Chem Soc Rev

Learning from Photobiology how to Design Molecular Devices with a Computer

Journal:	*Chemical Society Reviews*
Manuscript ID:	CS-TRV-01-2014-000037.R2
Article Type:	Tutorial Review
Date Submitted by the Author:	22-Apr-2014
Complete List of Authors:	Gozem, Samer; Bowling Green State University, Department of Chemistry Melaccio, Federico; Università di Siena, Department of Biotechnology, Chemistry and Pharmacy Luk, Hoi Ling; Bowling Green State University, Department of Chemistry Rinaldi, Silvia; Università di Siena, Department of Biotechnology, Chemistry and Pharmacology Olivucci, Massimo; Università di Siena, Department of Biotechnology, Chemistry and Pharmacy

Chem Soc Rev

RSCPublishing

Tutorial Review

Learning from Photobiology how to Design Molecular Devices with a Computer

qCite this: DOI: 10.1039/x0xx00000x

S. Gozem,[a] F. Melaccio[b] , H. L. Luk [a], S. Rinaldi[b] and M. Olivucci[a,b]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Biological photoreceptors and fluorescent proteins provide striking examples of how non-covalent interactions could be exploited for tuning the photochemistry and photophysics of organic chromophores. In this tutorial review we show how the construction of computer models of such natural supramolecular systems not only provides atomic-level information on the mechanisms of their function, but also principles useful for designing light-responsive components of artificial supramolecular systems. Using few complementary case studies, the intellectual process leading to the implementation of such an engineering target is followed up to the actual construction of a working prototype of a biomimetic molecular switch.

Introduction

Supramolecular photochemistry is an interdisciplinary field dealing with how spectroscopic and photochemical properties are influenced by non-covalent interactions.[1]. In Fig. 1 left we show a schematic representation of a system whose fluorescence is modulated by non-covalent bonding (e.g. hydrogen bonding, ionic interactions, π-stacking, or van der Waals forces) between two constitutive units **A** and **B**. Upon irradiation with light, **B** fluoresces when it is unbound, but its fluorescence is quenched when it binds to **A**. If **A** and **B** are conventional molecules of known structure it may be possible to trace the non-covalent interactions responsible for the change in the system properties. However, in larger systems such as those formed by a molecular cavity incorporating a fluorescent unit (Fig. 1 right), the quenching mechanism may be impossible to comprehend by looking at the **AB** structure. The systems provided by Nature and investigated by the field of photobiology belong to this class.

For systems such as fluorescent proteins or visual pigments, crystallographic structures provide landmark information. Indeed, they unveil how the protein cavity residues interact with internal chromophores. However, such structures cannot be systematically determined for all proteins and/or their mutants. Furthermore, they may not provide enough information to explain how residue-chromophore interactions control the function of the protein. In such cases, computer models represent powerful research tools. Indeed, they give access to an atomic-level representation of the system that makes it possible to describe the geometry of the interacting units, classify the non-covalent interactions between them, and draw hypotheses on how their properties are affected by these interactions. When using the advanced computer models described in the present tutorial, it becomes even possible to investigate the function of the system itself, to quantitatively reproduce observed properties or predict changes such as an increase in fluorescence quantum yield or, most remarkably, a change in chemical reactivity.

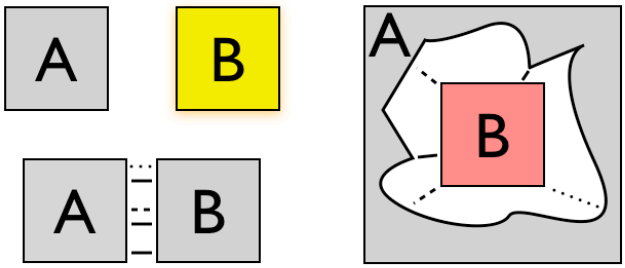


Fig. 1 Schematic representation of basic supramolecular systems. Left. **A** and **B** are two molecular units which may bind to form a binary system (**AB**). Right. A more complex binary system formed by a hollow host **A** (e.g. a protein cavity) and an emitting or reactive guest **B**. The units may bind using a number of different non-covalent interactions (represented by solid, dashed, and dotted lines).

Below, we start by presenting a set of "design principles" that will be used throughout the tutorial. These will be collected in a scheme related to the Jablonski diagram but featuring "paths" along energy curves rather than "jumps" between energy levels. In the second part we provide an overview of the computational tools available for building "functional" models of supramolecular systems, and explain the main considerations and decisions involved in building these models. We show that improvements in the available quantum chemical technology are needed to approach a quantitative description. In the third part of the tutorial we review some recent studies on light-responsive proteins to demonstrate how the design principles operate. Finally, in the fourth and last part of the tutorial we follow the route that led to the design of a biomimetic molecular switch that, ultimately, was prepared in the laboratory. This tutorial is by no means comprehensive (a

comprehensive review is not possible given the wide scope and limits in length). Instead, we walk the reader through some of the main considerations involved in the computational study of supramolecular systems, and through the intellectual process leading to the engineering of novel light-responsive systems.

Design Principles from Supramolecular Photochemistry

Upon absorbing a photon of light, a system is promoted from the ground state (GS) to an electronically excited state (ES). If we assume, for our purposes, that the system will end up on the lowest ES of the same GS spin multiplicity (this is the case in many systems), then the events following light absorption must be controlled by the ES, as well as the GS, potential energy surfaces (PESs). It is then the job of a computational chemist to understand how the molecular and electronic structure of the system might evolve along the PESs. The dimensionality of a PES is 3N-6 for non-linear systems (where N is the number of atoms), but often it is one important, but complex, geometrical coordinate, called the *reaction coordinate*, that explains much of the photophysics and/or photochemistry of the system.

The calculation of the reaction coordinate along the ES and GS PESs reveals fundamental information. For instance, at the GS energy minimum, usually taken as the coordinate origin, one obtains the vertical excitation energy (process **a** in Fig. 2) in terms of the ES-GS energy difference and therefore an estimate of the *wavelength of the absorption* maximum (λ^{a}_{max}). The change in potential energy along the ES reaction coordinate may either reveal a barrier (controlled by a transition state, TS) or not. The former case implies the existence of an ES energy minimum (FS) from which the *wavelength of the emission* (process **b** in Fig. 2) maximum (λ^{e}_{max}) may be computed.

For a polyatomic system, the reaction coordinate, as well as any structure change along the entire ES and GS PESs, are described in terms of coupled changes of 3N-6 internal coordinates such as bond lengths, angles and torsions (where N again is the number of atoms). Such changes can eventually lead to intersections where the ES and the GS potential energies become identical (or *degenerate*). When the intersecting states have the same spin multiplicity, such crossing points are called conical intersections (CIs). CIs are not isolated points but are part of a 3N-8 subspace known interchangeably in literature as a hyperline, seam of intersection, intersection space, or intersection subspace. This is because, at a CI, the energy degeneracy can only be lifted along two specific molecular modes (i.e. specific combination of internal coordinates) known as the branching place vectors (for more on CIs see ref. 2 and references therein). At a CI, a system decays rapidly from the ES to the GS (a common analogy is that a CI point acts as a "funnel" allowing molecules on the ES to funnel through to the GS). CIs are ubiquitous in polyatomic systems and are involved in many photochemical processes.

The shape, energy, and geometry of a CI often decide the *fate and timescale of a photochemical reaction*. In particular, if a CI is accessible from the Franck-Condon (FC) point, it provides a fast channel for population decay to GS. This would appear as a barrierless reaction path (path **1** in Fig. 2), which would reveal a short (subpicosecond) *ES lifetime* and display only transient fluorescence. On the other hand, in a barrier-controlled path (path **2** in Fig. 2) there would be a higher fluorescence quantum yield as a consequence of the longer ES lifetime that, in turn, will depend on the height of the ES barrier. Finally, if there is no CI accessible from the FC region (e.g. the CI is high in energy) the system will stop its evolution at an ES minimum, and decay to GS either radiatively or non-radiatively but with a slower and "conventional" *internal conversion* process (i.e. its rate will depend on the Fermi golden rule). After decay at a CI point, the system may follow one or more GS relaxation coordinates. One coordinate may describe a return to the GS reactant. This process defines a different CI-mediated *internal conversion* route (which is typically much faster than "traditional" internal conversion). Simultaneously, or in distinct cases, the reaction coordinate may proceed to a different GS species and therefore to a *photochemical reaction*.

Above we have explained how certain elements of the ES and GS PESs (energy minima, transition states and conical intersections as well as different "segments" of the reaction coordinate) are associated to a wanted or unwanted property of a light-responsive system (a specific λ^{a}_{max} or λ^{e}_{max}, a certain ES lifetime, the formation or quenching of a specific photoproduct). When focusing on the design of supramolecular systems, the mapping of the reaction coordinate helps one to learn how to control the PES shape by changing the type and pattern of the non-bonding interactions connecting the chromophore with its environment (be it solvent, the cavity of a protein or a man-made molecular cavity). The effects of the molecular environment on the PES can be investigated using appropriate computer models, which will be discussed in the next section.

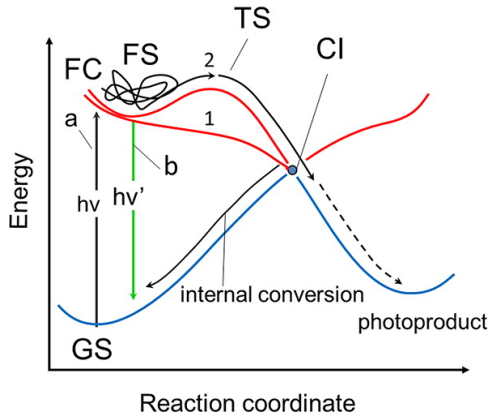


Fig. 2 Schematic representation of the barrierless (1) and barrier-controlled (2) ES path (a, absorption; b, emission; FC, Franck-Condon point; GS, ground state; FS, fluorescent state; TS, transition state; CI, conical intersection). Adapted from ref. 3

It must be stressed that, although informative, the approach above based on PES mapping does not usually provide predictions for quantities such as ES lifetimes and reaction quantum yields. These quantities require a description of the nuclear motion along the entire (3N-6 dimensional) valley defined by the reaction coordinate. In fact, as a consequence of such multidimensional motion a CI may be approached or exited not only along the reaction coordinate, but also along other coordinates (i.e. along both the branching plane vectors). Understanding the effects of other dimensions is a difficult task that could be investigated only with classical or semi-classical dynamics calculations (e.g. computing sets of trajectories starting on the ES PES) rather than the "static" calculations required for computing the reaction coordinate.

Computational Photochemistry and Photobiology

The computational investigation of photochemical processes in natural or artificial supramolecular systems requires the construction of models that satisfy two requirements. First, the need to respond to light implies that the models must be capable to represent the geometrical and electronic structure of both the GS and ES of a molecule or group of molecules. This is possible employing only a subset of electronic structure theory methods that have this capability. Second, since supramolecular systems correspond to molecules with hundreds, if not thousands, of atoms they require computationally affordable methods (i.e. requiring an affordable quantity of computer time). Unfortunately, these requirements are mutually exclusive. Typically, the electronic structure theory methods capable of accurately describing the ES electronic structure are computationally expensive and limited to systems containing tens of atoms.

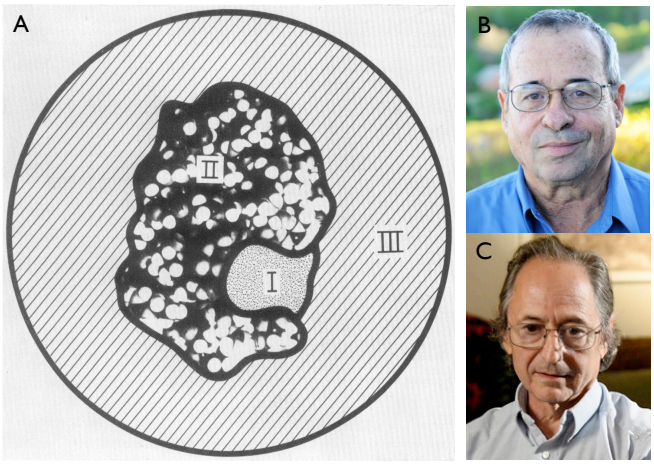


Fig. 3 A. A schematic representation of a multiscale approach. This figure is adapted from the original 1976 paper where this approach was first used by Warshel and Levitt, both recipients of the 2013 Nobel prize in chemistry.[4] Here, region I is described using accurate quantum mechanical methods, region II is described by more approximate but inexpensive molecular mechanics methods, and region III comprises the surrounding environment such as a solvent or lipid bilayer and may be described even more approximately by using, for instance, a dielectric continuum (this region is not always included). **B**. Arieh Warshel. **C**. Michael Levitt.

A solution to the above problem has been to separate a system into two (or more) parts that are treated at different levels (Fig. 3). This "multiscale" approach has been widely applied and has been recognized with the 2013 Nobel Prize in Chemistry. The use of multiscale approaches is well-suited for supramolecular systems that are, by definition, separable into distinct units interacting via non-bonding potentials, and where often it is only one unit (e.g., the chromophore) that is responsible for the spectroscopy and/or photochemistry of the system. Therefore only the chromophore needs to be treated with a suitable level of electronic structure theory, while the rest of the system is treatable at a less expensive level.

There are some alternatives to multiscale modelling that allow users to study the photochemistry of large systems. One possibility is to use semi-empirical methods, which are widely employed for the GS but have also successfully been applied to ES (e.g. see ref. 5). A second possibility lies in emerging electronic structure theory technologies that employ graphics processing units (GPUs) rather than the standard central processing unit (CPU) for performing the computation. These technologies allow for significant speed-ups making, in principle, possible to map the PES of an entire protein at a level of theory capable to describe the ES.[6] However, these two approaches are not yet widely applied (in the case of the latter that is because the approach is in its infancy). We would also not discuss the potential use of quantum computers, a fascinating development that may revolutionize the field by dramatically decreasing the computational cost of the most accurate methods available.[7]

The rest of this section is divided into three parts. In the first part, we present a simplified account of a few of the most common electronic structure theory (i.e. quantum chemical) methods that are currently used to study the properties of organic chromophores found in photochemistry and photobiology. Such chromophores may be studied isolated (i.e. in the gas-phase), or embedded in a specific environment. One may often study the isolated chromophore as a preliminary step towards the modelling, study, and comprehension of the full supramolecular system. In the second part, we discuss how quantum chemical methods are used in multiscale modelling allowing a user to study supramolecular and photobiological systems. In the third part, we outline some of the key steps involved in building a common type of multiscale model.

Electronic structure theory for excited states. There are several electronic structure theory methods capable of describing both GS and ESs. While an accurate method exists (i.e. the full configuration interaction, FCI, method, which solves the electronic part of the time-independent, non-relativistic Schrödinger equation exactly within a given basis set), it is computationally expensive and limited to systems made of just few atoms. Therefore, most practical electronic structure theories rely on approximations, or on an ansatz. Selecting a method (or combination of methods) becomes an issue of balancing cost and accuracy. Note that the cost of a method is not only related to its mathematical formulation, but depends also on the efficient algorithms and programming strategies used to optimize performance on the available computer architecture.

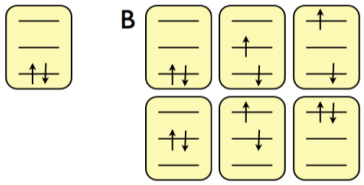


Fig. 4 A. Representation of a closed-shell electronic configuration describing the GS of a system with two electrons in the "basis" of three molecular orbitals. **B.** Representation of all possible configurations for a system with two electrons in three orbitals. An accurate method must use a wavefunction that includes all necessary Slater determinants. In practice, a multi-determinant wavefunction is constructed from "configuration state functions", which are linear combinations of Slater determinants corresponding to spin eigenstates.

One factor dominating the accuracy of electronic structure methods is the ability to account for electron correlation. This refers to the ability for an electron to interact with other electrons individually and not in a mean-field fashion as in the Hartree-Fock (HF) method where the electron correlation is, by definition, missing. The HF wavefunction corresponds to a

single *Slater determinant* describing a closed-shell electronic configuration exclusively (see Fig. 4A). Thus the missing electron correlation energy can be recovered including in the wavefunction determinants describing electronically excited and open-shell states (see Fig. 4B).

Electron correlation can be divided into two types: dynamical electron correlation and static (or non-dynamical) electron correlation. Static electron correlation becomes significant when a state has important contributions from more than one electronic configuration which is required, for instance, to describe a change in the "bonding state" of the system (as in the homolytic breaking of a covalent bond). This situation can be accounted for by including the few specific electronic configurations describing the change in bonding. Dynamical electron correlation, on the other hand, is related to the instantaneous interactions between electron pairs, and can only be accounted for by including many configurations, even including those with small weights in the system wavefunction. Of course, including all possible electron configurations in the wavefunction, as in the FCI method mentioned above, would account for both types of electron correlation, and provide the exact solution of the Schrödinger equation relative to the chosen atomic basis set. As such calculations are far too expensive for molecules of interest (the computational cost of FCI energies increases very rapidly with the number of basis functions), one practical solution has been to use "truncated" configuration interaction methods. These methods are defined by wavefunctions including only determinants describing electronic excitations up to a certain order with respect to the single HF determinant. One can therefore truncate the wavefunction after including all single excitations (CIS) or single and double excitations (CISD) with respect to the reference HF configuration (of course, higher excitations are also possible). However, while these methods may account well for static electron correlation, and for part of the dynamics electron correlation necessary to describe the GS (since they include excitations *relative to the GS configuration*), they typically do not account well for the dynamical electron correlation necessary for describing the ES.

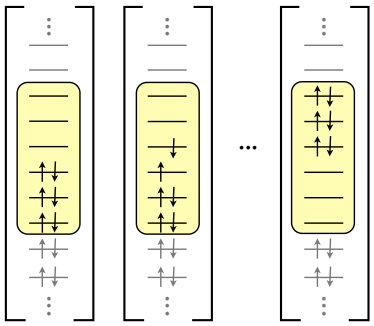


Fig. 5 A representation of the configurations included in the wavefunction of the CASSCF method with an active space of six electrons and six orbitals. The active space orbitals are highlighted in yellow, while inactive orbitals are greyed out.

Another important property of electronic structure methods is size-consistency. A size-consistent method would give the same electronic energy value for two non-interacting systems regardless if this is calculated as a sum of individual energies or of a composite of the two non-interacting systems. The performance of methods that are not size-consistent typically deteriorates with increasing size of the system they describe. One main problem of truncated configuration interaction methods, with the exception of CIS, is that they are not size-consistent [8].

The complete active space self-consistent field (CASSCF) method represents a valid alternative to CIS, CISD, etc. methods. Its wavefunction is defined by selecting a "reference or active space". This is defined by selecting a set of electrons and orbitals considered to be necessary for the description of the spectroscopic or reactivity property under investigation. For instance and typically, the description of a double bond isomerization requires an active space formed by all electronic configurations that may be formed selecting all π-electrons and π-orbitals of the conjugated system incorporating the reactive double bond. This choice is motivated by the need to allow for all possible variations in overlap between the set of p-orbitals forming the reacting π-system along the reaction coordinate or trajectory. Therefore the selection of the active space electrons and orbitals is a "chemical" problem, and often is not a straightforward problem (for more on the selection of an active space see ref. 9). As displayed in Fig. 5, in the CASSCF method all possible configurations of the active space are included in the wavefunction. The orbitals not incorporated in the active space (the inactive orbitals) remain either doubly occupied (the core space) or empty (the virtual space). Furthermore, all orbitals are optimized together with the configuration weights. Since this is just like performing a FCI limited to the active space, the CASSCF wavefunction is size consistent and accounts for static electron correlation, but not for all the dynamical electron correlation since the motion of the electrons outside the active space is not correlated.

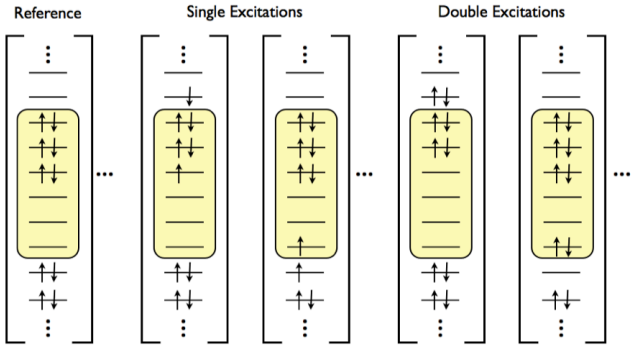


Fig. 6 A representation of the configurations included in the MRCI wavefunction taking one of the CASSCF configurations of Fig. 5 as the reference. In classical (uncontracted) MRCI, excitations are performed out of and into the active space from all inactive orbitals for each of the reference CASSCF. MRCISD would include both single and double excitations, as shown here.

While CASSCF is expected to perform better than CISD in describing chemical processes (such as those involving bond-breaking), it is not quantitatively accurate, and sometimes may not be even qualitatively accurate either. To overcome this problem CASSCF is often used as a reference wavefunction for more quantitative multi-reference methods. One such method is multi-reference configuration interaction (MRCI), which uses the same principle as truncated CI, but rather than having a single-reference configuration (i.e. the single HF determinant), it has several reference configurations. For instance, MRCISD would take each CASSCF configuration and expand the CASSCF wavefunction by adding single and double excited configurations on top of it (see Fig. 6). This accounts for a larger portion of electron dynamical correlation. MRCISD is a

variational method (i.e. its basis is the variational principle). A simple energy correction, derived from a perturbation theory analysis, the Davidson correction, may be applied to MRCISD (as well as to CISD) to estimate the energy of configuration interaction up to quadruple excitations from only single and double excitations in a computationally inexpensive way. The resulting Davidson corrected-MRCISD (MRCISD+Q) energy is considered to yield values close to FCI energy, and is therefore both qualitatively and quantitatively accurate. However, MRCISD+Q becomes very expensive beyond a certain number of reference configurations and basis functions, so its use is still limited to systems smaller than the typical organic chromophores absorbing and/or emitting in the visible. A less expensive alternative to MRCISD applicable to larger systems is achieved by accounting perturbatively for the missing dynamical electron correlation. Multi-reference second order perturbation (MR-PT2) methods account for interactions between CASSCF active and inactive electrons and therefore include a large part of the missing dynamical electron correlation.

One of the most important considerations when choosing an approximate electronic structure method for simultaneously calculating GS and ES energies (as typically required for photophysical and photochemical problems) is to find a balance between accuracy and computational cost. In fact, a certain method may be considered suitably accurate for describing one region of a PES of a given electronic state, but inaccurate for describing other regions of the same PES or the transition to other states. This is important when one wants to compute energy gaps between electronic states (as for processes **a** and **b** in Fig. 2) or energy differences between different points of the same PES (as for processes **1** and **2** in Fig. 2). Different PES points may be dominated by different electronic configurations and, therefore, require different amounts of electron correlation to be described accurately. In other words, it is important that the electronic structure method is capable to correctly account for *differential* electron correlation effects on the PES regions of interest. We have recently reported on a series of benchmark studies that tested several quantum chemical methods for their ability to describe PESs featuring a mixture of covalent, charge transfer, and diradical character, especially in the region of a conical intersection.[10] These studies have shown that the inclusion of dynamical electron correlation (largely missing in methods such as CISD and CASSCF) may result not only in quantitatively different, but also in qualitatively different description of GS and ES processes.

In computational photochemistry and photobiology one usually optimizes the chromophore at the CASSCF level of theory and then obtains its energy using MR-PT2. Most commonly, the complete active space second order perturbation theory (CASPT2), which is a type of MR-PT2 theory, has been used leading to often suitably accurate (e.g. with errors within 3 kcal mol^{-1} with respect to observed vertical excitation energies) CASPT2//CASSCF protocol.[11,12] Such protocol has been shown to benefit from a balanced cancellation of errors (at least for protonated and conjugated Schiff bases) that needs to be understood.[10a] Thus, while the CASTP2//CASSCF protocol has been successfully applied in many studies, it typically requires experienced users that are familiar with its pitfalls. In fact, this requirement is generally true of multi-configurational (such as CASSCF) and multi-reference (like MRCI and MR-PT2) methods, which require users capable to select variable such as the active space and the number of states to include in the calculation. CASSCF, MRCI, and MR-PT2 methods are therefore considered "unattractive" because they are not black-box methods.

Apart from multi-configurational and multi-reference methods, there are several types of single reference methods that are black-box methods capable of describing ESs. One of the most common is based on the coupled cluster (CC) theory. This theory accounts for electron excitations, but uses an exponential ansatz that operates, typically, on a single reference. Due to the use of an exponential ansatz in the wavefunction, CC methods are typically size-consistent (as long as the reference is also size-consistent). The computational cost of CC scales with the level of excitation. For instance, CCSD scales as N^6, CCSDT scales as N^8, and CCSDTQ scales as N^{10} (where S stands for singles, D for doubles, T for triples, Q for quadruples, and N is usually a relative measure of the system size – e.g., if one doubles the number of electrons and the number of basis functions, the calculation will take 2^n times as long per iteration in terms of CPU time). In many cases, triples or quadruples corrections are added perturbatively, which introduces these corrections in a cheaper way. As an example, CCSD(T), where the parentheses indicate that the triples correction has been added perturbatively, scales as N^7.

There are several extensions of CC methods that make them applicable to ESs, such as linear-response CC (LR-CC) and equation of motion CC (EOM-CC). EOM-CC is particularly powerful because different flavours of it have been developed that allow for calculations of excited states, open-shell species, ionization potentials, and electron affinities.[13] Unfortunately, the use of CC methods is again limited by their computational cost, as discussed above.

Finally, a black-box theory commonly used for computing excitation energies and accounting for both static and dynamical electron correlation is time-dependent density functional theory (TD-DFT). Methods based on TD-DFT are less expensive than multi-reference and CC methods, and therefore can be applied to larger molecules. In several cases TD-DFT methods have shown good accuracy in excitation or emission energy calculations. Obviously, the quality depends on the functional used which must be selected by the user. This is actually critical when the problem under investigation involves charge transfer states, charge separated states, or states with large weights of electronic configurations beyond single excitations.[8] In addition, TD-DFT methods are known to have several problematic issues in the vicinity of crossing points. However, new DFT methods are being developed to deal with such issues, and results have been encouraging.

Multiscale methods for supramolecular and photobiological molecules: The QM/MM approach. While in future it may be possible to model an entire light-responsive supramolecular system using a quantitatively accurate electronic structure theory, (e.g. with the advent of new technologies such as algorithms on the GPU[6] or quantum computers[7]) this is presently impossible. For this reason one employs a multiscale approach where a large system is divided into two or more interacting parts treated using different levels of theory. The quantum mechanics / molecular mechanics (QM/MM) approach[4] represents a rather basic multiscale approach involving the description of one part of the system with an electronic structure, and therefore quantum mechanical (QM) method and the remaining part at the inexpensive molecular mechanics (MM) level. Despite its simple form, QM/MM is far from being a black-box method and requires critical decisions that would affect the validity of the model. In the following two

sections, we will cover some of the main considerations involved when building a QM/MM model. However, the building protocol depends on the specific chemical problem being investigated. In order to ensure generality, we will not delve into such details here.

QM/MM methods can be classified as either subtractive or additive.[14] In additive schemes, the total QM/MM energy is defined as:

$$E_{Total} = E_{QM} + E_{MM} + E_{QM/MM}$$

While the first two terms represent the contributions coming from the atoms of the non-interacting QM and MM subsystems, the third term, the QM/MM coupling term, describes their interaction. The coupling term incorporate the effect of covalent bonding, van der Waals, and, most importantly, electrostatic interactions between the QM and MM atoms.

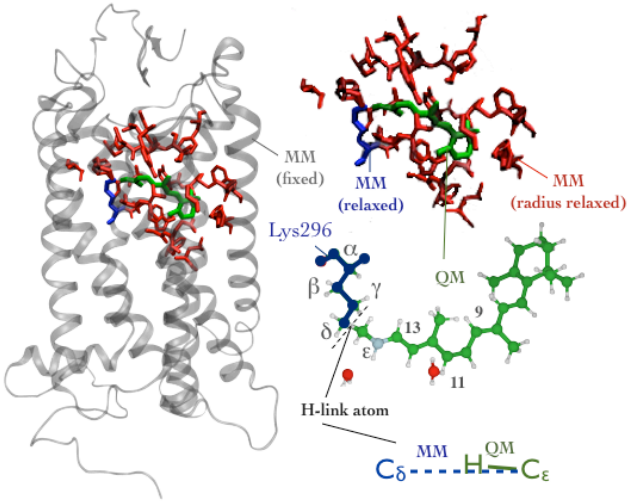


Fig. 7 A schematic representation of a QM/MM computer model of bovine rhodopsin based on the link-atom scheme. The "QM subsystem" (in green) is limited to the N-methyl 11-*cis* retinal protonated Schiff base moiety containing a π-system formed by six conjugated double bonds and corresponding to the macromolucule spectroscopically and chemically reactive part. It interacts with the MM "environment", the MM subsystem, via non-bonding interactions and a single covalent bond with the Lys296 residue side chain at the level of its C_δ atom. Thus the carbon corresponding to the N-methyl substituent of the QM subsystem is actually the Cε carbon of the "truncated" Lys296 side chain. In the model the Cε-H bond is required to be geometrically aligned with the C_δ-Cε segment. The MM subsystem may feature components with different properties. In the presented model part of the subsystem (in grey) is kept fixed during the geometrical relaxation (i.e. when finding the equilibrium structure of the entire supramolecular system), reaction coordinate computations, or trajectory computation. The remaining parts (red and blue) are instead relaxed or set into motion when computing, for instance, semi-classical trajectories and represent the supramolecular cavity (e.g. unit C in Fig. 1). Part of the MM cavity is special and correspond to the Lys296 side chain moiety (in blue) whose MM parameters are non-standard but modified to correctly describe the frontier through the Cδ-Cε bond.

The main differences among additive QM/MM models are related to the treatment of the electrostatic interactions and, in cases where there is a covalent bond between the QM and MM regions, the treatment of the frontier atoms (i.e. those lying at the border between the QM and MM subsystems). For the electrostatic interaction between the QM charge density and the MM atoms, three different approaches are possible: the mechanical, electrostatic and polarizable embedding schemes. In the mechanical embedding, the interactions between QM and MM regions are computed by Coulomb's law at the MM level using wavefunction-derived atomic charges on QM centers and parametrized charges on MM centers. This is the simplest treatment of the electrostatic coupling, but misses a component, since the MM charges do not polarize the QM wavefunction that is computed for an isolated QM moiety. On the contrary, in electrostatic embedding methods, the QM/MM electrostatic interaction is treated at the QM level by including the MM point charges in the electronic Hamiltonian of the QM subsystem. Thus, its electronic structure is polarized by the surrounding charge distribution. A distinct implementation of electrostatic embedding is to use the ElectroStatic Potential Fitted (ESPF) method,[15] which polarizes the QM wavefunction using the MM charges, but where the subsystem electrostatic interaction energy is calculated by using multipoles centered on the QM atoms and fitted to the electrostatic potential. However, in both mechanical and electrostatic embedding schemes, the MM part is not polarized by the QM part. To include this last contribution, one must use polarizable embedding schemes, which introduce polarizable MM charges and/or dipoles into the QM Hamiltonian, allowing for a mutual polarization of the two parts. Recent developments have focused on polarizable force fields for biomolecules, but require further testing to be systematically used in computational studies.[16,17]

In their simplest form, subtractive QM/MM schemes rely on the following energy expression:

$$E_{Total} = E_{QM}(small) + E_{MM}(full) - E_{MM}(small)$$

where "small" stands for the QM part and "full" is the entire system (both QM and MM regions). The main advantage of subtractive schemes is their simplicity, since no explicit QM/MM coupling is involved and energies can be obtained by performing three independent calculations. However, subtractive schemes require that force field parameters be available for the QM subsystem as well as the MM subsystem. This is due to the need to subtract the E_{MM}(small) energy term which has the same size as the QM subsystem and may generate issues when the QM subsystem corresponds to unconventional moieties whose MM parameters are not present in standard force fields. Moreover, traditionally, the QM/MM electrostatic interactions in subtractive schemes are treated at the MM level, and therefore the QM subsystem is not polarized by the MM charges (unless polarizable force fields are used in the E_{MM}(full) and E_{MM} (small) terms). However, the subtractive scheme has been made popular by the generalized version implemented by Morokuma and co-workers (Our own N-layered Integrated molecular Orbital and Molecular mechanics, ONIOM).[18] ONIOM has an added feature that allows for the use of more than two regions that are treated at progressively higher levels of accuracy. More recently, ONIOM has also been augmented with an electrostatic embedding treatment (ONIOM-EE) that allows the polarization of the QM subsystem.[19]

In some cases the QM subsystem may be covalently bound to the MM one. It is thus necessary to treat the "frontier" between the subsystems, which requires "cutting" one or more

covalent bonds. Different approaches are possible but we just mention three of them. In the link-atom scheme, an additional atom (usually a hydrogen atom) is capping the frontier QM atoms saturating their valence (see Fig. 7). The boundary-atom approach replaces the MM frontier atom with a special atom that appears in both the QM and MM subsystems. Finally, in the localized orbital scheme, the pending valences of the QM subsystem are saturated with *ad hoc* occupied frontier orbitals.[14]

Conventional QM/MM models suffer from several limitations. While, as explained above, van der Waals and electrostatic interactions, as well as bonded (covalent) interactions are accounted for (these are typically the most important interactions), there are other interactions that are neglected. These include interactions associated with the missing electrostatic polarization of the MM subsystem and the missing reciprocal exchange of the QM and MM subsystem electron densities (which may change the value of the van der Waals and electrostatic interactions). Finally, QM/MM models usually require specifically parameterized elements of the chosen MM force fields especially for describing the frontier region, which are not readily available (see legend of Fig. 7).

What does it take to build a QM/MM model? There are many considerations and decisions involved in building QM/MM models. The success of QM/MM models relies on the following main factors:

1. The availability of an experimental structure to be used as a template for model building (it is usually impossible to guess a valid template). X-ray crystallographic structures usually provide acceptable starting points. If they are not available, a crystallographic structure for a similar system may suffice (e.g. in biomolecules, for a system with a high degree of homology). In such cases, homology modelling can be used to construct a guess structure for the system of interest. Otherwise, the model building will have to rely on an educated guess, but this is only possible in well-ordered or relatively simple systems.
2. The partitioning of the QM/MM model into the QM and MM subsections. The decision on how to partition the system is crucial. In photobiological studies, the QM subsystem must at least include the chromophore. However, in cases where the chromophore is covalently linked to the MM subsystem it is important that the QM/MM frontier is at least one or two bonds removed from the chromophore such that it does not overpolarize it. In some cases, it may also be important to include other moieties in the QM subsystem. This might include moieties that exchange electronic density with the chromophore (e.g. an obvious case is when the chromophore is involved in a photoinduced electron transfer event), or moieties that have important interactions with the chromophore that cannot be adequately described at the level of the QM/MM interactions (e.g. pi-stacking). In some studies, several moieties near the chromophore are included in the QM subsystem to improve the description of the electrostatic environment around the chromophore. This approach is an effective way to account for all the important interactions of the chromophore with its environment, but is limited by the computational cost of treating an expanded QM subsystem.
3. The selection of methods to treat the QM and MM subsystems. This is often difficult as one does not have a clear best choice. Suitable QM methods have been above and are characterized by the balance between computational cost and accuracy required by the chemical problem. The MM subsystem, on the other hand, may be treated by any number of force field parameters. Commonly used MM force fields include AMBER, CHARMM and GROMOS for biological macromolecules, and OPLS for solvents. The best choice of MM force fields is not obvious. However, the focus of computational photochemistry and photobiology is on the chromophore and its interaction with the environment, so the treatment of the QM subsystem and QM/MM interactions are more important than the treatment of the MM subsystem.
4. Information about the protonation states of acidic and basic groups in the system. Often, if a model is built from an available crystallographic structure, the resolution is not high enough to reveal the positions of hydrogen atoms. Therefore, hydrogen atoms must be added to the model, and this involves making decisions regarding protonation states and hydrogen atom orientations. This is a tricky step in protein models, where there are usually several acidic and basic residues, and where there may exist a complex hydrogen-bonding network that depends on hydrogen atom placement. In such cases, one may rely on experimental results (e.g. FTIR studies) if available. Otherwise, there are computational methods that may be used to tentatively assign protonation states[20] and sample hydrogen bond orientations.[21] Furthermore, the problem does not always have a unique solution, and there are instances where this has led to disagreements in the literature. On the other hand, having the correct protonation states of all residues in a model is not always vital. For instance, residues that are remote from the chromophore are unlikely to have a large effect on the photochemical events of interest (if these are ionized there are usually counter-ions nearby). In cases where the residue in question is near the chromophore, it is advisable to test both the protonated and deprotonated models if its protonation state is unknown or debated.
5. The QM/MM model must provide a representation of the *average* conformation/structure of the system. A complex molecule does not usually correspond to a single conformer, but is rather represented by a number of interconverting conformers. This is certainly true for proteins and, therefore, biological photoreceptors where the side chains surrounding the QM subsystem may sample different conformers at room temperature. To address this issue one possibility is to use specific side chain conformers (e.g. derived by compiled conformer libraries and/or imposing structural constraints to the crystallographic structure, or by selecting highly populated conformers via molecular dynamics sampling). Another possibility is to model the average population with an ensemble of conformations that are randomly selected via molecular dynamics. The first approach is more widely adopted, but the second approach is superior, although more computationally expensive (requires the modelling of many structures rather than just one) and is hard to apply.
6. The possibility to validate the final model. This is done via a comparison of the values of computed and measured observables. In photochemical and photobiological systems, it is common to compare vertical excitation and emission energies to the corresponding observed λ^{a}_{max} and λ^{e}_{max} values, which are typically available.

Due to the complexity of supramolecular systems, it is difficult to produce a "correct" model (i.e. one that is approaching the chemical accuracy of ca. 1 kcal mol^{-1}). There is

always room for refinement, and with advancing technology and software availability, models are constantly being improved. Therefore, caution must always be exercised because differences in the building protocol, QM level of theory and even error cancellations may lead to different QM/MM models and results. For this reason it is advisable to use QM/MM models for comparative/trend studies (e.g. for studying changes among a set of different systems) rather than absolute properties (i.e. to measure an absolute value quantitatively). The potentially rewarding applications reveal a need for standardized building protocols that can be used systematically to produce sets of consistent QM/MM models by different labs worldwide. Work towards this aim is currently being pursued in our labs for QM/MM models of retinylidene proteins.

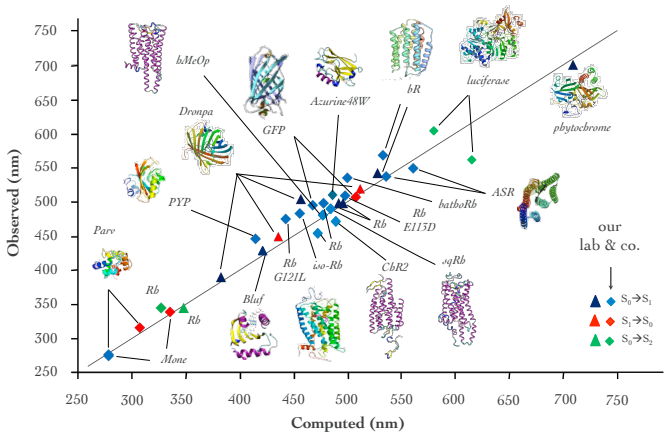


Fig. 8 QM/MM models of light-responsive proteins (including photoreceptor and fluorescent proteins). Comparison between observed and CASPT2//CASSCF/MM computed λ^{a}_{max} ($S_0 \rightarrow S_1$, $S_0 \rightarrow S_2$) and λ^{f}_{max} ($S_1 \rightarrow S_0$) excitation energies (rhombuses indicate data from the author and coworkers labs (the employed MM force fields are either AMBER or CHARMM). Triangles indicate data from other groups). The data include bovine rhodopsin (Rh),[12a-12d] bathorhodopsin (bathoRh),[12d,e] isorhodopsin (iso-Rh),[12e,f] the Rh mutants E113D and G121L,[12d] squid rhodopsin (sqRh),[12g] bacteriorhodopsin (bR),[12h] Anabaena sensory rhodopsin (ASR),[12h] human Melanopsin (hMeOp),[12i] Channel Rhodopsin (ChR2),[12j] green fluorescent protein (GFP),[12k] photoactive yellow protein (PYP),[12l] the E48W azurin mutant,[12m] blue-light using FAD (BLUF),[12n] the tryptophan containing proteins parvalbumin (Parv) and monellin (Mone),[12o] firefly luciferase,[12p] the GFP-like mutant Dronpa,[12q] and phytochrome.[12r]

Despite the limitations and drawbacks highlighted above, QM/MM models have provided valuable insight into many problems in photochemistry and photobiology (and, on a broader scale, in chemistry and biology in general). In several occasions, not only have they succeeded in explaining experimental observations, but they have predicted mechanistic features that were observed experimentally later on. Over the years, our group, together with others, have been using QM/MM models, with the QM part described by CASPT2//CASSCF level of theory, to consistently model a large and still growing number of photobiological systems. The results of the comparison of computed spectroscopic data with experiment are shown in Fig. 8.

In the next section, we present examples from the literature on how QM/MM models have been used for simulating spectral features and to understand how the environment may tune the photochemical and photophysical properties of a chromophore through non-covalent interactions.

Design Principles and Photobiology

Below, we focus on cases where the protein cavity appears to tune the excitation energy (see processes **a** and **b** in Fig. 2), the ES barrier (thus increasing or decreasing emission) or change the photochemical reactivity (see processes **1** and **2** in Fig. 2).

Case study 1: Colour tuning in retinylidene proteins. The Anabaena sensory rhodopsin (ASR) from the cyanobacterium *Anabaena* PCC 7120, is a member of a large family of proteins known as retinylidene proteins (RPs). RPs detect light using a retinal protonated Schiff base (rPSB) chromophore. rPSB typically exists in an all-*trans* form (see Fig. 9) in dark-adapted ASR and (typically) in microbial RPs, where it isomerizes to 13-*cis* upon photoexcitation. However, in other cases (such as visual RPs) rPSB exists in the 11-*cis* form and isomerizes to all-*trans* upon photoexcitation.

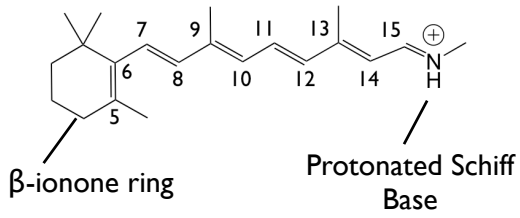


Fig. 9 The structure of the all-*trans* rPSB chromophore of ASR.

At its S_0 equilibrium structures, rPSB has its positive charge mostly localized near the Schiff base. Excitation to S_1 typically results in an increased delocalization of its positive charge along the π-conjugated framework, and therefore a part of the positive charge is transferred from the Schiff base moiety towards the β-ionone moiety (see Fig. 10A). Due to their difference in charge distributions, the $S_1 \rightarrow S_0$ excitation energies are sensitive to the surrounding electrostatic environment.[22] Therefore, a change in the protein environment (even a change as little as a single site mutation) may result in a RP that absorbs light of a different wavelength. The idea of changing the colour at which RPs absorb by altering its environment is known as *colour tuning* or *spectral tuning*. The mechanism of colour tuning in ASR is illustrated in Fig. 10B.[23] Due to the different electronic structure of the S_0 and S_1 states of rPSB, mutations that project a negative charge on the chromophore Schiff base region stabilize the S_0 state more than S_1, resulting in a blue-shift in the absorption, while residues that project a positive charge on the Schiff base region destabilize S_0 more than S_1, resulting in a red-shift in the absorption. A complimentary argument can also be made regarding a mutation near the β-ionone moiety. Computational studies have been used to understand and predict the effect of mutations on the λ^{a}_{max} of ASR. This is done by mapping the electrostatic potential of the protein on the rPSB chromophore, as shown in Fig. 11 for a red-shifting and a blue-shifting mutation. The mechanism of colour tuning has also been investigated extensively in other RPs (e.g. see work by Sekharan, Morokuma, and co-workers and by Fujimoto, Hasegawa, and co-workers[24])

Another interesting case of colour tuning has been explored in a recent study where researchers have used a bottom-up

approach to engineer an artificial set of RPs that mimic natural rhodopsins. Such RP mimics are based on the human cellular retinol binding protein II (hCRBPII), which normally hosts retinol in its cavity. Mutation of two amino acids in the hCRBPII sequence allows the binding of retinal to form rPSB. These artificial RPs are very different from natural RPs. Natural RPs are characterized by a seven α-helix transmembrane structure, while hCRBPII is soluble and dominated by a β-sheet structure forming a barrel shape. However, in spite of the structural differences, it was shown that colour tuning, using the same mechanism shown in Fig. 10 is possible in these proteins, which display a λ^{a}_{max} in the range of 425 to 644 nm.[25]

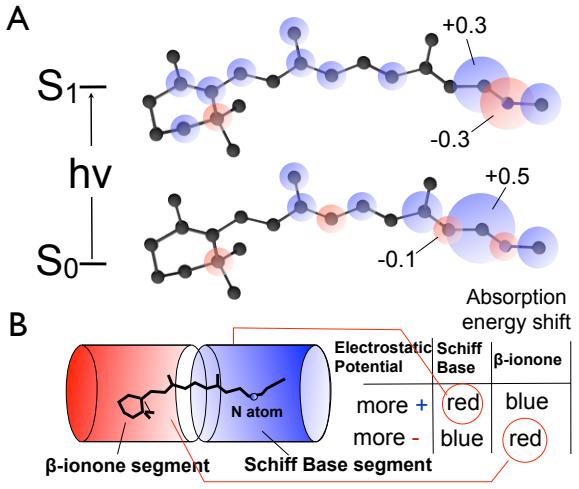


Fig. 10 Electrostatic effects upon photoexcitation in ASR. **A.** Change in the charge distribution along the backbone of the chromophore of ASR_{AT} upon photoexcitation from S_0 to S_1. The bubble diagram provides the value of the charge on the corresponding centre. The numbers in the diagram represent the minimum and the maximum values. **B.** A mutated ASR protein imposing a more positive potential (blue colour) on the chromophore Schiff base segment or a more negative potential on the β-ionone segment (red colour) will red shift the absorption energy. An inversion of the displayed electrostatic potential will lead to a blue shift. Adapted from ref. 23.

Case study 2: Turning on and off fluorescence in different environments. As illustrated in Fig. 2, fluorescence may occur when the ES PES displays a minimum, but not when an efficient decay channel is available (e.g. a barrierless path to a CI). The molecular environment surrounding a chromophore can affect these PES features and block, for instance, a decay channel (e.g. by steric constraints or electrostatic stabilization). One example is provided by the green fluorescent protein (GFP) chromophore. The GFP chromophore is highly fluorescent in the protein cavity while its close derivative displays no fluorescence in a solvent cavity at room temperature.[26,27] Combined ultrafast fluorescence spectroscopy and computational studies of the chromophore in solution and in the gas-phase respectively point to the existence of an ES deactivation pathway accessible via twisting about double and single bonds.[26-29] Such a pathway becomes inaccessible when the chromophore is bound to the protein environment.[26,28,29]

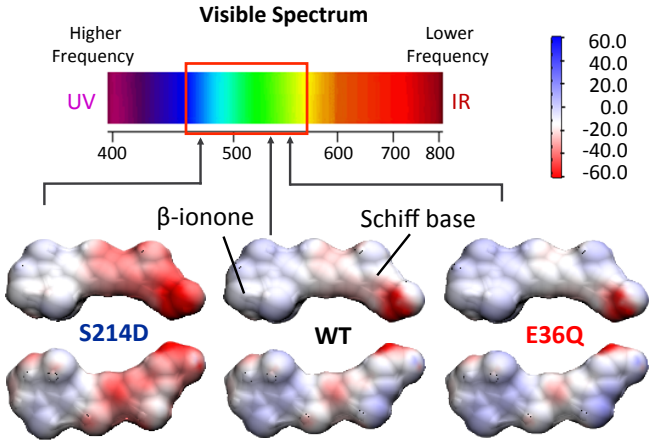


Fig. 11 QM/MM modeling based analysis of the predicted absorption energy change in 13-cis form of ASR upon a blue-shifting mutation (S214D) and a red-shifting mutation (E36Q). The electrostatic potential (in units of kT/e) imposed by the protein on the van der Waals surface of the chromophore for the wild-type (WT) and the mutants are shown. Adapted from ref. 23.

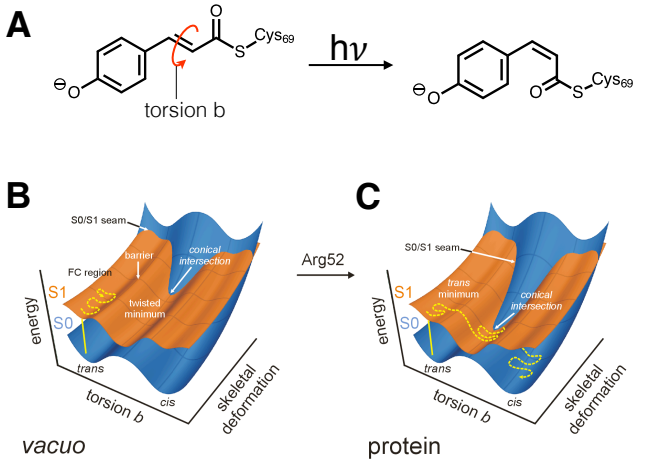


Fig. 12 A. The trans→cis photoinduced isomerization of the PYP *p*-coumaric acid chromophore (other possible isomerizations due to space-saving motions[31] are not shown here). B. Schematic representation of the PYP S0 and S1 PESs along the chromophore trans→cis isomerization coordinate (torsion b in part A) and a stretching deformation coordinate (a concerted length change of single and double bonds). The yellow line represents the average photochemical reaction path followed by the chromophore. Motion along torsion b connects the chromophore planar Frank-Condon geometry to a ca. 90° twisted intermediate. Non-radiative decay occurs at any point along the surfaces intersection hyperline. C. The same PESs for the gas-phase chromophore. By comparison with part B, notice that the electrostatic field of the protein stabilizes the S_1 state of the twisted chromophore, moving the seam closer to the twisted S_1 minimum also causing a decrease of the S_1 energy barrier, leading to the minimum. Adapted from ref. 30.

A second example is provided by the photoactive yellow protein (PYP); a photoreceptor of the eubacterium *Halorhodospira halophila*. Its chromophore is the anion of the *trans* isomer of a *p*-coumaric acid moiety that, upon photoexcitation, undergoes a double bond isomerization (see Fig. 12A). It is proposed that the isomerization leads to changes in the protein structure ultimately triggering a negative phototactic response in the organism. A computational study[30] of the PYP photoactivation has indicated a distinct behaviour of the chromophore in the protein cavity and in the gas-phase. In the protein cavity the chromophore encounters a small barrier along the ES isomerization coordinate (see Fig. 12C) and then reaches a CI from which it may decay back to the *trans* isomer or isomerize to the *cis* isomer. The isomerization is enhanced by the electrostatic stabilization of the negatively charged chromophore on the ES by a nearby positively charged protein residue, Arg52, which is located close to the phenyl ring on the chromophore. This stabilizes some regions of the ES PES and makes the CI more accessible. (see Fig. 12B and C) Moreover, due to the steric interactions with the surrounding residues, the chromophore isomerization occurs by a space-saving motion as experimentally observed in picosecond X-ray crystallography study.[31] In the gas phase, however, the CI is displaced far from the FC region and the barrier between the FC and CI points is higher (see Fig. 12B). This leads to an increase the ES lifetime and, therefore, to a higher fluorescence which is indeed observed. Note that this is in stark contrast to GFP where the protein environment induces higher fluorescence rather than quenching it.

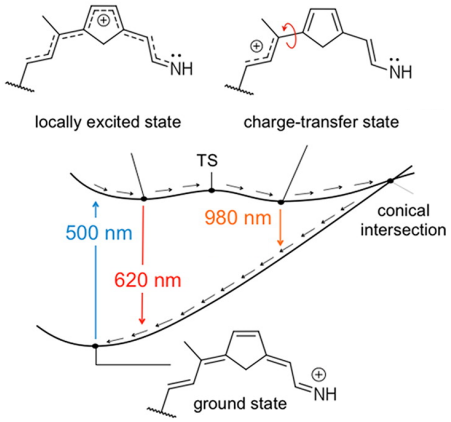


Fig. 13 Schematic representation of the origin of the 620 nm fluorescence λ^{e}_{max} generated by an analog of Rh incorporating an artificial 11-*cis*-locked chromophore. The three resonance formulas describe the electronic structure and charge distribution characterizing the S_0 and two different fluorescent states located on the spectroscopic (S_1) state of the protein. Fluorescence from the charge transfer state (CT) is hypothetical, whereas emission from the locally excited state (LE) has been observed experimentally. The distributed charge of the LE minimum is explained by stabilization via mixing of alternative charge transfer configurations tentatively associated higher excited states (not shown). The mixing is removed by double bond twisting towards the CT state. The LE→S_0 transition has a large oscillator strength and, therefore, does not seem to display a diradical S_2 state component. Figure adapted from ref. 3.

Another example of turning on fluorescence can be found in RPs. Biological evolution has shaped RP cavities to speed up and increase the efficiency of the photoisomerization of their chromophores. Nevertheless, using the design principles operating above, it appears to be possible to convert RPs into fluorescent proteins. An increase of the ES lifetime and fluorescence quantum yield has been indeed achieved in an unnatural bovine rhodopsin (Rh) featuring a chemically modified rPSB chromophore. Naturally, Rh (see also Fig. 7) has an 11-*cis* rPSB chromophore which photoisomerizes to the corresponding all-*trans* isomer by following a barrierless S_1 path leading to a CI (see path **1** in Fig. 2). As a consequence, it does not display fluorescence. However, by replacing the natural chromophore with a derivative where the 11-*cis* double bond is incorporated in a five-membered ring, the fluorescence quantum yield increases appreciably.[32] A CASPT2//CASSCF based QM/MM Rh model[3] indicates that: (i) locking the 11-*cis* bond creates a barrier along the S_1 C11=C12 double bond isomerization path which leads to an increased fluorescence lifetime, (ii) in such locked Rh the isomerization may be diverted to the C9=C10 double bond, (iii) along the isomerization path the charge transfer character increases in both rPSB and 11-*cis*-locked rPSB resulting in a displacement of the Schiff base positive charge towards the β-ionone side of the chromophore (Fig. 13), (iv) the fluorescent minimum in locked rPSB is located near the FC region and has a LE electronic character rather than a CT character (see Fig. 13 and its legend for label definitions), (v) the LE region also exist in the unlocked Rh although it is not associated with an S_1 energy minimum, and finally (vi) mutations that stabilize the LE electronic structure with respect to the rest of the S_1 surface (by stabilizing the delocalized positive charge in the central part of the chromophore) may deepen the corresponding S_1 minimum. Therefore, certain set of mutations may be able to deepen the S_1 minimum in the locked chromophore or create such a minimum in the natural chromophore, which would result in an increased fluorescence quantum yield. The latter scenario is interesting because it suggests that visual rhodopsins, which are light sensors, may be converted into fluorescent proteins. Indeed, fluorescent RPs that incorporate a natural rPSB have been shown to exist, and may operate under mechanisms similar to that outlined in Fig. 13 for artificial chromophores (see for instance the proteorhodopsin optical proton sensor[33] and archearhodopsin 3[34]).

Molecular Switches From Design Principles

One wonders if the different PES properties discussed above for proteins can be engineered in a light-responsive unnatural supramolecular system that may be ultimately prepared in the laboratory. Such challenge has been undertaken in our lab leading to the computer design and characterization of photoactive molecular components of such systems. These are light-driven switches based on a N-alkyl-indanylidene-pyrrolinium (NAIP) framework.[35,36] NAIP switches (from now on NAIPs) have also been prepared[37-39] and shown to display, in a solvent cavity, properties similar to visual pigments such as Rh.[40,41]

As schematically illustrated in Fig. 14A, a NAIP is formed by two cyclic units connected via a single exocyclic double bond forming an extended conjugated moiety. Such a moiety contains an alkylated Schiff base function similar, in electronic structure, to the protonated Schiff base chromophore of RPs. Thus, in principle, different features observed in such proteins (including Rh) may be mimicked through a suitable choice of substituents. For instance, a carboxylate counterion inserted in the pyrrolinium ring creates an intramolecular ion-pair mimicking the ion-pair between the protonated Schiff base and the Glu113 carboxylate counterion in Rh (see Fig. 14A and

14B[35]). Electron withdrawing and electron releasing substituents in the *para* and *ortho* position of the phenyl in the indanylidene unit would mimic the color tuning effects obtained by changing the amino acid composition of the retynilidene protein cavity (see Fig. 10 and 11). Finally, the conformational constrains imposed on the conjugated moiety by the two cyclic units, results in a single allowed isomerization mode (i.e. the isomerization of the exocyclic double bond) which mimics the highly selective 11-*cis* to all-*trans* isomerization imposed by the tight protein cavity of Rh (see Fig. 14B).

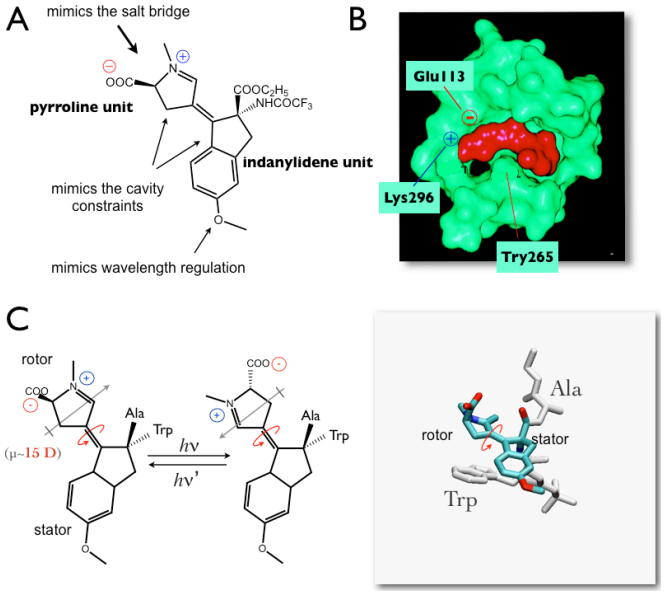


Fig. 14 A prototype biomimetic light-driven molecular switch. A. The structure of a NAIP switch.[35] B. Molecular cavity and PSB11 chromophore of the visual pigment Rh (see also Fig. 7). C. Photoisomerization of the "rotor" unit (the charged pyrrolinium ring) of the system. The light conversion of the E form (left) into the Z form (right) induces a dipole moment inversion when the re-orientation of the indanylidene stator is constrained (e.g. by incorporation in a peptide backbone). A tube model of a designed Ala-NAIP-Trp system is shown on the right.

NAIPs have been designed with the idea of creating an alternative to common light-driven molecular switches (e.g. those based on the azobenzene chromophore[43]) or, in the case of chiral NAIPs, to light-driven single-molecule molecular motors (e.g. those based on crowded alkenes[44]). However, NAIPs have also served another purpose. Indeed, these systems can be used to test how successfully QM/MM models can be employed for designing components of light-responsive supramolecular systems. For instance, it has been proposed that a NAIP featuring both amino and carboxylic substituents may be rigidly incorporated in a peptide backbone.[35] In Fig. 14C we present a case in which the indanylidene unit is incorporated in the peptide backbone as an unnatural α-amino acid (with a quaternary α-carbon) and, therefore, may be seen as a "stator" relative to the light-driven pyrrolinium "rotor". As illustrated in the same figure the rotor, corresponding to a zwitterion, will control the orientation of the system dipole and, in turn, may be used to control the properties (i.e. the conformation) of the supramolecular system hosting the switch.

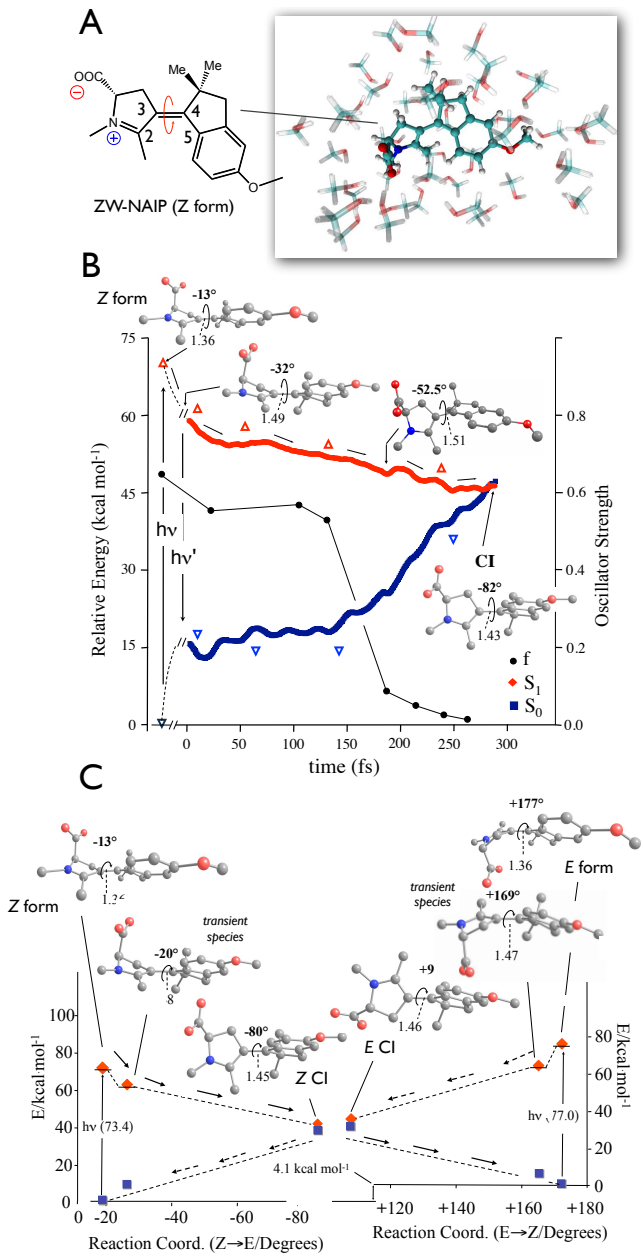


Fig. 15 Computational design of a zwitterionic NAIP (ZW-NAIP) working in methanol solution. A. The QM/MM model comprises a 20x20x20 Å^3 cubic box filled with methanol molecules representing the switch environment. B. Energy profiles (full symbols) along scaled-CASSCF/AMBER (i.e. with a QM subsystem treated at the scaled-CASSCF level and the MM subsystem treated using the AMBER force field. Scaled-CASSCF energies approximate CASPT2 energies)[35] trajectory simulating the ES Z→E isomerization event leading to decay at a conical intersection (CI). The system is predicted to reach the CI in less than 300 fs. Open symbols indicate CASPT2//CASSCF/AMBER (i.e. with a QM subsystem treated at the CASPT2//CASSCF level and the MM subsystem treated using the AMBER force field) energies, demonstrating that the scaled energy profiles qualitatively account for dynamic electron correlation. The CASSCF/AMBER value of the oscillator strength (f) along the trajectory is also given (full circles). C. ES (diamonds) and GS (squares) relative energies of ZW-NAIP structures characterizing its Z→E / E→Z photochromic

cycle. The values in degrees labeled on the structures refer to the torsional angle, while bond lengths are given in Å. The GS energy difference between the energy of the Z and E form is calculated using averaged solvent electrostatic potential / molecular dynamics (ASEP/MD)[42] computations accounting for the effect of the room temperature equilibrium average solvent electrostatics and configuration.

To support the conjectures above, QM/MM computer models of NAIPs in solution may be used to go one step forward along the design process. For this reason, we constructed a QM/MM model of the zwitterionic NAIP (ZW-NAIP) switch of Fig. 15A and put it at the centre of a solvent box representing a methanol solution. In the model we used a switch featuring methyl groups rather than the amino acid moiety of Fig. 14. The solvent molecules immediately surrounding the switch were relaxed during the simulations in order to interact with the solute. This model was then used to probe spectroscopic and reactivity properties. For instance, the λ^{a}_{max}, the transient emission and the detailed mechanism of the light-triggered isomerization process have been investigated via S_1 trajectory computations which also provided an estimate of the time-scale required for the event itself (see Fig. 15B). Notice that these properties are directly related to ES PES features summarized in Fig. 2 and therefore to the design principles discussed above.

A desirable feature of a supramolecular device is the existence of a "functional" state which can be turned on and off by irradiation with different wavelengths. This is, by definition, a basic property of light-driven molecular switches that feature two photochemically interconvertible states. Therefore, both the Z→E and E→Z light-induced isomerizations of ZW-NAIP have been computationally investigated. It was shown that after the less stable E form has been produced via irradiation of the Z form (following the isomerization dynamics of Fig. 15B), it may be reconverted to Z by irradiation with a more blue-shifted wavelength. The energetics associated to the Z/E photochromic cycle has been computed (see Fig. 15C). The predicted difference in excitation energies of the two isomers, is only 3.6 kcal mol^{-1} yielding just a few nm difference in the corresponding λ^{a}_{max} (389 and 371 nm for the Z and E form respectively). Both forms then evolve along the S_1 PES ultimately reaching distinct conical intersection points and relaxing to the photoproducts.

The synthesis of the ZW-NAIP of Fig. 15, and of other NAIPs with different substitution patterns,[37-39] has opened the possibility to compare the properties predicted using QM/MM models with spectroscopic measurements.[40,41] A set of data measured for ZW-NAIP in methanol is summarized in Fig. 16A. These include stationary absorption (with a λ^{a}_{max} of 397 and 391 nm for the Z and E form respectively) and transient absorption, fluorescence and infrared spectral data. Also, the photochromic behaviour of ZW-NAIP has been demonstrated via irradiation at monochromatic wavelengths of 340 nm and 455 nm yielding photostationary states characterized by Z/E ratio of 1:0.5 and 1:8.5 respectively. The data demonstrate a substantial agreement with the computed quantities and mechanism. In particular, both the Z→E and E→Z isomerization occurs on an ultrafast timescale. For the Z→E process the observed ES decay occurs with a timescale of 230 fs and the observed initial GS absorption appearance time is 380 fs. These times appear to be consistent with the computed 270 fs time required to reach the CI (see Fig. 15B and Table 1).

Table 1. Comparison between observed transient spectral data for Rh and two different switches (ZW-NAIP and MeO-NAIP).

	Rh	ZW-NAIP (Fig. 15A)	MeO-NAIP (Fig 15B)
Emission Decay	>100 fs	230 fs	200 fs
Product Appearance	200 fs	395 fs	330 fs
Vibrational Oscillations	60 cm^{-1}	60 cm^{-1}	70 cm^{-1}

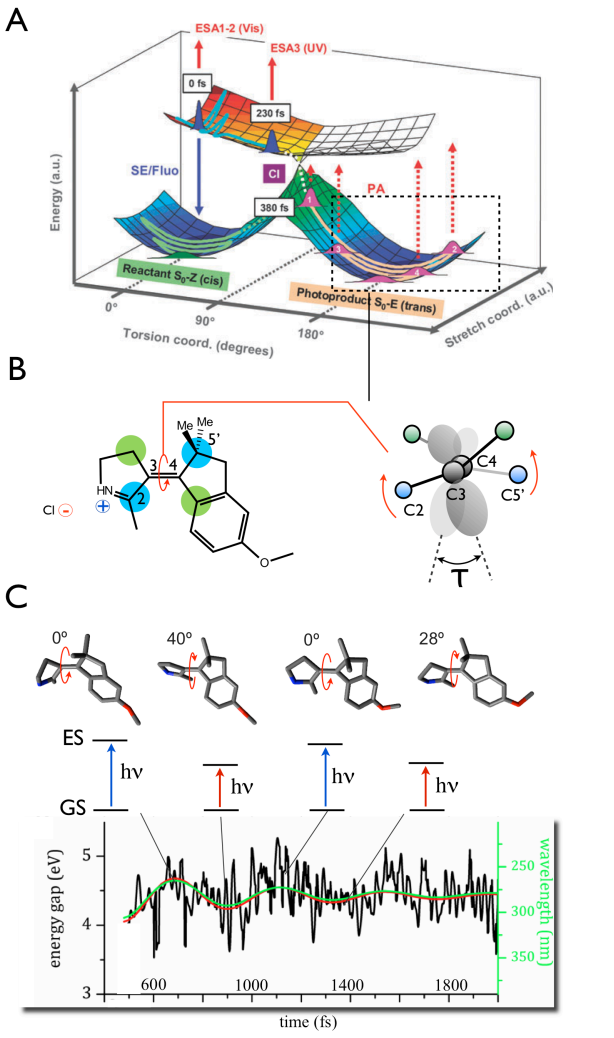


Fig. 16 Light triggered dynamics of a NAIP switch. A. Evolution and timing for ZW-NAIP proposed on the basis of time-resolved spectroscopic data. Immediately after photoexcitation, ES absorption (ESA) and stimulated emission/Fluorescence (SE/Fluo) are observed. While the molecules leave the fluorescent window, the ESA band spectrally shifts, reflecting motion along the torsional coordinate. After a spectroscopically dark period, associated with the passage through the conical intersection (CI), the molecules populate the GS of both isomers producing spectral modulations in the GS absorption signal (indicated by turning points 1 to 4, along the reaction coordinate). B. Qualitative relationship between the geometrical variable τ and the π-bonding for the switch MeO-NAIP (which is similar to ZW-NAIP but with the intramolecular counterion replaced by a solvated Cl-). C. Result of a 2000 fs semi-classical trajectory computation for MeO-NAIP. These oscillations

have a frequency which reproduce the damped oscillations observed in the experiment.[41]

The measured data support the idea that NAIPs mimic RPs. For instance, their ultrafast light-driven isomerization parallels the ultrafast isomerization observed for the visual pigment Rh, whose ES decays, as also stated above, in ca. 100 fs and has a primary photoproduct appearance time of ca. 200 fs. Most remarkably, studies carried out on a set of different NAIPs, have revealed that the switch of Fig. 15A and also the switch of Fig. 16 (where the intramolecular counterion is replaced by a solvated Cl^-), show a series of dumped oscillations in the transient absorption spectra immediately after ES decay. As shown in Table 1, 60 cm^{-1} modulations have been also observed in Rh where they were shown to be a consequence of the isomerization event.[45] The modulations are associated with coherent nuclear motion in the photoproduct population (see Fig. 16A): a phenomenon associated with the ability of NAIPs (and RPs) to funnel the absorbed light energy into one or few specific GS vibrational modes.

In spite of the interest that the generation of a vibrationally coherent motion raises when thinking to possible applications, the experimental observations do not provide information on the nature of the coherent motion. In order to address this issue, we have constructed a QM/MM model of MeO-NAIP (see Fig. 16B) that, due to a reduced computational cost, allows for the computation of a long semi-classical trajectory (ca. 2 ps). The time evolution of the excitation energy, starting 550 fs after photoexcitation, is displayed in Fig. 16C. It can be seen that the simulated GS→ES excitation energy changes, as a function of time, in a periodic fashion indicating that the system would display modulations in the absorption signal. Most importantly, the modulation period (ca. 500 fs) appears to be in line with the observed one (i.e. the 70 cm^{-1} frequency of Table 1 corresponds to a ca. 500 fs period). The analysis of the corresponding molecular motion along the trajectory in Fig. 16C, reveals energy gap changes correlated with the value of the dihedral angle τ (see Fig. 16B). The value of τ is, roughly, proportional to the overlap between the π-systems residing on the two cyclic units and, in turn, describing the extent of π-bonding across the central double bond. Therefore the oscillations in the ES-GS energy gap are assigned to the oscillation along a mode changing the extent of π-bonding, a quantity directly related to the molecular switching function.

Conclusions and Perspectives

In this review, we have attempted to equip the reader with design principles and provide information on the computational methods available for investigating supramolecular systems of the kind displayed in Fig. 1B. Since such methods have been extensively used to investigate photobiological systems, we have highlighted how their computer models could reveal the underlying design principles. Finally, we showed how the same models have been employed to develop a prototype biomimetic molecular switch displaying important similarities with RPs in a common solution environment. As the synthesis and the spectroscopic characterization of the system indicated, several biomimetic properties have been demonstrated especially through the comparison with the visual pigment Rh.

A first and obvious direction of future research must deal with the fact that while we have shown that it is possible to design and prepare functionalized NAIPs, the synthesis and characterization of supramolecular materials of the type illustrated in Fig. 14 has not been achieved yet. One route towards this goal has however already been opened by Sampedro, Woolley and coworkers[46] who have reported the preparation and characterization of a light-switchable grafted peptide. As shown in Fig. 17 such system is based on a benzylidene-pyrroline system analogue to NAIP switches that was originally synthesized in our lab.[47]

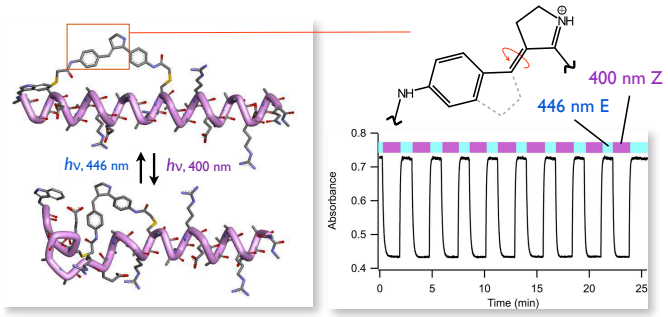


Fig. 17 Light-driven control of the conformation of a peptide. Left. Illustration of the cross-linked peptide with the photoswitch in the helix stabilizing E-form (top) and showing disruption of helical structure by the Z-form (bottom) of the photoswitch. Right. Structure of the phenylidene-pyrroline switch (top). The difference with a NAIP framework is shown as a dashed ring. Multiple rounds of photoswitching of the peptide (bottom) with alternating violet (400 nm) and blue (446 nm) light in aqueous solution at 20°C. Adapted from ref. 46.

A second, also obvious, direction for future research is the development of tools for the simulation of the light-induced dynamics of entire molecular populations. These developments will provide information on the quantum yield of the light-driven isomerization process in light-responsive proteins, with obvious impact in photobiology, and on the design of efficient artificial supramolecular systems. Semi-classical trajectories of the Tully surface-hopping type, are being applied to increasingly larger systems and to longer time scales even when the required ES forces are computed using expensive electronic structure theories such as MRCI.[48,49] This is not new, as semi-classical trajectories have been computed for some time, but now their popularity in photochemistry and photobiology is increasing as computer clusters and server farms become more and more easily accessible. Accordingly, semi-classical trajectories are rapidly replacing systematic static studies (e.g. mapping PES paths) even if static and dynamics studies are actually complementary. However, again, dynamic studies require an accurate potential energy surface to give meaningful results.

In Fig. 18 we report the data obtained by our group for a minimal model of the rPSB chromophore of Fig. 9.[10f] The data demonstrates the wealth of information obtained by computing a set of one hundred CASPT2 semi-classical trajectories representing the evolution of a population. Information such as: (i) average excited state lifetimes, (ii) average photoproduct appearance time, (iii) quantum yields as well as more sophisticated quantities such as (iv) the geometry and time distribution of ES decay points and (v) the time evolution of the charge distribution (i.e. of the electronic structure) may be represented in a single graph. It is hoped that, with rapidly evolving hardware technologies,[6,7] such information will be soon systematically available for QM/MM models of proteins, molecular devices in solution and supramolecular constructs.

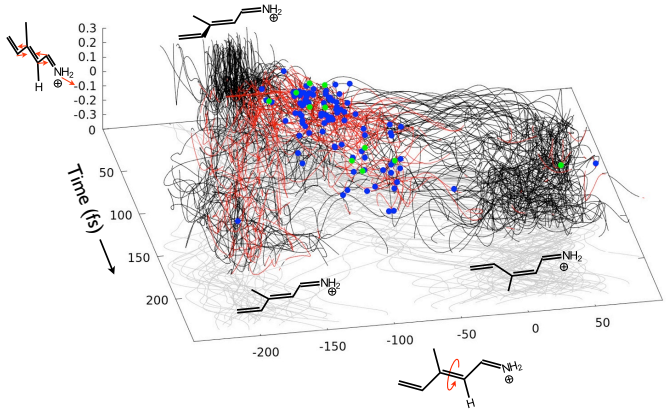


Fig. 18 Semi-classical trajectory computation for a reduced three double bond model of rPSB (shown in figure). The evolution of two key geometrical parameters, the stretching (vertical axis) and the torsional (horizontal axis) deformation of the carbon backbone associated to the double bond isomerization are reported as a function of time. The blue dots represent ES→GS hop points, The green dot GS→ES hop points and the red colour of certain segments of the trajectories represents regions where the positive charge has been translocated towards the hydrocarbon end of the backbone. This figure is based on trajectories reported in ref. 10f.

A third direction of development is the construction of faster single molecule rotary *motor*s (instead of just switches). Computational evidence for the existence of rotary motion in RPs has been provided for ASR the aforementioned microbial rhodopsin, which has been shown to display photochromic behaviour. The chiral cavity hosting its all-*trans* chromophore induces a positive helicity in the chromophore structure that, in turn, isomerizes unidirectionally (in contrast with visual rhodopsins).[12f] These findings suggest that chiral NAIP-like structures may provide a basis for the preparation of biomimetic single-molecule motors to be incorporated in novel, light-responsive supramolecular materials. Computational research in this direction has already started. For instance Sampedro, Frutos and coworkers[50] have reported on the computer design of a gas-phase fluorenylidene-pyrroline chiral system that is predicted to be capable of unidirectional isomerization more efficiently than the crowded olefin motors based on a stilbene-like framework. In our own group a genuine NAIP-based molecular motor, based on a chiral version of the MeO-NAIP structure, has being computationally designed and prepared in the laboratory.

Interest in computational photobiology and in the computational design of biomimetic photo-responsive systems should increase, as people start to look more and more towards light as a source of molecular-level energy, control, or information. These needs can only be fulfilled by continuing to expand the ability of traditional and emerging QM methods to treat ES and to be incorporated in protocols for constructing QM/MM or other type of multiscale models. The need for faster and more reliable algorithms and protocols for evaluating the GS and ES potential energies and their derivatives remains a considerable bottleneck for applications together with the overcoming of the limitations of the present QM/MM models. These remain challenges to theoreticians who are constantly struggling to overcome the problem of balancing quantitative accuracy and computational cost. New computer hardware technologies that are currently emerging may play an important role in the expansion of the field.

Acknowledgements

We are grateful to Alessio Valentini and Luis Manuel Frutos for discussion and for anticipating the results of Fig. 18. This work was supported in part by the National Science Foundation (NSF) under Grant CHE-1152070 and the Human Frontier Science Program Organization under Grant RGP0049/2012CHE09-56776. M.O. is also grateful to the Center for Photochemical Sciences and School of Arts and Sciences of the Bowling Green State University and to the FP7 grant PIOF-GA-2012-332233 PINADBio for further support. The European Cooperation in Science and Technology Action CM1002 (CODECS) is also acknowledged. The authors are indebted to the NSF-Extreme Science and Engineering Discovery Environment (XSEDE) and Ohio Supercomputer Center for granted computer time.

Notes and references

[a] Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43402, United States.
[b] Dipartimento di Chimica, Università di Siena, via De Gasperi 2, I-53100, Siena, Italy.

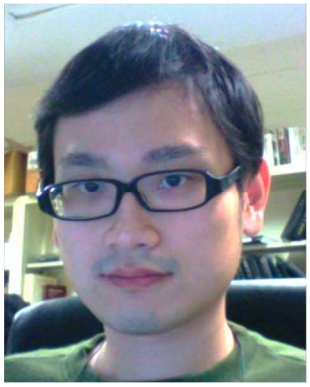
1. V. Balzani, *Supramolecular Photochemistry,* Springer, 1987.
2. F. Bernardi, M. Olivucci and M. A. Robb, *Chem. Soc. Rev.*, 1996, **25**, 321-328.
3. E. N. Laricheva, S. Gozem, S. Rinaldi, F. Melaccio, A. Valentini and M. Olivucci, *J. Chem. Theo. Comput.*, 2012, **8**, 2559-2563.
4. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227-49.
5. M. R. Silva-Junior and W. Thiel, *J. Chem. Theo. Comput.*, 2010, **6**, 1546-1564.
6. C. M. Isborn, N. Luehr, I. S. Ufimtsev and T. J. Martínez, *J. Chem. Theo. Comput.*, 2011, **7**, 1814-1823.
7. D. Lu, B. Xu, N. Xu, Z. Li, H. Chen, X. Peng, R. Xu and J. Du, *Phys. Chem. Chem. Phys.*, 2012, **14**, 9411-20.
8. A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009-37.
9. V. Veryazov, P. Malmqvist and B. O. Roos, *Int. J. Quant. Chem.*, 2011, **111**, 3329-3338.
10. (a) S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli and M. Olivucci, *J. Chem. Theo. Comput.*, 2012, **8**, 4069-4080; (b) Gozem, A. I. Krylov and M. Olivucci, *J. Chem. Theo. Comput.*, 2012, **9**, 284-292; (c) S. Gozem, F. Melaccio, R. Lindh, A. I. Krylov, A. A. Granovsky, C. Angeli and M. Olivucci, *J. Chem. Theo. Comput.*, 2013, **9**, 4495-4506; (d) M. Huix-Rotllant, M. Filatov, S. Gozem, I. Schapiro, M. Olivucci and N. Ferre, *J. Chem. Theo. Comput.*, 2013, **9**, 3917-3932; (e) X. Xu, S. Gozem, M. Olivucci and D. G. Truhlar, *J. Phys. Chem. Lett.*, 2012, **4**, 253-258. (f) S. Gozem, F. Melaccio, A. Valentini, M. Filatov, M. Huix-Rotllant, N. Ferré, L. M. Frutos, C. Angeli, A. I. Krylov, A. A. Granovsky, R. Lindh, and M. Olivucci. *Submitted for Publication.*
11. D. Roca-Sanjuán, F. Aquilante and R. Lindh, *WIREs Comput. Mol. Sci.*, 2012, **2**, 585-603.
12. (a) A. Altun, S. Yokoyama and K. Morokuma, *J. Phys. Chem. B*, 2008, **112**, 6814-27; (b) A. Altun, S. Yokoyama and K. Morokuma, *J. Phys. Chem. B*, 2008, **112**, 16883-90; (c) T. Andruniów, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 17908-13; (d) P. B. Coto, A. Strambi, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 17154-9; (e) A. Strambi, P. B. Coto, L. M. Frutos, N. Ferré and M. Olivucci, *J. Am. Chem. Soc.*, 2008, **130**, 3382-8; (f) D. Polli, O. Weingart, D. Brida, E. Poli, M. Maiuri, K.M. Spillane, A. Bottoni, P. Kukura, R.A. Mathies, G. Cerullo, and M. Garavelli,

Angew. Chem. Int. Ed., 2014, **53**, 2504-7; (g) S. Rinaldi, F. Melaccio, S. Gozem, F. Fanelli and M. Olivucci, *Proc. Natl. Acad. Sci. U.S.A*, 2014, **111**, 1714-9; (h) P. Altoè, A. Cembran, M. Olivucci and M. Garavelli, *Proc. Natl. Acad. Sci. U.S.A*, 2010, **107**, 20172-7; (i) A. Strambi, B. Durbeej, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. U.S.A*, 2010, **107**, 21322-6; (j) Work in progress in our lab; (k) A. Sinicropi, T. Andruniow, N. Ferré, R. Basosi and M. Olivucci, *J. Am. Chem. Soc.*, 2005, **127**, 11534-5; (l) P. B. Coto, S. Martí, M. Oliva, M. Olivucci, M. Merchán and J. Andrés, *J.Phys. Chem. B*, 2008, **112**, 7153-6; (m) C. Bernini, T. Andruniów, M. Olivucci, R. Pogni, R. Basosi and A. Sinicropi, *J. Am. Chem. Soc.,* 2013, **135**, 4822-33; (n) A. Udvarhelyi, T. Domracheva. Personal communication; (o) S. Pistolesi, A. Sinicropi, R. Pogni, R. Basosi, N. Ferré and M. Olivucci, *J. Phys. Chem. B,* 2009, **113**, 16082-90; (p) Navizet, Y. J. Liu, N. Ferré, H. Y. Xiao, W. H. Fang and R. Lindh, *J. Am. Chem. Soc.*, 2010, **132**, 706-12; (q) X. Li, L. W. Chung, H. Mizuno, A. Miyawaki and K. Morokuma, *J. Phys. Chem. B*, 2010, **114**, 1114-26; (r) P. Altoe, T. Climent, G. C. De Fusco, M. Stenta, A. Bottoni, L. Serrano-Andrés, M. Merchán, G. Orlandi and M. Garavelli, *J. Phys. Chem. B*, 2009, **113**, 15067-15073.

13. A. I. Krylov, *Annu. Rev. Phys. Chem.*, 2008, **59**, 433-62.
14. H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198-229.
15. N. Ferré and J. G. Ángyán, *Chem. Phys. Lett.*, 2002, **356**, 331-339.
16. J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549-64.
17. C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *J. Chem. Theo. Comput.*, 2009, **5**, 1838-1848.
18. M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357-19363.
19. T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma and M. J. Frisch, *J. Chem. Theo. Comput.*, 2006, **2**, 815-826.
20. M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theo. Comput.*, 2011, **7**, 525-537.
21. L. Zhang and J. Hermans, *Proteins*, 1996, **24**, 433-438.
22. A. Cembran, F. Bernardi, M. Olivucci and M. Garavelli, *J. Am. Chem. Soc.*, 2004, **126**, 16018-37.
23. F. Melaccio, N. Ferré and M. Olivucci, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12485-95.
24. Many studies have investigated colour tuning in RPs. See for example; S. Sekharan, A. Altun and K. Morokuma, *Annu. Rep. Comput. Chem. 7*, 2011, **7**, 215; K. Fujimoto, S. Hayashi, J. -Y. Hasegawa and H. Nakatsuji, *J. Chem. Theo. Comput.*, 2007, **3**, 605-618.
25. W. Wang, Z. Nossoni, T. Berbasova, C. T. Watson, I. Yapici, K. S. Lee, C. Vasileiou, J. H. Geiger and B. Borhan, *Science*, 2012, **338**, 1340-3.
26. S. R. Meech, *Chem. Soc. Rev.*, 2009, **38**, 2922-34.
27. H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi and F. I. Tsuji, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13617-22.
28. M. E. Martin, F. Negri and M. Olivucci, *J. Am. Chem. Soc.*, 2004, **126**, 5452-64.
29. P. Altoè, F. Bernardi, M. Garavelli, G. Orlandi and F. Negri, *J. Am. Chem. Soc.*, 2005, **127**, 3952-63.
30. G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. De Visser, H. J. Berendsen, M. Olivucci, A. E. Mark and M. A. Robb, *J. Am. Chem. Soc.*, 2004, **126**, 4228-33.
31. Y. O. Jung, J. H. Lee, J. Kim, M. Schmidt, K. Moffat, V. Srajer and H. Ihee, *Nat. Chem.*, 2013, **5**, 212-20.
32. H. Kandori, S. Matuoka, Y. Shichida, T. Yoshizawa, M. Ito, K. Tsukida, V. Balogh-Nair and K. Nakanishi, *Biochem.*, 1989, **28**, 6460-7.
33. J. M. Kralj, D. R. Hochbaum, A. D. Douglass and A. E. Cohen, *Science*, 2011, **333**, 345-8.
34. J. M. Kralj, A. D. Douglass, D. R. Hochbaum, D. Maclaurin and A. E. Cohen, *Nat. Methods*, 2012, **9**, 90-5.
35. A. Melloni, R. Rossi Paccani, D. Donati, V. Zanirato, A. Sinicropi, M. L. Parisi, E. Martin, M. Ryazantsev, W. J. Ding, L. M. Frutos, R. Basosi, S. Fusi, L. Latterini, N. Ferré and M. Olivucci, *J. Am. Chem. Soc.*, 2010, **132**, 9310-9.
36. A. Sinicropi, E. Martin, M. Ryazantsev, J. Helbing, J. Briand, D. Sharma, J. Léonard, S. Haacke, A. Cannizzo, M. Chergui, V. Zanirato, S. Fusi, F. Santoro, R. Basosi, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17642-7.
37. F. Lumento, V. Zanirato, S. Fusi, E. Busi, L. Latterini, F. Elisei, A. Sinicropi, T. Andruniów, N. Ferré, R. Basosi and M. Olivucci, *Angew. Chem. Int. Ed.*, 2007, **46**, 414-20.
38. L. Rivado-Casas, D. Sampedro, P. J. Campos, S. Fusi, V. Zanirato and M. Olivucci, *J. Org. Chem.*, 2009, **74**, 4666-74.
39. R. Rossi Paccani, D. Donati, S. Fusi, L. Latterini, G. Farina, V. Zanirato and M. Olivucci, *J. Org. Chem.*, 2012, **77**, 1738-48.
40. J. Briand, O. Bräm, J. Réhault, J. Léonard, A. Cannizzo, M. Chergui, V. Zanirato, M. Olivucci, J. Helbing and S. Haacke, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3178-87.
41. J. Léonard, I. Schapiro, J. Briand, S. Fusi, R. R. Paccani, M. Olivucci and S. Haacke, *Chem.: Eur. J.*, 2012, **18**, 15296-304.
42. I. Fdez Galván, M. L. Sánchez, M. E. Martin, F. J. Olivares del Valle and M. A. Aguilar, *Comput. Phys. Comm.*, 2003, **155**, 244-259.
43. G. Mayer and A. Heckel, *Angew. Chem. Int. Ed.*, 2006, **45**, 4900-21.
44. N. Koumura, R. W. Zijlstra, R. A. van Delden, N. Harada and B. L. Feringa, *Nature*, 1999, **401**, 152-5.
45. O. Weingart and M. Garavelli, *J. Chem. Phys.*, 2012, **137**, 22A523.
46. M. Blanco-Lomas, S. Samanta, P. J. Campos, G. A. Woolley and D. Sampedro, *J. Am. Chem. Soc.*, 2012, **134**, 6960-3.
47. D. Sampedro, A. Migani, A. Pepi, E. Busi, R. Basosi, L. Latterini, F. Elisei, S. Fusi, F. Ponticelli, V. Zanirato and M. Olivucci, *J. Am. Chem. Soc.*, 2004, **126**, 9349-59.
48. M. Barbatti, G. Granucci, M. Persico, M. Ruckenbauer, M. Vazdar, M. Eckert-Maksić and H. Lischka, *J. Photochem. Photobiol. A: Chem.*, 2007, **190**, 228-240.
49. M. Barbatti, *WIREs Comput. Mol. Sci.*, 2011, **1**, 620-633.
50. C. García-Iriepa, M. Marazzi, F. Zapata, A. Valentini, D. Sampedro and L. M. Frutos, *J. Phys. Chem. Lett.*, 2013, **4**, 1389-1396.

Samer Gozem was born in 1986 in Aleppo, Syria. He obtained his BSc in Chemistry (2008) from the American University of Beirut in Lebanon, and his PhD in Photochemical Sciences (2013) from Bowling Green State University in the US. There, he worked with Massimo Olivucci on investigating the thermal and photochemical isomerization of the chromophore of visual pigments in different protein environments by employing multiscale QM/MM methods. Currently, he is a postdoctoral fellow working with Prof. Anna I. Krylov at the University of Southern California.

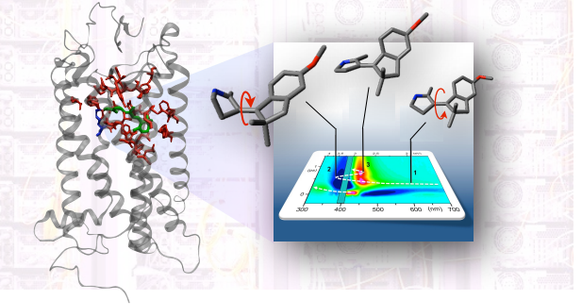
Federico Melaccio was born in 1984 in Poggibonsi, Italy. He obtained both his BSc in Chemistry (2006) and his MSc in Chemistry for Sustainable Development (2008) at the University of Siena, Italy. He obtained his PhD in 2011, working in the lab of Prof. Olivucci in the same university, on the computational simulation of spectroscopic properties of protein mutants. Currently, he is a postdoctoral fellow in the lab of Prof. Olivucci, working on the development of automated tools for computational chemistry simulations.



Hoi Ling (Calvin) Luk was born in Hong Kong. He attended the Ohio State University and received his B.S. degree in the spring of 2006 and PhD degree in summer of 2012 under the direction of Prof. Matthew S. Platz and Professor Christopher M. Hadad. Currently, he is a post-doctoral fellow working with Prof. Massimo Olivucci in Bowling Green State University focusing on the QM/MM computational study of Anabaena Sensory Rhodopsin and its mutants.

Silvia Rinaldi was born in 1985 in Orvieto, Italy. She received her MSc in Biotechnology from the University of Perugia in 2009, working in the nanotechnology field. She then got her PhD in 2013 working in the lab of Prof. Olivucci at the University of Siena. During this period her research interests focused on the computational modeling of photoreactive proteins (in particular, rhodopsin-like receptors) in order to study their spectroscopy and photoreactivity. Currently, she is a postdoctoral fellow at the Institute of Chemistry of Molecular Recognition (ICRM-CNR) in Milano, Italy.

Massimo Olivucci was born in Forlì, Italy. He received his PhD in 1989 with a Thesis in Theoretical Organic Chemistry at the University of Bologna, Italy with Fernando Bernardi, just before starting postdoctoral research with Michael A. Robb at King's College London, UK. He started his independent academic career in 1992, ultimately becoming full professor of Organic Chemistry at the University of Siena, Italy in 2001. In 2006 he was also appointed Research Professor at the Centre for Photochemical Sciences, Bowling Green State University, USA, where he is in charge of the Laboratory for Computational Photochemistry and Photobiology.



Learning how to model photo-responsive proteins may open the way to the design of light-powered biomimetic molecular devices

Five Key learning points

1. Learn a set of molecular engineering criteria based on the connection between the structure of ground and excited state potential energy surfaces and the spectroscopy (absorption and emission), photophysics (excited state lifetime and internal conversion) and photochemistry (selectivity) of desired light-responsive systems.

2. Learn the architecture of computer models suitable for the design of light-responsive organic supramolecular systems through a discussion of *ab initio* electronic structure theory methods and basic multiscale (QM/MM) protocols available in contemporary computing resources.

3. Learn about the biological and chemical information provided by QM/MM models of light-responsive proteins with emphasis on the control of color tuning, excited state lifetime and photochemical reactivity.

4. Learn how to computationally design and characterize a biomimetic molecular switch capable of mimicking, in common polar solvents, the spectroscopy and excited state dynamics of retinylidene proteins.

5. Learn current research trends in computational photochemistry and photobiology which may have an impact in the comprehension, simulation and design of light-responsive molecular materials.