry, Dover Publications, New York, 1989.
[2] H. Köppel, W. Domcke, L.S. Cederbaum, Adv. Chem. Phys. 57 (1984) 59.
[3] G.A. Worth, L.S. Cederbaum, Annu. Rev. Phys. Chem. 55 (2004) 127.
[4] J. von Neumann, E. Wigner, Z. Physik 30 (1929) 467.
[5] G.J. Atchity, S.S. Xantheas, K. Ruedenberg, J. Chem. Phys. 95 (1991) 1862.
[6] G.J. Atchity, K. Ruedenberg, J. Chem. Phys. 111 (1999) 2910.
[7] D.R. Yarkony, J. Chem. Phys. 112 (2000) 2111.
[8] H. Koppel, J. Gronki, S. Mahapatra, J. Chem. Phys. 115 (2001) 2377.
[9] H. Muller, H. Koppel, L.S. Cederbaum, New J. Chem. 17 (1993) 7.
[10] S. Matsika, D.R. Yarkony, J. Phys. Chem. A 106 (2002) 2580.
[11] I. Burghardt, J. Chem. Phys. 114 (2001) 89.
[12] M. Thachuk, M.Y. Ivanov, D.M. Wardlaw, J. Chem. Phys. 109 (1998) 5747.
[13] J. Onuchic, P. Wolynes, J. Phys. Chem. 92 (1998) 6495.
[14] M. Desouter-Lecomte, J. Lorquet, J. Chem. Phys. 71 (1979) 4391.
[15] A. Warshel, Nature 260 (1976) 679.
[16] A. Toniolo, S. Olsen, L. Manohar, T.J. Martinez, Faraday Discuss. 127 (2004) 149.
[17] K.K. Baeck, T.J. Martinez, Chem. Phys. Lett. 375 (2003) 299.
[18] R.J. Bartlett, M. Musial, Rev. Mod. Phys. 79 (2007) 291.
[19] I. Tavernelli, U.F. Rohrig, U. Rothlisberger, Mol. Phys. 103 (2005) 963.
[20] A. Dreuw, M. Head-Gordon, Chem. Rev. 105 (2005) 4009.
[21] J.D. Coe, T.J. Martinez, J. Phys. Chem. A. 110 (2006) 618.
[22] J.D. Coe, T.J. Martinez, J. Am. Chem. Soc. 127 (2005) 4560.
[23] O. Weingart, A. Migani, M. Olivucci, M.A. Robb, V. Buss, P. Hunt, J. Phys. Chem. A 108 (2004) 4685.
[24] M. Boggio-Pasqua, M.J. Bearpark, P.A. Hunt, M.A. Robb, J. Am. Chem. Soc. 124 (2002) 1456.
[25] A. Sanchez-Galvez, P. Hunt, M.A. Robb, M. Olivucci, T. Vreven, H.B. Schlegel, J. Am. Chem. Soc. 122 (2000) 2911.
[26] H.J. Werner, W. Meyer, J. Chem. Phys. 74 (1981) 5794.
[27] G.A. Worth, M.A. Robb, Adv. Chem. Phys. 124 (2002) 355.
[28] M.K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, Wiley, New York, 2000.
[29] B.O. Roos, Acc. Chem. Res. 32 (1999) 137.
[30] P. Celani, H.J. Werner, J. Chem. Phys. 119 (2003) 5044.

[31] J.D. Coe, B.G. Levine, T.J. Martinez, J. Phys. Chem. A 111 (2007) 11302.
[32] M. Boggio-Pasqua, M.J. Bearpark, M. Klene, M.A. Robb, J. Chem. Phys. 120 (2004) 7849.
[33] T.W. Keal, A. Koslowski, W. Thiel, Theor. Chem. Acc. 118 (2007) 837.
[34] J.A. McCammon, B.R. Gelin, M. Karplus, P.G. Wolynes, Nature 262 (1976) 325.
[35] H.J.C. Berendsen, Science 271 (2001) 954.
[36] F. Jensen, Introduction to Computational Chemistry, Wiley, New York, 2001.
[37] A. Warshel, M. Levitt, J. Mol. Biol. 103 (1976) 227.
[38] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. 91 (1987) 6269.
[39] M.J. Field, P.A. Bash, M. Karplus, J. Comp. Chem. 11 (1990) 700.
[40] P. Amara, M.J. Field, Theor. Chem. Acc. 109 (2003) 43.
[41] D.M. Philipp, R.A. Friesner, J. Comput. Chem. 20 (1999) 1468.
[42] J. Gao, P. Amara, C. Alhambra, M.J. Field, J. Phys. Chem. A 102 (1998) 4714.
[43] K.J. Hellingwerf, J. Hendriks, T. Gensch, J. Phys. Chem. A 107 (2003) 1082.
[44] M. Boggio-Pasqua, M.A. Robb, G. Groenhof, J. Am. Chem. Soc. 131 (2009) 13580.
[45] E.V. Gromov, I. Burghardt, J.T. Hynes, H. Köppel, L.S. Cederbaum, J. Photochem. Photobiol. A 190 (2007) 241.
[46] N. Mataga, H. Chosrowjan, Y. Shibata, Y. Imamoto, F. Tokunaga, J. Phys. Chem. B 104 (2000) 5191.
[47] N. Shimizu, H. Kamikubo, Y. Yamazaki, Y. Imamoto, M. Kataoka, Biochemistry 45 (2006) 3542.
[48] P. Changenet-Barret, P. Plaza, M.M. Martin, H. Chosrowjan, S. Taniguchi, N. Mataga, et al., Chem. Phys. Lett. 434 (2007) 320.
[49] K. Takeshita, Y. Imamoto, M. Kataoka, K. Mihara, F. Tokunaga, M. Terazima, Biophys. J. 83 (2002) 1567.
[50] G. Groenhof, L.V. Schäfer, M. Boggio-Pasqua, H. Grubmüller, M.A. Robb, J. Am. Chem. Soc. 130 (2008) 3250.
[51] G. Groenhof, M.F. Lensink, H.J.C. Berendsen, A.E. Mark, Proteins 48 (2002) 212.
[52] L.V. Schäfer, G. Groenhof, A.R. Klingen, G.M. Ullmann, M. Boggio-Pasqua, M.A. Robb, H. Grubmüller, Angew. Chem. Int. Ed. 46 (2007) 530.
[53] L.V. Schäfer, G. Groenhof, M. Boggio-Pasqua, M.A. Robb, H. Grubmüller, PLoS Comput. Biol. 4 (2008) e1000034.
[54] T.A. Schüttrigkeit, T. von Feilitzsch, C.K. Kompa, K.A. Lukyanov, A.P. Savitsky, A.A. Voityuk, M.E. Michel-Beyerle, Chem. Phys. 323 (2006) 149.
[55] S. Habuchi, P. Dedecker, J.I. Hotta, C. Flors, R. Ando, H. Mizuno, A. Miyawaki, J. Hofkens, Photochem. Photobiol. Sci. 5 (2006) 567.
[56] E. Fron, C. Flors, G. Schweitzer, S. Habuchi, R. Ando, H. Mizuno, J. Am. Chem. Soc. 129 (2007) 4870.