

## TOPICAL REVIEW

# Theoretical description of protein field effects on electronic excitations of biological chromophores

**Daniele Varsano**

S3 Center, CNR Institute of Nanoscience, via Campi 213/A, 41125 Modena, Italy

E-mail: [daniele.varsano@nano.cnr.it](mailto:daniele.varsano@nano.cnr.it)

**Stefano Caprasecca**

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via G. Moruzzi 13, 56124 Pisa, Italy

E-mail: [stefano.caprasecca@for.unipi.it](mailto:stefano.caprasecca@for.unipi.it)

**Emanuele Coccia**

Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila, via Vetoio, 67100 L'Aquila, Italy & Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC Univ Paris 06, CNRS, 4 place Jussieu, F-75005 Paris, France

E-mail: [coccia@lct.jussieu.fr](mailto:coccia@lct.jussieu.fr)

**Abstract.** Photoinitiated phenomena play a crucial role in many living organisms. Plants, algae, and bacteria absorb sunlight to perform photosynthesis, and convert water and carbon dioxide into molecular oxygen and carbohydrates, thus forming the basis for life on Earth. The vision of vertebrates is accomplished in the eye by a protein called rhodopsin, which upon photon absorption performs an ultrafast isomerisation of the retinal chromophore, triggering the signal cascade. Many other biological functions start with the photoexcitation of a protein-embedded pigment, followed by complex processes comprising, for example, electron or excitation energy transfer in photosynthetic complexes. The optical properties of chromophores in living systems are strongly dependent on the interaction with the surrounding environment (nearby protein residues, membrane, water), and the complexity of such interplay is, in most cases, at the origin of the functional diversity of the photoactive proteins. The specific interactions with the surrounding environment often lead to a significant shift of the chromophore excitation energies, compared with their absorption in solution or gas phase. The investigation of the optical response of chromophores is generally not straightforward, from both experimental and theoretical standpoints; this is due to the difficulty in understanding diverse behaviours and effects, occurring at different scales, with a single technique. In particular, the role played by *ab initio* calculations in assisting and guiding experiments, as well as in understanding the physics of photoactive proteins, is fundamental. At the same time, owing to the large size of the systems, more approximate strategies which take into account the environmental effects on the absorption spectra are also of paramount importance. Here we review the recent advances in the first-principle description of electronic and optical properties of biological chromophores embedded in a protein environment. We show their applications on paradigmatic systems, such as the light-harvesting complexes, rhodopsin and green fluorescent protein, emphasising the theoretical frameworks which are of common use in solid state physics, and emerging as promising tools for biomolecular systems.

## 1. Introduction

The interaction between light and living organisms [1–3] currently represents one of the most puzzling and exciting research areas in the scientific community, because of its several connections with medicine, biology, chemistry and physics. Light is used by organisms to secure the energy needed by life processes, to accumulate information on the surrounding environment, and to establish a mutual interaction with other individuals.

Photoactive proteins uphold the absorption of light and trigger, at the molecular scale, the conversion of photon energy into chemical energy, by exploiting the optical properties of the chromophores, which are molecules embedded in the protein environment. The absorption properties of photoactive proteins are often determined by specific interactions between the embedded chromophores and the surrounding protein environment. Indeed, the functionality differentiation of such proteins, as well as their efficiency, closely depend on the fine-tuning mechanisms of the optical spectrum due to the environment.

In the peridinin-chlorophyll *a*-protein (PCP) complex [4], a water-soluble protein trimer deriving from marine dinoflagellate *Amphidinium carterae* algae, four identical peridinin molecules surrounding one chlorophyll absorb light at slightly different wavelengths, due to their different geometry and to the effect of the close environment in the active centre of the protein. This maximises the efficiency of sunlight absorption and of its conversion. Subtle differences in the geometry and environment of each peridinin also play a role in the energy transfer process to the nearby chlorophyll, and in the photoprotective triplet-triplet energy transfer [5–7].

The light absorption of photopsins, photoreceptors present in the cone cells of the retina, is strongly affected by the interaction between the chromophore and the protein environment. Different cone cells contain opsins differing in few aminoacids. These extremely localized changes in the primary structure are responsible for the different wavelength of light absorption of the photopsins as retinal bound pigments, giving rise to the colour vision in animals [8].

Another emblematic case is the well-known Green Fluorescent Protein (GFP) [9]. Thanks to the fluorescence properties and inertness when attached to other proteins, the use of GFP is nowadays an ubiquitous imaging tool for fluorescence microscopy in molecular biology [10]. From the wild-type GFP firstly isolated from the jellyfish *Aequorea victoria*, in the last two decades a large variety of mutants have been engineered, which are characterized by absorption energies and fluorescence peaks covering almost the whole visible spectrum. Also in this case, the protein environment plays a crucial role in the colour tuning [11].

With the advent of ultrafast time-resolved spectroscopy, the experimental study of photo-triggered reactions on a subpicosecond timescale has become accessible. In recent years, techniques such as transient absorption

and nonlinear optical spectroscopies [12] have been successfully applied to photoreceptors, unravelling biological reactions and providing information on structures, electronic properties and interaction with the solvent or the protein environments. These classes of experiments have permitted to model the reaction dynamics and to reveal energy- and electron-transfer pathways occurring in photoactive biosystems. Time-resolved spectra are very often difficult to interpret and theoretical assistance is needed to dissect the wealth of information contained. Electronic structure simulations are becoming routine tools to assist the experimentalists in understanding and rationalising data on electronic, dynamical and reactivity properties of molecules. A primary goal of electronic structure calculations is to achieve a detailed knowledge of the molecular mechanisms. A more ambitious and challenging aim would be to go beyond the experiments and be predictive. To achieve this, quantum-mechanical calculations, accurate and computationally affordable at the same time, are necessary. However, even considering the dramatic development of high-performance computing facilities, accurate first-principle calculations on biomolecular systems still represent an open challenge, because of the complexity of the algorithms of the more sophisticated computational methods, combined with the size of the systems of interest. Clearly, light-induced processes, as absorption and energy transfer are, always require a quantum treatment because of their intrinsically electronic nature; on the other hand, biological systems are composed of thousands or even millions atoms, making a complete quantum-mechanical description of the system impossible. However, since the region where the light absorption and the following reactions occur is usually limited to a small portion of the entire system, different strategies have been developed to overcome the problem, by introducing reasonable approximations that make calculations possible. In all the so-called *multiscale* methodologies, only a small sub-region of the protein (the active site) is treated using an accurate quantum-mechanical description, while a lower-level method is used for the remainder (the environment). The degree of accuracy and reliability of the theoretical description depends on various factors: first, on the quality of the high-level method used to describe the excitations; but also on the multiscale strategy adopted to take the effect of the environment into account. The choice of how the partition of the entire system into active site and environment is made is also important, and in many cases not at all straightforward; it can strongly affect the final results.

Many computational strategies have been developed to include environment effects on an active site, ranging from continuum models like the polarisable continuum model [13], quantum-mechanical embedding [14] and discrete hybrid quantum-classical model (QM/MM). QM/MM, introduced in 1976 by Warshel and Levitt [15], is probably the most popular multiscale strategy which allows the study of quantum phenomena, taking also into account the environment in a classical way. The QM/MM methods have been extensively developed and successfully

applied to study chemical reactions in proteins [16, 17]. The relevance of such development is testified by the 2013 Nobel Prize in Chemistry awarded to Warshel, Levitt and Karplus for “*the development of multiscale models for complex chemical systems*” [18–20]. Continuum methods, QM/MM methods and quantum-mechanic embedding schemes have been also extended to the treatment of excited states, thus enabling the possibility to investigate the effect of the environment on the light absorption and electronic energy transfer in biological chromophores.

The above mentioned strategies are based on various assumptions and account for the environmental effects in different ways, as it will be described in the following sections. As sketched in Fig 1, depending on the approximations adopted for the description of the environment, one can treat systems ranging from around 100 atoms, with fully quantum-mechanical methods, to entire proteins, up to  $10^6$  atoms, using classical force fields. It is important to stress that the accurate description of excited-state properties is computationally more challenging than the ground state, and that QM/MM schemes based on the point-charge representation of the protein environment could be insufficient for a proper description of the bathochromic shifts.

In order to obtain a reliable representation of the optical properties of biological chromophores, besides the essential need to include the environmental effects on the active part of the system, crucial attention has to be paid to the choice of the level of theory used to describe the optical excitation and the ground-state structure of the active site. For a large variety of chemical and biochemical systems, Density Functional Theory (DFT), thanks to its favourable computational cost and accuracy, is the most used tool for the calculation of ground-state properties of medium-sized to large molecules in all areas of chemistry, physics, and biology [24]. Due to its good performance, DFT is successfully applied within conventional QM/MM schemes and embedding methods and used as a quantum engine in *ab initio* molecular dynamics [25]. Unfortunately, no comparably reliable and cheap method exists for the calculation of the molecular excitations. Excited-state calculations from first principles are computationally cumbersome, and often characterised by a bad scalability with respect to the system size: this aspect limits the applicability of the most accurate quantum chemistry (post Hartree-Fock) methods to biomolecular systems. A good compromise between accuracy and computational cost for excited-state calculations is provided by time-dependent DFT (TDDFT) [26], which is nowadays the most applied framework for the calculation of optical properties of biochromophores. TDDFT is easily combined to molecular dynamics [27, 28] and also coupled to continuum schemes [29–31] and hybrid QM/MM [32] and QM/QM schemes [33].

Nevertheless, despite the great success of TDDFT for the calculation of absorption spectra and excited-state properties, various drawbacks exist, which are particularly serious when dealing with biological chromophores. These

include the poor description of long-range charge-transfer excitations [34], and the failure in describing states with double-excitation character [35].

The present topical review covers recent developments in the description of protein-field effects on optical excitations in paradigmatic biological chromophores, with a particular emphasis on the application of accurate computational frameworks that have been developed and successfully applied in solid state physics, such as the quantum Monte Carlo [36, 37] and many-body perturbation theory [38], which in the very last years have been proposed with success as promising tools for the study of biomolecular systems.

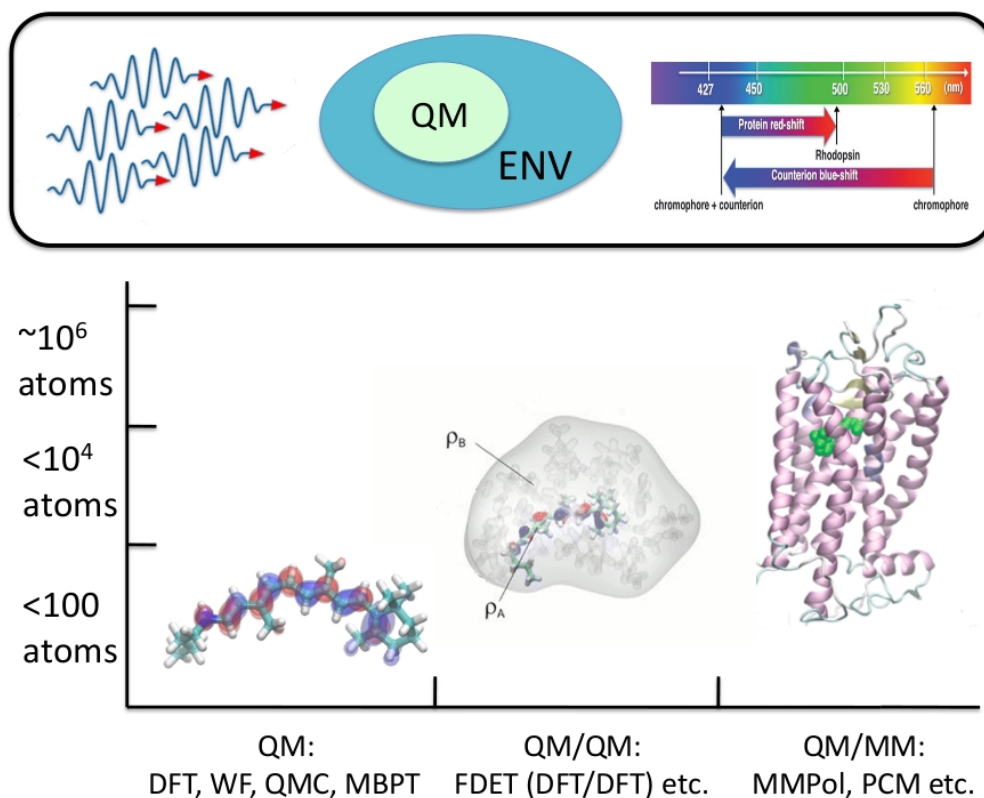
In Section 2 we introduce the most used methodologies available to deal with the structural electronic and optical properties of the active site, with a general description of the quantum Monte Carlo (2.2) and many-body perturbation theory (2.3). In Section 3 we report the different strategies to include environmental effects on the structures and excitation energies, introducing QM/MM methods (3.1), QM/QM methodologies (3.2), QM approaches for large systems (3.3) and techniques to take into account excitonic effects (3.4). In Section 4 we review recent efforts in the study of the environment effects on paradigmatic biochromophores, namely: light harvesting complexes (4.1), rhodopsin (4.2) and the Green Fluorescent Protein (4.3). In Section 5 we give our conclusions and final remarks.

## 2. Theoretical methods for the active site

The study of protein field effects on the electronic excitations of chromophores necessarily requires a quantum treatment of the chromophore itself, combined with an efficient but reliable description of the full protein system, which may include up to  $10^6$  atoms. For an appropriate representation of the vertical electronic absorption of the chromophore, the theoretical method needs to be chosen carefully, paying attention both to the description of the ground-state structure, and to the computation of the excited-state properties.

Since the application of any electronic structure method to the entire protein system is impractical, the first decision to make concerns the definition of a boundary, separating a smaller subsystem (the active site) from the environment. The active site comprises the optically active pigment, and eventually includes the surrounding residues that may participate in the excitation process, in order to avoid an unphysical description of the electronic properties. This subsystem is described by a high-level method (usually within the Born-Oppenheimer approximation), while the environment is commonly treated by classical force fields. Once this boundary is defined, the optical properties are extracted from excited-state calculations on the quantum subsystem, where environmental effects are taken into account in an approximate way.

Starting from crystallographic X-ray or NMR data, or from trajectories obtained by carrying out classical



**Figure 1:** Cartoon of different schemes to include the environment and typical system sizes for *in silico* light absorption of chromophores in complex protein systems. Adapted from Refs [21], [22] and [23] with permission from the American Chemical Society.

molecular dynamics simulations, the whole set of atomic coordinates is then refined within the chosen multiscale model. The criterion that must guide both the choice of the boundary, and the selection of the most appropriate multiscale model is the correct account of the electronic correlation, which is particularly essential for the nuclear and electronic structures of conjugated molecules.

In this Section the most widespread electronic-structure methods will be rapidly reviewed, pointing out their advantages and drawbacks when they are applied to the study of large molecules, starting from density and wave-function based methods. An account of quantum Monte Carlo methods and many-body perturbation theory will be given. The former are based on the stochastic sampling of the electronic coordinate space, and considered a “third way” in electronic structure calculations. For the latter, particular care will be devoted to the computation of optical excitations within the GW/Bethe-Salpeter framework.

### 2.1. Density and wave-function based methods

Accurate electronic and geometric structures of biochromophores are necessary to correctly interpret the exper-

imental observations on optically active proteins, and to rationalise them in terms of chromophore-environment interactions. In particular, the description of the electronic correlation is a fundamental ingredient to the precise account of chemical and physical processes. Many wave-function methods have been therefore proposed over the years in order to go beyond the simple approximation given by the Hartree-Fock approach [39, 40], which fails in treating systems where the electronic correlation plays an important role. Because of the size of the systems treated in this review, particular attention will be devoted to the search of a satisfactory trade-off between accuracy and computational effort, keeping in mind that the chromophores usually contain hundreds of electrons. For an all-encompassing presentation of quantum chemistry and density functional theory approaches, the interested reader is referred to the several textbooks available, for instance Refs. [26, 39–42].

Since the exact solution of the Schrödinger equation can not be obtained for the ground state of  $N$ -electron systems, a series of approximations must be carefully introduced. Wave-function methods are based on the explicit definition of an  $N$ -electron wave function

describing the system, according to a hierarchic addition (in terms of contribution to the total energy) of terms in the one-electron representation. Examples of wave-function methods are the configuration interaction (CI) [39], the Møller-Plesset perturbation theory [43–47] and coupled-cluster approaches [39, 40, 48, 49].

A conceptually different scheme to describe the electronic structure is represented by DFT [41], which employs the electron density as the key quantity to compute the properties of the system. In DFT, exchange and correlation effects are included via an exchange-correlation functional. Several approximations, based either on fits of experimental data or on theoretical models, define sets of functionals to be used in calculations on atoms, molecules and solids. Algorithms and exchange-correlation functionals are being continuously developed, in order to overcome known drawbacks and limitations [50], such as those concerning the proper description of dispersion forces and self-interaction problems [51]. Thanks to such development, DFT is approaching the “chemical” accuracy of  $\sim 1$  kcal/mol, and its favourable computational cost makes it the standard tool for medium-large size systems, well beyond the range of applicability of more accurate wave-function methods.

Electronic correlation is also essential for the investigation of excited-state properties. The calculation of optical properties of biochromophores also represents an open issue for theoreticians, because of the size of the molecular target. Excited-state methods, in analogy to the ground-state techniques, can be divided into wave-function-based and electron-density-based ones. The first class of approaches is based on the inclusion of excited configurations in the wave-function expansion, generated by promoting one electron from an occupied to a virtual orbital. The reference state is usually the Hartree-Fock ground state. For instance, CIS [52] and CISD [53] are characterised by the inclusion of singly and doubly excited configurations, respectively, and represent a simple approximation to the excited-state calculation. In the complete active space self-consistent field method (CASSCF) [54], the wave function is explicitly multiconfigurational. Perturbative corrections (at the second order) to CASSCF, improving the description of the dynamic correlation, are provided by the CASPT2 [43, 44, 55] and NEVPT2 [45–47] techniques. The parametrised DFT/MRCI [56, 57] explicitly exploits the advantages of both DFT and a multiconfigurational expansion of the wave function, while coupled-cluster [48, 49] derived approaches are based on a single-reference wave function. Such methods are very accurate but often too demanding in terms of computational cost for large chemical systems, containing hundreds of electrons [58].

As an alternative, semi-empirical techniques such as Zerner’s intermediate neglect of differential overlap (ZINDO)[59] or the modified neglect of diatomic overlap (MNDO) [60] can be applied to pigments of biological interest, at the cost of a loss of accuracy [61].

TDDFT [26, 42] can be considered in many cases as a reasonable trade-off between accuracy and computational

effort. A large variety of functionals are available. Several specific limitations of the method are known (see Ref. [58] for a complete discussion of theoretical issues and consequent repercussions on the study of biochromophores), mainly related to long-range charge-transfer excitations [62] and transitions with a double-excitation character [35]. Both problems can be encountered in the study of electronic excitations in pigments. Despite these known issues, TDDFT, both in its real-time [63] and linear response frequency domain (Casida equations) [64] implementation is widely employed for excited-state calculations of biochromophores. As in the case of the ground-state DFT, continuous development of exchange-correlation functionals is ongoing, and, even though the degree of accuracy is not comparable with its ground-state counterpart, some of the known problems, e. g., the partial failure in describing charge-transfer excitations, can be mitigated by employing hybrid and range separated functionals [65].

## 2.2. Quantum Monte Carlo methods

Quantum Monte Carlo methods (QMC)[36, 37] represent a valid alternative to the wave-function and density based methods for the description of the electronic structure of atoms, molecules and solids, since an explicit description of the electronic correlation is provided. QMC is characterised by a good scalability with respect to the system size ( $N^d$ , with  $3 < d < 4$ ,  $N$  being the number of electrons) [36, 66], comparable with that of DFT, and by the use of algorithms that can be massively parallelised [66], making them extremely suitable for the PetaScale architectures, as Blue Gene and hybrid CPUGPU machines. These ingredients justify the growing number of applications of QMC in the study of problems of chemical and physical interest. In particular, an accurate determination of structural and electronic parameters is required for the study of optical properties of biological chromophores.

The Variational Monte Carlo (VMC) method combines the Monte Carlo integration and a variational principle for the ground state energy. The VMC energy  $E_{\text{VMC}}$  [36] is computed as the minimum of the expectation value of the electronic Hamiltonian  $\hat{H}$ , over the variational parameters  $\mathbf{p} = \{p_i\}$  of a trial wave function  $\Psi_{\text{T}}$ , given a specific nuclear configuration  $\mathbf{R}$ :

$$E_{\text{VMC}} = \min_{\mathbf{p}} E[\Psi_{\text{T}}(\mathbf{x}, \mathbf{p}, \mathbf{R})], \quad (1)$$

where

$$E[\Psi_{\text{T}}] = \frac{\int \Psi_{\text{T}}(\mathbf{x}) \hat{H} \Psi_{\text{T}}(\mathbf{x}) d\mathbf{x}}{\int \Psi_{\text{T}}^2(\mathbf{x}) d\mathbf{x}}, \quad (2)$$

given a real  $\Psi_{\text{T}}$ . In the VMC framework, the integral above is written in terms of the local energy  $E_{\text{L}} = \hat{H} \Psi_{\text{T}} / \Psi_{\text{T}}$ , and of a probability density  $\frac{|\Psi_{\text{T}}|^2}{\int |\Psi_{\text{T}}|^2}$ :

$$E[\Psi_{\text{T}}] = \frac{\int \Psi_{\text{T}}^2(\mathbf{x}) E_{\text{L}}(\mathbf{x}) d\mathbf{x}}{\int \Psi_{\text{T}}^2(\mathbf{x}) d\mathbf{x}}. \quad (3)$$

The integral in Eq. 3 is computed as a sum over a set of points  $\mathbf{x}$  in the configurational space of the electronic Cartesian and spin coordinates, generated stochastically by the probability density  $\frac{|\Psi_T^2|}{\int |\Psi_T^2|}$ . Dependence of  $\Psi_T$  on  $\mathbf{p}$  and  $\mathbf{R}$  has been omitted in Eqs. 2 and 3 (and in the following) for sake of clarity.

The VMC estimate can be improved by adopting the fixed-node (FN) projection Monte Carlo techniques, which provide the lowest possible energy, with the constraint that the wave function  $\Phi_{\text{FN}}$  has the same nodal surface of an appropriately chosen guiding function [36, 67], that typically is the variationally optimised function  $\Psi_T$ . The most commonly used fixed node projection Monte Carlo methods are the diffusion Monte Carlo (DMC) [67] and the lattice regularised diffusion Monte Carlo [68] (LRDMC).

In any QMC approach for electronic structure calculations, the wave function  $\Psi_T$  is defined as the product of a fermionic term  $\mathcal{D}$ , providing the nodal structure, and a bosonic one, named Jastrow factor,  $\mathcal{J}$ , explicitly describing the interparticle correlation [36, 69]:

$$\Psi_T(\mathbf{x}) = \mathcal{D}(\mathbf{x}) \times \mathcal{J}(\mathbf{x}). \quad (4)$$

Various choices in literature are found for the  $\mathcal{D}$  and  $\mathcal{J}$  terms: we will briefly review here the main features of the CASSCF-like expansion and of the Antisymmetrised Geminal Power for the fermionic part of  $\Psi_T(\mathbf{x})$ .

*2.2.1. CASSCF-like wave function* In the CASSCF-like expansion of the wave function [36] the fermionic term is given by

$$\mathcal{D}(\mathbf{x}) = \sum_I c_I C_I(\mathbf{x}), \quad (5)$$

where  $C_I(\mathbf{x})$  are configuration state functions (CSFs) and  $c_I$  the linear coefficients of the expansion. Each CSF is a symmetry-adapted linear combination of (excited) Slater determinants of single-electron orbitals, which are expanded over a basis set of functions. Molecular orbitals and the set of  $c_I$  coefficients are obtained by a previous CASSCF calculation, but only a limited number of CSFs is taken into account for the QMC wave function, selected by the value of the corresponding  $c_I$ , since the dynamical correlation is described by the Jastrow factor. This approach is completely general, and can be applied to any spin symmetry.

*2.2.2. Jastrow Antisymmetrised Geminal Power* The Jastrow Antisymmetrised Geminal Power (JAGP) trial wave function, based on Pauling's resonating valence bond representation, [70–72] is defined as the product between the Antisymmetrised Geminal Power (AGP)

$$\mathcal{D}(\mathbf{x}) = \Psi_{\text{AGP}}(\mathbf{x}) \quad (6)$$

and a Jastrow factor  $J(\mathbf{r})$ , which only depends on the Cartesian electronic coordinates  $\mathbf{r}$ . For non-polarised molecular systems of  $N$  electrons ( $N/2 = N^\uparrow = N^\downarrow$ ) and  $M$  nuclei the AGP term is written as

$$\Psi_{\text{AGP}}(\mathbf{x}) = \hat{A} \prod_i \Phi_G(\mathbf{x}_i^\uparrow; \mathbf{x}_i^\downarrow) \quad (7)$$

where  $\hat{A}$  is the antisymmetrisation operator and  $\Phi_G$  the geminal pairing function for the electrons  $i$  and  $j$ :

$$\Phi_G(\mathbf{x}_i; \mathbf{x}_j) = \phi_G(\mathbf{r}_i, \mathbf{r}_j) \frac{1}{\sqrt{2}} \left( |\uparrow\rangle_i |\downarrow\rangle_j - |\downarrow\rangle_j |\uparrow\rangle_i \right). \quad (8)$$

The spatial function  $\phi_G(\mathbf{r}_i, \mathbf{r}_j)$  is a linear combination of products of atomic orbitals:

$$\phi_G(\mathbf{r}_i, \mathbf{r}_j) = \sum_{A,B} \sum_{\mu,\nu} \lambda_{\mu_A \nu_B} \psi_{\mu_A}(\mathbf{r}_i) \psi_{\nu_B}(\mathbf{r}_j) \quad (9)$$

where the indexes  $\mu_A$  and  $\nu_B$  refer to the basis sets centred on nuclei  $A$  and  $B$ , respectively.

In order to describe spin-polarised systems, it is necessary to introduce the so-called generalised Antisymmetrised Geminal Power (GAGP) wave function [73, 74]. Independent “molecular orbitals” built as linear combinations of atomic orbitals are added to the closed-shell AGP. This ansatz allows one to treat high-spin and radical states using the same computational protocol applied to the closed-shell singlet states [75]. It has been shown [69, 76] that the  $\Psi_{\text{AGP}}$  can be seen as a combination of a subset of even electronic excitations, using a multiconfigurational one-electron orbitals expansion. The AGP ansatz for singlet states corresponds to the subsector of the Hilbert space with seniority number  $\Omega = 0$  [77, 78]. The seniority number indicates the number of unpaired electrons in a given configuration, and therefore  $\Omega = 0$  corresponds to a closed-shell configuration. Such an ansatz can recover most of the static correlation, which plays a central role in conjugated [79, 80] and diradical [69, 75, 76] molecular systems.

*2.2.3. The Jastrow factor* The Jastrow factor is usually written as  $J = e^U$ , where  $U$  consists of several terms accounting for the 2-body, 3-body and 4-body interactions between the electrons and the nuclei [69, 81]:

$$U = U_{\text{en}} + U_{\text{ee}} + U_{\text{een}} + U_{\text{eenn}}, \quad (10)$$

where  $U_{\text{en}}$ ,  $U_{\text{ee}}$ ,  $U_{\text{een}}$  and  $U_{\text{eenn}}$  are the electron-nucleus, electron-electron, electron-electron-nucleus and electron-electron-nucleus-nucleus contributions. The leading contribution is the homogeneous two-electron interaction term  $U_{\text{ee}}$ , which only depends on the electron pair distance, and is used to satisfy the electron-electron cusp condition in all-electron calculations. The one-electron interaction term  $U_{\text{en}}$  describes the electron-nucleus correlation and satisfies the nuclear cusp condition for all-electron calculations. The  $U_{\text{een}}$  and  $U_{\text{eenn}}$  functions describe an inhomogeneous two-electron interaction, correcting the description introduced by  $U_{\text{ee}}$ . Furthermore,  $J$  is usually a spin-independent function to avoid spin contamination [69]. Functional forms of the various contributions can be found in Refs [69, 81].

*2.2.4. Optimisation methods* A crucial aspect in any QMC approach for electronic structure is the optimisation of the wave-function, according to the minimisation of the

total energy, of the variance or of a combination of the two [82], usually performed at VMC level.

At variance with what is routinely done in traditional quantum chemistry or DFT calculations, the full set of parameters is optimised, including linear coefficients and exponents of the atomic basis sets, together with the Jastrow parameters and CSF linear coefficients (if a CASSCF-like expansion is used). This approach leads to a rapid convergence of the variational results with the size of the basis set.

The most efficient procedure for the wave function optimisation exploits the powerful features of the linear method [83–85]. The basic idea is to expand the wave function  $\Psi_T(\mathbf{p})$  to the first order in the variational parameters  $\mathbf{p}$  around the current values of the parameters  $\mathbf{p}_0$  and to minimise the expectation value of the electronic Hamiltonian [84]. The parameter variations  $\Delta\mathbf{p} = \mathbf{p} - \mathbf{p}_0$  minimising the energy computed with the linearised wave-function are the lowest eigenvalues solving the generalised eigenvalue equation

$$\mathbf{H}\Delta\mathbf{p} = E\mathbf{S}\Delta\mathbf{p}, \quad (11)$$

where  $\mathbf{H}$  and  $\mathbf{S}$  are the Hamiltonian and the overlap matrices defined in the space of the wave function and of its first derivatives with respect to each  $\mathbf{p}$ .

Even more importantly for the topic of this review, an efficient computation of ionic forces within VMC has been recently achieved for electronic-structure calculations using the JAGP ansatz and the adjoint algorithmic differentiation (AAD) [66, 79, 80, 86–89]. The AAD allows one to automatically obtain exact derivatives of any complex function, provided that the dependence between variables is given [86]. Using AAD, the overall computational overload for calculating forces is only a factor four larger with respect to the single-energy evaluation, independently on the system size [86, 87].

The calculation of forces within VMC (using the standard steepest descent method or following directions maximizing the ratio between signal and noise) allows to obtain very accurate relaxed structures in systems where the correlation plays a crucial role, and standard DFT functionals are not accurate enough. Examples of large conjugated moieties optimised using this computational protocol are the retinal [88] and peridinin [89] chromophores, and polyacetylene chains [80]. In these systems, small variations in structural parameters such as the bond length alternation (BLA) have an important effect on the optical response. [88–90]. VMC ionic forces have been also extended to the molecular dynamics for the study of the properties of bulk water [91] and the phase transition of hydrogen in extreme temperature and pressure conditions [92, 93].

### 2.2.5. Excited states within quantum Monte Carlo

The extension of QMC to the excited states is not a straightforward task, since QMC is originally a family of ground-state methods. A general theoretical treatment of electronic excitations in the QMC framework is still missing, even though few works on the application of

QMC for excitations are present in literature: for instance, singlet and triplet energies for the benchmark  $\text{CH}_2$  diradical [94], and the low-lying singlet excited states of biochromophores [95, 96].

A systematic extension of QMC to the treatment of excited states would be highly desirable, for the same reasons already cited for ground-state properties, *i.e.*, a successful combination between a correlated ansatz (such as the JAGP), a good scaling with the system size, and the massive use of high-performance computing facilities. For instance, a formal analogy between the linear method for wave-function optimisation and linear-response theory exists. The generalised eigenvalue equation of the linear method (Eq. 11) coincides with the eigenvalue equation of linear-response theory in the Tamm-Dancoff approximation [94], corresponding to a CIS approximation for excited states. In the linear method, the highest eigenvalues of Eq. 11 can be interpreted as an estimate of excited-state energies [94]. The main difficulties arise from the necessity of defining improved estimators for the excited-state energies, to reduce the statistical error affecting them. Excitation energies can be also accessed by QMC using the the state-average scheme for excited states, based on the alternated optimisation of the linear coefficients of the CSF expansion and the optimisation of the linear and nonlinear coefficients of the orbitals and of the Jastrow factor [95, 96].

### 2.3. Absorption from many-body perturbation theory

An alternative approach to calculate fully *ab initio* charged excitations (photoemission spectra) and neutral excitations (absorption spectra) comes from Green’s function many-body perturbation theory (MBPT) [38, 97]. In particular, the GW approximation and Bethe-Salpeter formalism (BSE) [98, 99] have been developed in the last twenty years and successfully used to describe with high accuracy quasiparticle energies and optical excitations in several materials (bulk semiconductors, surfaces, 2D materials, polymers), including polyenic chains where TDDFT with local and semilocal approximations dramatically fails [100]. The BSE formalism has recently started being applied to the calculation of optical properties in gas-phase molecular systems [89, 90, 101–106], showing satisfactory results, comparable or even more accurate with respect to range-separated TDDFT calculations [65, 107–109]. Most importantly, GW/BSE has been shown to provide a reliable parameter-free description of both intra-molecular and inter-molecular charge-transfer and standard Frenkel excitations [101, 102, 105, 110, 111]. Concerning biological chromophores and biomolecules in general, GW/BSE showed a remarkable agreement with gas-phase experiments [89, 112, 113], and has been applied by Yin *et al.* [114] to study charge-transfer excited states in aqueous DNA, including explicit water molecules, and by Varsano *et al.* [90] to describe protein field effects on the retinal chromophore, by coupling it with a classical MM environment (see Sec. 4.2 for further details). In the following we recall the main procedure to obtain a GW/BSE absorp-

tion spectrum, and we refer to Ref. [38] for a theoretical review, and to Ref. [115] for technical details. Charged (electron addition/removal)  $n^{\text{th}}$  excitations, as measured by direct/indirect photoemission experiments, are naturally accessible in Green's many-body theory (also known as quasiparticle levels) and can be obtained as the solution of a generalised eigenvalue equation:

$$\left( -\frac{\hbar^2 \nabla^2}{2m} + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) \right) \psi_n(\mathbf{r}) + \int \Sigma(\mathbf{r}, \mathbf{r}', E^{\text{QP}}) \psi_n(\mathbf{r}') d\mathbf{r}' = E_n^{\text{QP}} \psi_n(\mathbf{r}), \quad (12)$$

where  $V_{\text{H}}(\mathbf{r})$  is the Hartree potential and  $\Sigma(\mathbf{r}, \mathbf{r}', E)$  is the nonlocal and energy-dependent self-energy operator which contains all the interactions beyond the Hartree contribution. The formal expression for the self-energy is given by a set of coupled Hedin's equations [98]. In the GW approximation the self-energy is expressed as:

$$\Sigma(\mathbf{r}, \mathbf{r}', E) = \frac{i}{2\pi} \int e^{-i\omega 0^+} G(\mathbf{r}, \mathbf{r}', E - \omega) W(\mathbf{r}, \mathbf{r}', \omega) d\omega \quad (13)$$

where  $G$  is the one-body Green's function:

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_n \frac{\psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}')}{E - E_n + i0^+ \text{sgn}(E - \mu)} \quad (14)$$

with  $\mu$  the chemical potential and  $W$  the dynamical screened Coulomb potential:

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int \frac{\epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega)}{|\mathbf{r}'' - \mathbf{r}'|} d\mathbf{r}'', \quad (15)$$

with  $\epsilon$  the dielectric matrix, calculated using the random phase approximation [116]. In practice, Eq. 13 can be solved starting from a DFT calculation, calculating the energy correction at first order, assuming the Kohn-Sham wavefunctions  $\psi_{\text{KS}}$  as a good approximation of quasiparticle states in the so called  $G^0W^0$  approximation, or adopting partial self-consistency in the eigenvalues and in the Green's function (scGW<sup>0</sup>), to fully self-consistent GW (scGW). In the last years the impact of the self-consistency or the starting point in  $G^0W^0$  approximation in molecular systems has been extensively studied [117–121]. The GW approximation then gives access to quasiparticle energies. In order to study absorption spectra, MBPT, using two-body Green's functions, permits to take explicitly into account the (screened) electron-hole interaction via the Bethe-Salpeter equation. The Bethe-Salpeter equation requires the quasiparticle energies and the screened interaction ( $W$ ) as calculated in the GW approximation and can be cast in an eigenvalue problem, providing the excitation energies by diagonalising an excitonic Hamiltonian defined in the basis of a two-body electron-hole state [38]:

$$\begin{pmatrix} R & C \\ -C^* & -R^* \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = E^{\text{exc}} \begin{pmatrix} A \\ B \end{pmatrix}, \quad (16)$$

where  $A$  and  $B$  are the coefficients of the excited state under study:

$$\psi^{\text{exc}}(\mathbf{r}_e, \mathbf{r}_h) = \sum_{\alpha\beta} A_{\alpha\beta} \psi_\alpha(\mathbf{r}_e) \psi_\beta(\mathbf{r}_h) + B_{\alpha\beta} \psi_\alpha(\mathbf{r}_h) \psi_\beta(\mathbf{r}_e) \quad (17)$$

having energy  $E^{\text{exc}}$ , where the indexes  $\alpha$  and  $\beta$  run respectively on the occupied and virtual orbitals, and  $\mathbf{r}_e$  and  $\mathbf{r}_h$  are the electron and hole positions. Here the products  $\psi_\alpha(\mathbf{r}_e) \psi_\beta(\mathbf{r}_h)$  and  $\psi_\alpha(\mathbf{r}_h) \psi_\beta(\mathbf{r}_e)$  indicate excitation and de-excitation respectively. In the excitonic Hamiltonian,  $R$  ( $-R^*$ ) are the resonant (antiresonant) terms between electron hole excitations, while  $C$  ( $-C^*$ ) are the couplings between resonant (antiresonant) transitions. Neglecting the coupling parts results in the often used Tamm-Dancoff approximation. The generic element reads:

$$R = D + 2K^{R,x} + K^{R,d} \quad (18)$$

$$C = 2K^{C,x} + K^{C,d}. \quad (19)$$

Here  $D$  corresponds to the quasiparticle energy differences between occupied and virtual orbitals calculated in GW approximation as explained before:  $D = (E_\alpha^{\text{QP}} - E_\beta^{\text{QP}}) \delta_{\alpha\alpha'} \delta_{\beta\beta'}$ . The bare exchange term and screened direct terms  $K^{R,x}$  and  $K^{R,d}$  of the electron-hole interaction kernel for the resonant transition read:

$$K_{(\alpha,\beta)(\alpha',\beta')}^{R,x} = \int \int d\mathbf{r} d\mathbf{r}' \psi_\alpha^*(\mathbf{r}) \psi_\beta(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_{\beta'}(\mathbf{r}') \psi_{\alpha'}^*(\mathbf{r}') \quad (20)$$

and

$$K_{(\alpha,\beta)(\alpha',\beta')}^{R,d} = \int \int d\mathbf{r} d\mathbf{r}' \psi_\beta^*(\mathbf{r}) \psi_{\beta'}(\mathbf{r}) W(\mathbf{r}, \mathbf{r}') \psi_\alpha(\mathbf{r}') \psi_{\alpha'}^*(\mathbf{r}'). \quad (21)$$

Similarly, the terms for the coupling part are given by  $K_{(\alpha,\beta)(\beta'\alpha')}^x$  and  $K_{(\alpha,\beta)(\beta'\alpha')}^d$ .

As for the GW approximation, the impact of the (partial) self-consistency in the quasiparticle energies plugged in Eq.21, and the exchange correlation functional in the underlying DFT calculation has been recently benchmarked [108, 109].

#### 2.4. Beyond PES minima points and temperature effects

An important aspect in modelling the absorption of photoactive proteins is to take into account the thermal fluctuations of the protein system, leading to structural distortions of the chromophore and of the surrounding environment, which affect the optical properties of the active site. Accounting for the dynamical properties of the protein allows one to go beyond the static model, *i.e.*, the direct use of crystallographic coordinates (kept fixed) [122], or the geometry optimisation within a QM/MM scheme (see Refs [21] and [88] as prototypical examples for the bovine rhodopsin).

The thermal sampling is usually included in the calculations by performing classical molecular dynamics (MD) or *ab-initio* QM/MM MD in which a large part of the protein is treated at the classical level of theory using force fields. Absorption calculations are then carried out



on structures extracted from the computed (semi-)classical trajectories. In such a way, one can compute the vertical excitations on an ensemble of representative structures and estimate the broadening of the absorption spectrum due to dynamical effects [23, 123, 124].

### 3. Introducing microenvironmental effects acting on the chromophore

A large number of strategies exist to represent the diverse effects that environments of different types have on the chromophores. In a very straightforward manner, one could simply include some atoms of the environment together with those of the chromophores, neglecting the rest. However, this approach, referred to as *minimal environment*, should be used with care, as it may fail in representing the whole effect of the surrounding, while greatly increasing the computational requirements of the calculation. A much better alternative is constituted by a large number of models, belonging to the wide family of *multiscale models* [18–20] (see Figure 2). The idea behind them is that a molecular system undergoing some process (such as electronic excitation) can often be studied reasonably well by explicitly considering only a small part of it (the active site, M), while the remaining part (the environment, E) does not actively take part in the process, but rather acts as a perturbation. When this is true, one can employ expensive methods, such as quantum-mechanical approaches, to appropriately describe the process undergone by the active site; this is possible owing to its relatively small number of atoms; by contrary, the environment can be treated at a lower level, provided that its effect on M is properly taken into account. This is particularly important in the cases reviewed here, since environment effects in protein systems can affect markedly the optical properties of the chromophores.

Before reviewing the main families of multiscale models available, it should be mentioned that the partition of the system into active site and environment may not be trivial, and it might have relevant effects on the description of the whole process. It is often sufficient to rely on one’s chemical intuition, but there are cases where a more accurate analysis is required. This is particularly true when there are chemical bonds between M and E, which therefore need to be artificially cut. In these cases, the definition of active site itself is subject to interpretation; one should make sure that the region where the process under study takes place is not affected by the cuts.

Among the multiscale models available, the so-called QM/classical ones are widely used because they combine simplicity, versatility and effectiveness. As the name suggests, the active site is treated using the electronic structure methods of quantum chemistry, while the environment is described classically. Alternatively, the environment may also be treated using quantum mechanical models, but a way must be devised to avoid performing one calculation on the whole system, which would be unfeasible.

The most relevant interaction term between M and

E is the electrostatic one: indeed, the electric field that the environment generates on the active site can polarise its electron density and subsequently modify its response, often markedly, depending on its strength. It is also possible to take into account the possibility for the environment to polarise as a response to the presence of M, thus changing the effect it exerts on it, but this second order polarisation term is more subtle and often less relevant. Repulsion interactions are particularly important in systems embedded, especially in protein matrices. Finally, hydrogen bonds represent a class on their own: when present, they constitute the strongest interaction term and can sometimes require a redefinition of the M/E boundary. Intermolecular hydrogen bonding can play a seminal role in many photophysical and photochemical processes, since a large organization of the interaction network can occur in the electronic excited state, as reported by Zhao and Han [125]. Light absorption by a hydrogen-bonded system triggers a change in the spatial and electronic structure of donor and acceptor molecules, due to the differences of charge distribution in the ground and excited states. If the intermolecular hydrogen bond is strengthened in the excited state, the hydrogen bond induces a red-shift to the absorption spectrum. In the case of intermolecular hydrogen bond weakening in the electronic excited state with respect to the ground state, the absorption spectrum is instead characterized by a blue-shift.

Prototypical examples are given by TDDFT studies of the formation of intermolecular hydrogen bond between photoexcited protochlorophyllide a and methanol [126], and the photoinduced electron transfer from alcoholic solvents to chromophores [127].

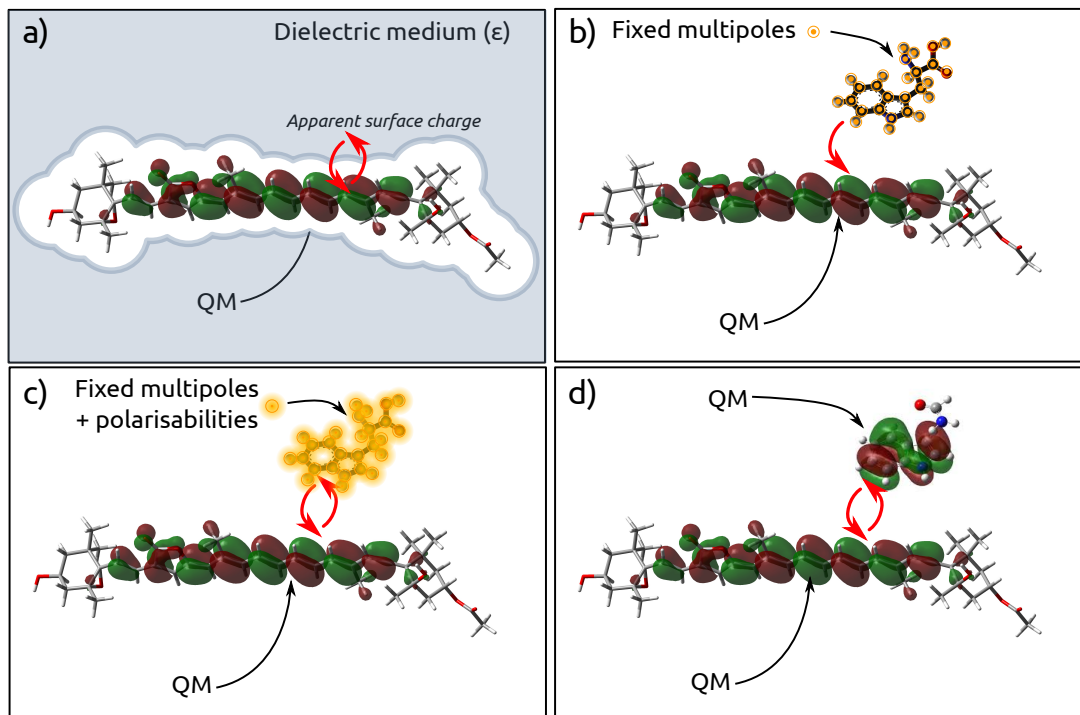
#### 3.1. QM/classical models

The QM/classical models resort to classical physics to represent the environment and its interaction with the QM system. Two families can be identified: the so-called QM/continuum models and the QM/discrete ones. In both cases, the electrostatic effect of the environment on the QM system M is taken into account through an explicit environment term  $\hat{V}^E$  in its Hamiltonian:

$$\hat{H} = \hat{H}^\circ + \hat{V}^E \quad (22)$$

where  $\hat{H}^\circ$  is the Hamiltonian of the isolated system. The expression of  $\hat{V}^E$  varies widely depending on the QM/classical model employed. Non-electrostatic interactions, by contrast, are not included in the Hamiltonian, and their effect is usually included as a purely classical correction to the obtained eigenvalues.

There are several types of continuum models available, among which the polarisable continuum model (PCM) developed in Pisa by Tomasi, of which various formulations exist [13, 128]; the Surface and Simulation of Volume Polarisation for Electrostatics model (SS(V)PE [129]); and the Conductor-like Screening Model (COSMO [130]). They resort to a description of the environment in terms of apparent surface charges; in the PCM model, for



**Figure 2:** Schematic representation of different approaches available to include the environment effects on the electronic properties of the active site, here represented as a peridinin. The environment can be represented at different levels: a) as a polarisable continuum; b) with non-polarisable multipoles; c) with multipoles and polarisabilities; d) at QM level. The red arrows point towards the parts of the system which are polarised by the other.

instance, the environment is described as a structureless dielectric medium, characterised by its dielectric constant. The active site is placed in a molecular cavity,  $\Omega$ , built using interlocking spheres centred on the atoms, which separates it from the surrounding medium. The polarisation of the latter as a response to the presence of M is then described in terms of an apparent surface charge spread over the molecular cavity. The determination of the surface charge requires the solution of a classical electrostatic problem, which is solved in practice by discretising the cavity into tiles (called *tesserae*), over which the set of charges representing the environment polarisation are placed:

$$\mathbf{T}(\epsilon, \Omega)\mathbf{q} = \mathbf{V}(\rho_M) \quad (23)$$

In the matrix equation,  $\mathbf{q}$  is the vector collecting the polarisation charges,  $\mathbf{V}$  the electrostatic potential generated by the electron density  $\rho_M$  of the subsystem M on the *tesserae*, and  $\mathbf{T}$  is the PCM matrix, which only depends on the cavity topology and on the environment dielectric constant  $\epsilon$ . The dimension of  $\mathbf{T}$  is  $N_{\text{tess}} \times N_{\text{tess}}$ , where  $N_{\text{tess}}$  is the number of tesserae. The polarisation charges obtained enter the Hamiltonian through  $\hat{V}^E$ , which, in this case, reads:

$$\hat{V}^E = \sum_i^{N_{\text{tess}}} q_i \hat{V}(\mathbf{r}_i) \quad (24)$$

where  $\mathbf{r}_i$  is the position of the  $i$ -th tessera and  $\hat{V}$  the electrostatic potential operator. The presence of the PCM

charges thus polarises the electron density  $\rho_M$ , on which in turn the charges depend. If the QM approach already makes use of an iterative solution (*e.g.* HF or DFT), then the mutual polarisation problem can be solved naturally within the SCF cycles.

The strong points that have made PCM, and continuum models in general, the preferred and most widely used choice when environment effects need to be accounted for, are firstly due to their computational inexpensiveness. The time required to solve the SCF problem with PCM is typically hardly more expensive than in vacuo. Even when large cavities are treated, and therefore large PCM matrices are formed, Eq. (23) can be solved with an iterative procedure, thus avoiding the calculation of the inverse matrix. A further advantage of continuum models is that they are able to fully characterise the average effect of the environment; for homogeneous solvents, one calculation is usually sufficient to account for the whole environment effect, and compared directly with the experimental results.

There is however one major shortcoming of continuum models, which is particularly important in the cases studied here: being the environment described as a structureless medium, all the *structured* interactions, depending on the particular configuration of the environment atoms, are neglected. A typical example is that of hydrogen bonds, but in general any non-isotropic environment may pose problems. Alternative approaches to continuum models are the discrete ones, where the structural

information of the environment is maintained [16, 17, 131–133]. The environment molecules are here represented with a classical force field, where electrostatic interactions are reproduced using a set of atomic multipoles. Non-electrostatic M/E interactions include van der Waals and bonded terms. The former are typically introduced with a Lennard-Jones potential, while the latter can pose some problems, and are tackled using various schemes, such as the link atom, boundary atom, and localised orbitals schemes [16]. In most QM/discrete implementations, only atomic point charges are used to reproduce the charge distribution of the environment molecule (*e.g.* ESP or RESP charges), although there are cases when multipoles up to the quadrupoles are considered [134]. Different QM/discrete schemes exist: if all M/E interactions are treated classically through the force field, the scheme is called *mechanical embedding* (ME). A more refined and widely used one, named *electrostatic embedding* (EE) scheme, by contrast, explicitly considers the electrostatic perturbation of the QM system due to the environment, whence a correction to the Hamiltonian stems:

$$\hat{V}^E = \sum_i^{N_{\text{MM}}} \left[ q_i \hat{V}(\mathbf{r}_i) + \boldsymbol{\mu}_i \cdot \hat{\mathbf{E}}(\mathbf{r}_i) + \dots \right] \quad (25)$$

where  $N_{\text{MM}}$  is the number of classical sites,  $\{q\}$  and  $\{\boldsymbol{\mu}\}$  are sets of atomic point charges and dipoles, and  $\hat{V}$  and  $\hat{\mathbf{E}}$  the electrostatic potential and electric field operators, which couple the QM and classical parts. It follows that the resulting electron density of M becomes polarised by the environment. This has proven to be of fundamental importance for a correct account of environment effects, particularly when biological environments are considered [16, 135]. Recent developments have also included the possibility for the environment itself to polarise as a response to the presence of M, analogously to what happens in PCM [133, 136, 137]. In the resulting *polarisable embedding* (PE) schemes, the environment polarisation can be represented in terms of induced dipoles [138–140], fluctuating charges [141] or Drude oscillators [142]. In the former approach, generally referred to as QM/MMPol, the classical atoms are also assigned an atomic polarisability, which induces a dipole as a response to the electric field:

$$\mathbf{K}(\boldsymbol{\alpha}, \mathbf{R}) \boldsymbol{\mu}^{\text{ind}} = F(\rho_{\text{M}}, \boldsymbol{\Theta}). \quad (26)$$

In the last equation (notice the similarity with eq. 23)  $\boldsymbol{\mu}^{\text{ind}}$  is a vector collecting the induced dipole moments,  $\mathbf{K}$  is the MMPol matrix, depending on the positions and polarisabilities of the classical atoms ( $\mathbf{R}$  and  $\boldsymbol{\alpha}$ ), and  $F$  is the electric field generated by the system M, and by the classical terms of the surrounding environment molecules ( $\boldsymbol{\Theta}$ ). This polarisable embedding scheme has been developed within coupled cluster theory [143], CASSCF [144], TDDFT [139, 140, 145, 146], and recently quantum Monte Carlo [147]. Ground state gradients have also been implemented [148], as well as hybrid fully polarisable QM/QM/MM model called polarisable density embedding (PDE) [149]. The possibility for the environment

molecules to self-polarise is forbidden in some schemes and allowed in others, where it may lead to the so-called *polarisation catastrophe*. To avoid this, the interactions between two classical sites may be scaled or screened, depending on their distance and connectivity [150]. On top of this, a damping of the QM/MM interaction can also be introduced to avoid overpolarisation [151, 152]. The different polarisation schemes also differ by their parametrisation of the sets of fixed multipoles and atomic polarisabilities. The induced dipoles obtained with eq. 26 enter the Hamiltonian through an additional polarisation term; the mutual polarisation between M and the environment is again solved in a self-consistent procedure, like for PCM. The increase in computational requirements that this implies, compared with non-polarisable treatments, is justified by the observation that the introduction of environment polarisation has shown to have decisive effects, particularly when electronic excitations in structured protein systems are studied [153], and is needed to recover agreement with full QM approaches [143, 154]. Attention however must be paid to the consistency between the QM treatment and the polarisability parametrisation [155].

The advantage provided by QM/discrete treatments in describing the short-range directional interactions is counterbalanced by the strong dependence of their results on the particular environment configuration chosen. In order to capture the average effect of the environment, a large number of environment configurations must be sampled. To do so, it is common to first carry out a MD for a certain amount of time, and perform QM/discrete calculations on certain snapshots [156]. This greatly increases computational requirements and time. By contrast, QM/continuum approaches immediately provide averaged effects. Mixed QM/discrete/continuum approaches, also fully polarisable, have been developed to fully take advantage of the ability of discrete and continuum treatments in representing short and long range environment effects, respectively [141, 157, 158].

When vertical electronic excitations are studied within a polarisable environment, it is necessary to take into account that non-equilibrium situations may arise. These are due to the fact that not all of the environment response will be able to instantaneously equilibrate to the excitation: in general, only some “fast” components of the environment polarisation, associated with its electronic degrees of freedom, will, whereas the “slow” components of the polarisation, associated with the environment nuclear degrees of freedom, will necessarily remain in equilibrium with the initial state.

In PCM, this is done in practice by considering a *dynamic* response of the environment, depending on the optical dielectric constant  $\epsilon_{\infty}$ , and an *inertial* response, which also depends on the static  $\epsilon$ . Equation (23) is therefore solved after building the matrix  $\mathbf{T}$  with the appropriate dielectric constant. In polarisable QM/classical models, by contrast, the separation is more natural, since the fixed multipoles can be directly associated with the nuclear environment polarisation,

while the polarisation introduced through the atomic polarisabilities, fluctuating charges or Drude oscillators is already the electronic one.

Different formalisms can be employed to introduce environment response to electronic excitations. Most notably, when DFT is used, it has been shown that two alternative formulations (the state-specific and linear-response ones — SS and LR), which are equivalent in vacuo, differ once a polarisable environment is introduced [31, 159, 160]. In particular, the relaxation of the environment polarisation, upon the electronic excitation, depends on the difference between initial and final electron densities,  $\Delta\rho$ , in the SS formulation. It has been shown that this recovers the correct electrostatic response [159, 161], and is especially indicated when excitations with a large charge-transfer character are being studied [162]. By contrast, in the LR formulation, the relaxation includes the dynamical response of the environment to the transition density,  $\rho^{\text{tr}}$ , and accounts for the dispersion interaction. Both SS and LR terms should ideally be included; the latter proves to be of fundamental importance in the study of excitations in the Green Fluorescent Protein, as illustrated below.

In the TDDFT LR scheme, which can be easily extended to TDHF, CIS or semiempirical treatments, the electronic excitations are determined by solving the Casida equation (Eq. 27), where  $(\mathbf{X}_n^\dagger \mathbf{Y}_n^\dagger)^\dagger$  are the transition vectors and  $\omega_n$  the corresponding excitation frequencies:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} = \omega_n \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix}. \quad (27)$$

The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are the Hessian of the electronic energy and contain an explicit term  $\mathbf{C}^{\text{E}}$  due to the polarisable environment. In the case of MMPol, this reads:

$$\mathbf{C}_{ai,bj}^{\text{E}} = - \sum_p^{N_{\text{MM}}} \left[ \int d\mathbf{r} \phi_i(\mathbf{r}) \phi_a^*(\mathbf{r}) \frac{\mathbf{R}_p - \mathbf{r}}{|\mathbf{R}_p - \mathbf{r}|^3} \right] \mu_p(\phi_j \phi_b^*) \quad (28)$$

where indices  $i, j$  ( $a, b$ ) run over occupied (virtual) molecular orbitals  $\phi$ , and  $p$  over the  $N_{\text{MM}}$  induced dipoles  $\mu$ .

### 3.1.1. Coupling QMC and MBPT to the environment

Within the classical description of the environment with fixed multipoles (Figure 2b), a electrostatic embedding QMC/MM interface has been recently implemented [88] in the TurboRVB package [72] for the ground-state geometry optimisation of the chromophore in presence of the external field given by the protein. The accurate QMC methodology, based on the JAGP ansatz, is coupled to a classical model, where the electrostatic potential is only due to point charges on the  $N_{\text{MM}}$  atoms. The approach used in the CPMD code [163, 164] to treat the electrostatic coupling between the electronic density of the QMC subsystem and the point charges describing the large MM subsystem has been followed. Using a simple additive scheme,

$$\hat{H} = \hat{H}_{\text{QMC}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QMC/MM}}, \quad (29)$$

with  $\hat{H}_{\text{QMC}}$  the full electronic Hamiltonian for the active site M and  $\hat{H}_{\text{MM}}$  the classical force field describing the environment E, the resulting boundary Hamiltonian  $\hat{H}_{\text{QMC/MM}}$  is defined by bonded ( $\hat{H}_{\text{B}}$ ) and non-bonded terms ( $\hat{H}_{\text{NB}}$ ):

$$\hat{H}_{\text{QMC/MM}} = \hat{H}_{\text{B}} + \hat{H}_{\text{NB}} \quad (30)$$

where  $\hat{H}_{\text{NB}}$  is explicitly given by

$$\hat{H}_{\text{NB}} = \sum_{i \in \text{MM}, j \in \text{QMC}} E_{\text{vdW}}(R_{ij}) + \quad (31)$$

$$\sum_i^{N_{\text{MM}}} q_i \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_i|} v_i(|\mathbf{r} - \mathbf{R}_i|) + \sum_{i \in \text{MM}, j \in \text{QMC}} \frac{q_i Z_j}{R_{ij}}$$

and  $\hat{H}_{\text{B}}$  by

$$\hat{H}_{\text{B}} = \sum_{i \in \text{MM}, j \in \text{QMC}} [E_{\text{bond}} + E_{\text{angles}} + E_{\text{dihedrals}}]. \quad (32)$$

$E_{\text{vdW}}(R_{ij})$  in Eq. 31 is the short-range van der Waals term, with  $R_{ij}$  the distance between nuclei  $i$  and  $j$ . The electrostatic potential of the QMC/MM border ( $\sum_i q_i \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_i|} v_i(|\mathbf{r} - \mathbf{R}_i|)$  in Eq. 31) is taken into account by including the interaction between the electronic density  $\rho(\mathbf{r})$  and the point charges  $q_i$  located at  $\mathbf{R}_i$ , where  $v_i(|\mathbf{r} - \mathbf{R}_i|)$  is a screening function to treat non-bonded short-range interactions [163]. The Coulomb term between the MM charges  $q_i$  and the effective charges  $Z_j$  on the QM nuclei ( $\sum_{i \in \text{MM}, j \in \text{QMC}} \frac{q_i Z_j}{R_{ij}}$  in Eq. 32) is also added. Terms in  $\hat{H}_{\text{B}}$  (Eq. 32) represent the the harmonic stretching ( $E_{\text{bond}}$ ) and bending ( $E_{\text{angles}}$ ) potentials, and the torsional potential ( $E_{\text{dihedrals}}$ ).

In the same spirit, MBPT calculations, namely GW/BSE, have been also extended to the QM/MM scheme in order to study the effect of protein electrostatic field on the optical properties of the retinal chromophore [90]. The MBPT/MM scheme has been implemented by interfacing the MBPT code Yambo [115] with CPMD [164]. In this framework Eq. 13 is modified by introducing the classical electrostatic field  $U^{\text{QM/MM}}$ :

$$\begin{aligned} & \left( - \frac{\hbar^2 \nabla^2}{2m} + U^{\text{QM}}(\mathbf{r}) + U^{\text{QM/MM}}(\mathbf{r}) + V_H(\mathbf{r}) \right) \psi_{\text{QP},n}^{\text{QM/MM}}(\mathbf{r}) + \\ & + \int \Sigma(\mathbf{r}, \mathbf{r}', E_{\text{QP},n}^{\text{QM/MM}}) \psi_{\text{QP},n}^{\text{QM/MM}}(\mathbf{r}') d\mathbf{r}' = E_{\text{QP},n}^{\text{QM/MM}} \psi_{\text{QP},n}^{\text{QM/MM}}(\mathbf{r}) \end{aligned} \quad (33)$$

Eq. 33 differs from a full-QM equation (Eq.13) in the expression of the external potential, that has been split in two contributions:  $U^{\text{QM}}$  is the external potential induced by the ions of the quantum subsystem, while  $U^{\text{QM/MM}}$  comes from the electrostatic coupling of the QM part with classical one. The coupling of the QM region with the environment is therefore directly included in the fundamental equations for the calculation of the quasiparticle eigenvalues and eigenvectors, finally affecting the quasiparticle energies and the absorption spectrum. Actually the ingredients needed for the construction of

the self-energy in the  $G^0W^0$  approximation are built using eigenvalues  $E_{\text{KS},n}^{\text{QM/MM}}$  and eigenvectors  $\psi_{\text{KS},n}^{\text{QM/MM}}$  coming from a coupled DFT QM/MM calculation. The  $\psi^{\text{QM/MM}}$  coming from the coupled QM/MM calculation together with the quasiparticle energies  $E_{\text{QP},n}^{\text{QM/MM}}$  calculated in the coupled scheme (Eq. 33) are used to build the Bethe-Salpeter effective excitonic two-body Hamiltonian of Eq.16. The main assumption for interfacing a MBPT approach with a classical force field is that the optical properties of the chromophore do not involve the electronic structure of the environment, so special care has to be devoted to the choice of the QM/MM partitioning.

The many-body GW formalism has been very recently coupled to the PCM [165] by Duchemin *et al.* extending the calculation of electron-addition and electron-removal energies of molecules in solution, and allowing to obtain quasiparticle energies renormalised by the effects of the ground state and non-equilibrium polarisation of the solvent. This method has been validated by calculating solvatochromic shifts with respect to the ionisation potential of nucleobases, resulting in very good agreement with  $\Delta\text{SCF}$  calculations (difference between the total energies of the neutral and ionised system) performed at DFT and coupled cluster level, with the advantage that, beyond the case of ionisation potential and electron affinity, the energy shift of all occupied/virtual energy levels can be obtained on the same footing. The development of Ref. [165] widens the capability of MBPT as a potentially very powerful tool for the calculation of electronic and optical properties of molecules in biological photoactive systems.

### 3.2. QM/QM embedding

The use of QM/QM embedding techniques [14,166] represents a step further in the study of complex protein systems. In QM/QM schemes, different quantum chemical methods are employed for different parts of the system. For instance, a highly sophisticated wave-function method can be used for a (small) subsystem, while the description of the environment is done at the less expensive DFT level. The same method can be also employed for partitioning the system using the same theoretical approach, as in the DFT/DFT embedding schemes, where different functionals are chosen for the subsystems. The key element in these methods is the embedding operator generated by the environment, which affects the embedded wave function or density. It couples the quantum-mechanical operators describing the active site to the environment.

In a DFT/DFT embedding scheme, the total electron density is partitioned, and the subsystem corresponding to the active site is polarised by the environment by means of an effective embedding potential, which only depends on the frozen electron density of the environment. This approach is called frozen density embedding theory (FDET) [14], and it will be briefly reviewed here in view of its application to the systems discussed in Section 4.

FDET provides a universal formalism that is in

principle exact and that, unlike QM/MM schemes, does not rely on any empirical parametrisation. It is based on the partition of the total electron density  $\rho_{\text{tot}}(\mathbf{r})$  into the electron densities of the two subsystems I and II:

$$\rho_{\text{tot}}(\mathbf{r}) = \rho_{\text{I}}(\mathbf{r}) + \rho_{\text{II}}(\mathbf{r}). \quad (34)$$

The total energy of the system (I+II) becomes a bifunctional of both  $\rho_{\text{I}}(\mathbf{r})$  and  $\rho_{\text{II}}(\mathbf{r})$ , and nonadditive terms arise for the exchange-correlation and for the kinetic functionals [14]. The electron density  $\rho_{\text{I}}(\mathbf{r})$  in subsystem I is obtained by minimising the total energy with respect to  $\rho_{\text{I}}(\mathbf{r})$ , with  $\rho_{\text{II}}(\mathbf{r})$  kept frozen in the subsystem II. Using the Kohn-Sham formalism and the constraint that the number of electrons in the subsystem I is conserved, the resulting Kohn-Sham equations are characterised by an unknown effective potential which is related, in a noninteracting system, to an electron density corresponding to that of the interacting system obtained by minimising the energy bifunctional. The Kohn-Sham effective potential can be written as the sum of the Kohn-Sham potential of the isolated subsystem I and the embedding potential depending on  $\rho_{\text{I}}(\mathbf{r})$ , and  $\rho_{\text{II}}(\mathbf{r})$

$$v_{\text{eff}}^{\text{FDET}}[\rho_{\text{I}}(\mathbf{r}), \rho_{\text{II}}(\mathbf{r})](\mathbf{r}) = v_{\text{eff}}[\rho_{\text{I}}(\mathbf{r})](\mathbf{r}) + v_{\text{eff}}^{\text{emb}}[\rho_{\text{I}}(\mathbf{r}), \rho_{\text{II}}(\mathbf{r})](\mathbf{r}). \quad (35)$$

$v_{\text{eff}}[\rho_{\text{I}}(\mathbf{r})](\mathbf{r})$  contains the nuclear, Coulomb and exchange-correlation potentials, while  $v_{\text{eff}}^{\text{emb}}[\rho_{\text{I}}(\mathbf{r}), \rho_{\text{II}}(\mathbf{r})](\mathbf{r})$  describes the interaction of the subsystem I with the frozen density and nuclei of subsystem II.

The embedded active region M (conventionally the subsystem I) usually comprises the chromophore and all the important residues needed for a balanced description of the electron density. Reasonable approaches are used for computing  $\rho_{\text{II}}(\mathbf{r})$  (representing the environment E), including approximated DFT calculations. FDET allows one to carry out large-scale quantum chemical calculations at the density functional theory level, since the computational effort is drastically reduced by treating the surrounding residues as a frozen electron density. The modelling of the chromophore-protein interaction is therefore proposed at full quantum mechanical level, using system sizes with  $\sim 400$  atoms [22]. Excited-state methods can be coupled to FDET to perform absorption calculations, as done for the rhodospin using the linear-response TDDFT [22].

Neglecting the dynamic response of the environment ( $\rho_{\text{II}}(\mathbf{r})$  is kept frozen also in the response framework) could lead to not negligible effects in FDET/TDDFT calculations [167–169].

### 3.3. Subsystem DFT

The FDET is a special case of partitioning the complex molecular target into small subsystems within DFT [170, 171].

Generally, the basic idea of the subsystem DFT is to fragment the system on the basis of the corresponding densities

$$\rho(\mathbf{r}) = \sum_I \rho_I(\mathbf{r}). \quad (36)$$

Treating subsystems separately reduces the overall scaling in the calculations, and consequently the computational effort, at the expense of (usually) noticeable approximations. The subsystem densities are represented through molecular orbitals  $\{\phi_i\}$  of noninteracting particles for each subsystem [171].

The main issue of this approach is that the noninteracting kinetic energy is no longer defined as in the Kohn-Sham method, since  $T_s$  requires knowledge of the molecular orbitals  $\{\phi_i\}$  of the full system. As a first approximation,  $T_s$  can be estimated by the sum of all subsystem noninteracting kinetic terms [171]

$$T_s[\{\phi_i\}] \sim \sum_I T_s[\{\phi_i I\}]. \quad (37)$$

The energy is minimised with respect to the Kohn-Sham orbitals of each subsystem  $I$ , with the constraint that the other electron densities are kept fixed, leading to Kohn-Sham equations for each fragment. The embedding potential coming from this coupled minimisation depends on nonadditive exchange-correlation and kinetic terms, as already stated for FDET. Approximate models for these potentials should be carefully chosen and benchmarked in order to get reliable results, see Ref. [171] for a detailed review of the various attempts present in literature.

The embedding potential is exact in the limit of exact functionals. However, the electrostatic term, which is usually the dominant one in the chromophore-environment interaction, is always exact in the subsystem DFT (and FDET) formulation.

From the energy minimisation point of view, FDET can be seen as an approximation of the general subsystem DFT, since the variational freedom (within Levy's DFT approach) is only restricted to the active site density in presence of a frozen environment density. As already pointed out, in the standard FDET the dynamic response of the environment is missing. Such a description is reasonably accurate for localised excitations of single chromophores, but it could lead to only a poor representation of excitonic effects. Beyond the neglect of the environment response framework, a general formalism within the response TDDFT theory has been introduced by Neugebauer to compute excitations delocalised over several subsystems, physically corresponding to excitonic transitions of strongly coupled monomeric pigments, as found in the light harvesting complexes [168] (see Section 3.4 for a more detailed introduction to excitonic effects in protein absorption).

Chemical intuition is supposed to be used for a proper partition of the molecular target. Formally, when an approximate functional for the nonadditive kinetic energy is employed, the many ways in which the system can be fragmented are not equivalent. One can overcome this difficulty by imposing that the embedding potential is the same for each subsystem [172, 173].

### 3.4. Environment effects in multichromophoric systems

In multichromophoric systems, the excitation energies of the chromophores (*site* energies) represent only one of the

three independent ingredients which is necessary to take into account. Indeed, depending on the particular system under study, two more factors can be relevant, namely the electronic coupling between chromophores, and the dynamical chromophore-protein interaction. While in the modelling of site energies and couplings the environment can be introduced as a perturbation, the latter term is an explicit coupling between the motions of environment and chromophores. These terms allow the study of coherences, which have been observed to play a relevant role in light-harvesting pigment-protein complexes. Note that the term *coherence* can signify both the quantum notion of a delocalised excited state arising from a superposition of localised excitations; and the classical notion of temporal and spatial correlations of spectral components, arising from vibrational motions [174].

The electronic coupling is an important parameter when an excitonic treatment of a multichromophoric system is carried out, and also determines energy transfer rates between chromophores. In this case, the Hamiltonian is expressed in the basis of the local excitations  $\{|n\rangle\}$  as:

$$\hat{H} = \sum_n E_n |n\rangle \langle n| + \sum_{m,n} V_{mn} |m\rangle \langle n| \quad (38)$$

where  $E_n$  are the localised (site) energy, and  $V_{mn}$  the electronic couplings.

There are various approaches to determine the couplings, and to estimate how the environment can affect them. In the point-dipole approximation (PDA), the coupling between two chromophores A and B is described in terms of an interaction between their transition dipoles:  $V_{AB} = \mu_A^{\text{tr}} \mu_B^{\text{tr}} \kappa R_{AB}^{-3}$ , where  $R_{AB}$  is the distance between them and  $\kappa$  is an orientation factor. In general, the PDA fails when the inter-chromophore distance is small, and the dipole approximation is no longer valid. More reliable extensions consist in representing the transition density by means of atomic charges (and higher multipoles), computed in different ways [175–177]. Alternatively, the full definition of Coulomb coupling in terms of transition densities can be used [178–180]:

$$V_{AB}^{\text{Coul}} = \int d\mathbf{r}_1 d\mathbf{r}_2 \rho_A^{\text{tr}}(\mathbf{r}_1) r_{12}^{-1} \rho_B^{\text{tr}}(\mathbf{r}_2). \quad (39)$$

A full-QM approach to the definition of coupling has been developed within the frozen-density embedding method of subsystem DFT [14, 168], where the coupling between two excitations  $K_A$  and  $L_B$ , localised on subsystems A and B, respectively, is computed as:

$$V_{K_A L_B} = \int d\mathbf{r} \sum_{(ia)_A} 2U_{(ia)_A, K_A} \omega_{(ia)_A}^{1/2} \phi_{i_A}(\mathbf{r}) v_{A, L_B}^{\text{ind}}(\mathbf{r}) \phi_{a_A}(\mathbf{r}) \quad (40)$$

where indices  $i$  and  $a$  refer to occupied and virtual orbitals, respectively;  $\omega$  is their energy difference; the subscripts A and B indicate the localisation on the two subsystems;  $U$  contains the eigenvector for the transition  $K_A$  on A, obtained in the uncoupled subsystem TDDFT calculation;  $v_{A, L_B}^{\text{ind}}$  the potential induced on A by the excitation  $L_B$ .

The presence of an environment can affect the electronic coupling both indirectly, by modifying the chromophore transition densities, and directly, through an explicit solvent term. This has often been introduced as a constant screening factor  $s$ , expressed in terms of an effective dielectric constant  $\epsilon_{\text{eff}}$  (for instance in the Förster or Onsager models [181]). More recently, an explicit environment contribution to the electronic coupling has been derived for polarisable QM/classical models, where it arises naturally [139, 179]. Such contribution accounts for the interaction between the transition density of one chromophore and the polarisation response of the environment to the transition density of the other chromophore. For instance, in the polarisable QM/MM, it reads:

$$V_{\text{AB}}^{\text{MMPol}} = - \sum_i^{N_{\text{MM}}} \int d\mathbf{r} \rho_{\text{A}}^{\text{tr}}(\mathbf{r}) \frac{\mathbf{r} - \mathbf{R}_i}{|\mathbf{r} - \mathbf{R}_i|^3} \cdot \boldsymbol{\mu}_i(\rho_{\text{B}}^{\text{tr}}). \quad (41)$$

Other terms should be introduced in the electronic coupling, such as exchange and overlap ones. These can be calculated in terms of the transition densities of the individual chromophores [182], or through diabatisation schemes recovering the full coupling from QM calculations on the whole AB system [183]. It should be noted however that, in the majority of the cases, the Coulomb and the environment interactions are the predominant terms to the coupling between singlet states. Within subsystem DFT, an extension has been introduced to account explicitly for solvent effects, which needs the inclusion of a large number of states [184].

The environment-induced screening of the electronic interaction between chromophores has been shown to be more complex than in the simplest Förster or Onsager pictures [185]. Even a continuum-based description of the environment showed that, within photosynthetic systems (including PSII, PE545, PC645 and LHCII), the effective screening varies markedly with the inter-chromophore distance, displaying an exponential decay [155]. It has also been stressed that, in protein environments, where specific and directional chromophore-residue interactions are likely to establish, it is important to capture this anisotropy by means of an explicit description of the environment structure [185]. Indeed it has been shown that in several pigment-protein complexes the chromophores can be differently affected by the dielectric heterogeneity of the protein environment [186]. Effective dielectric permittivities ranging between 1.4 and 2.6 have been calculated for the PE545 light-harvesting antenna complex [155] (compared with the value of 2 commonly used in continuum descriptions). In some studies on artificial complexes, it has been shown that the presence of a rigid bridge separating the chromophores can actually be able to enhance rather than screen the coupling between the chromophores, through both classical polarisation and orbital-mediated effects [187, 188].

## 4. Light absorption in biological systems

In this section we review recent advances on computational studies aimed to understand the role of the protein field in the light absorption and in the photophysics of the chromophores of paradigmatic systems: the light harvesting complexes in Sec.4.1, the retinal chromophore in Sec.4.2 and the wild-type Green Fluorescent Protein in Sec.4.3.

### 4.1. Photosynthesis

A wide variety of light harvesting (LH) pigment protein complexes (PPCs) exist, in very different photosynthetic organisms (bacteria, algae and higher plants). They absorb light through antenna complexes, which possess optimised spectral and spacial cross sections. Once absorbed, light is transferred and eventually ends up in reaction centres, where it is converted to chemical energy through charge separation processes [189]. Both the transfer and the conversion are extremely efficient, despite spanning various time and length scales. In order to model the energy absorption and transfer processes, as well as the charge separation, it is necessary to correctly account for the presence of the protein environment. Indeed, the latter does not only affect the electronic properties of the individual chromophores, but is also able to tune the electronic interactions among them. Moreover, while the energy transfer process was originally believed to proceed through an incoherent hopping mechanism, long-lasting coherence effects have been recently observed experimentally in several photosynthetic systems, even at room temperature [190–193]. This has shown that the dynamic interaction among the chromophores, as well as that between the chromophores and the environment, may be of paramount importance.

The pigments most commonly found in LH systems are chlorophylls (Chls) and bacteriochlorophylls (BChls), which can be of different kind. They are all characterised, in their native form, by the coordination of a  $\text{Mg}^{\text{II}}$  ion, a (bacterio)chlorine ring with four pyrrole-like units, and a polyisoprenoid alcohol chain called phytyl chain [3]. The absorption spectrum of (B)Chls presents a weak and an intense band, called Q and B (or Soret), respectively. The bands are characterised by two nearly degenerate transitions each, referred to as x and y. An important exception is given by the PCP complex [4], where the role of main harvester pigment is played by the peridinin carotenoid.

Several studies have been carried out to assess the performance of different theoretical methods in reproducing the excitation energies of (bacterio)chlorophylls and porphyrins [194], and carotenoids [195]. A recent study on the Q-band of BChl *a* in the Fenna-Matthews-Olson (FMO) complex of green sulphur bacteria, including the effect of the environment, found that the ground state is characterised by a near-multiconfigurational character [196]. In the study, the semiempirical ZINDO method [197] performed reasonably well to predict site energies

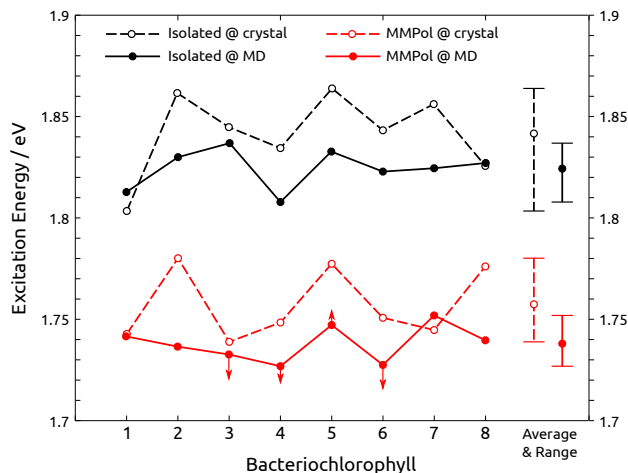
and transition dipole moments of the isolated BChl, and poorly when considering the environment. On the other hand, TDDFT provided robust and reliable results, with the hybrid functional PBE0 [198] ranking best, followed by B3LYP [199] and the long-range corrected CAM-B3LYP [200]. In general, the amount of exact exchange in the functional is important.

It has been shown that the geometric structure can affect quite markedly the resulting excitation energies of Chls and BChls [58]. For this reason, it is often recommended not to trust structures directly taken from X-ray diffraction, but to perform geometry optimisations first, although different optimisation methods generally provide different equilibrium structures, and the ability of the force fields in providing accurate geometries should be verified. Apart from methodological precautions, this observation also suggests that one of the roles played by the protein environment in light-harvesting systems originates directly from the structural deformations it causes, as clearly shown by the absorption spectrum of the nonequivalent peridinin molecules in PCP [195, 201, 202].

The optical properties of PPCs are determined by those of the individual pigments, and by their interplay. [203–205] In many cases, the spectra are determined by the tuning of the site energies by the protein environment. In other cases, the excitonic couplings, also affected by the protein, are dominant, as in LH2, which leads to a partial or complete delocalisation of the excited states. Intermediate cases, where both the chromophore-protein and chromophore-chromophore interactions are relevant, have been also studied theoretically. See for instance Ref. [205], where the environment tuning of the BChl energies within the LHC-II light harvesting complex of green plants has been analysed using TDDFT, suggesting the existence of an intrinsic energy transfer pathway from lumen to stroma. Different systems can help assess the reliability of the models that account for the surrounding environment in the calculation of energies and couplings. The dynamical interaction, usually taken into account through the spectral density, will be briefly presented.

**4.1.1. Fenna-Matthews-Olson complex** Various studies have tried to dissect and analyse the different effects of the environment on the BChl units of FMO [58, 204]. Indeed, the optical properties of the system arise directly from the tuning of the site energies by means of the surrounding protein residues. Being able to reproduce the correct energy ladder is a challenging task for theoretical chemistry, because of the many factors involved, and FMO is therefore an ideal test system. FMO was the first light-harvesting pigment-protein complex to be structurally identified [206], and forms a homotrimer, with eight BChl *a* molecules in each monomer [207]. The BChl units are placed in specific binding pockets, like in most protein complexes, creating individual microenvironments.

Computational studies, carried out on the crystal structure, have obtained spectra in good agreement with the experimental ones [208, 209]. These studies are based on electrostatic shift calculations [210], where the



**Figure 3:** Excitation energies computed in Ref. [186]. Data with and without environment (at MMPol level) are shown, both for the crystal and the MD geometries. The vertical arrows show the correction due to non-electrostatic short-range effects, which were recovered by including some of the closest residues at QM level.

protein environment and the chromophore are described in terms of fixed charges, and the resulting electrostatic interaction generates a chromophore-specific shift to the common excitation energy. Despite the relatively simple approach, the results showed that the major features of the FMO spectra arise from the BChl interaction with the surrounding aminoacid residues.

More recently, Gao *et al.* [211] employed a multiscale treatment based on non-polarisable QM/MM calculations carried out on snapshots extracted from a classical MD. They showed that the inclusion of the environment induces large variations in the site energies of the 8 BChls. In particular, they found that BChl 8 has the highest site energy, and may be responsible for the transfer of energy from the baseplate to the reaction centre. At the same time, BChl 3 acts the energy sink, having the lowest energy, as confirmed also by experiments [212]

In a more complex multiscale treatment by Jurinovich *et al.* [186], the nature of the environment-induced shift has been dissected into its main components. In this article, calculations were carried out both on the X-ray structure, and on a set of uncorrelated snapshots extracted from a purely classical MD. The 8 BChls were treated separately at TDDFT level, while the effect of the protein environment was introduced at the classical level, using different approaches. From the comparison of the results, the authors were able to draw conclusions on the nature of the environment effect. In particular, the protein environment was either completely neglected; or described as a structureless continuum; or described in terms of discrete classical polarisable atoms; or partially included at QM level. The study was focussed on the  $Q_y$  excitation. The authors were able to estimate that the purely geometric effects diversify the excitation

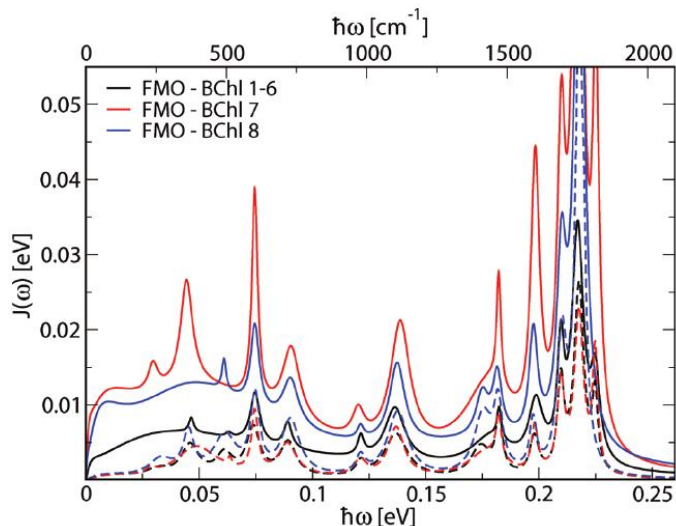


energies of the 8 BChls; the corresponding variations are rather large if computed on the X-ray structure, but reduce to a range of 0.03 eV when the average over the MD snapshots is performed, as shown in Figure 3. The direct environment effect due to the explicit account of electrostatic interaction with the surrounding protein causes a red-shift by 0.09 eV on average, which tends also to reduce the energy difference induced by the structural deformations. It was observed that BChl 3 is particularly stabilised by two  $\alpha$ -helices at short distance, as also reported by Renger [209], and shows the largest stabilisation overall by means of the environment. However, despite being identified as the trapping site by several studies [211,213], it does not display the lowest energy. To ascertain whether short-range non-electrostatic interactions, such as dispersion, between the BChls and the close residues could play a role, the authors included some of these residues at QM level, obtaining small corrections. Indeed, they found that BChls 3, 4 and 6 were stabilised by  $\sim 10$  meV. This correction, calculated only at the crystal structure, was applied to the MD-averaged results, as shown in the figure, where it is clear that BChl 3 is the most stabilised by it. The authors were also able to point out that, in most cases, the polarisable discrete approach was able to recover both specific electrostatic interactions at short-range, and the bulk effect of the environment.

An alternative study, carried out in 2013 by König *et al.* [203], is based on the extension of the FDET to coupled excited states, called subsystem TDDFT [168], as reported in Section 3.3. It has the advantage that the environment is also treated at QM level, upon the division of the system into fragments. The authors assessed the importance of a preliminary geometry optimisation of the BChl units within their (frozen) binding pockets, whereas the geometries from an unconstrained optimisation, or directly from the X-ray structure tended to under- and over-estimate the site energy differences. They showed that a careful study of the protonation of residues is of fundamental importance, as charged sites can largely affect the energies. They observed that the inclusion of the phytyl chain greatly increases the energy spread among BChl units, from 22 to 37 meV.

The same authors also estimated electronic couplings among chromophores, using an extension of FDET [168]. However, this approach cannot properly include the effective screening of the couplings by means of the protein environment. Although the issue of couplings will be better treated for the LH2 system, it is worth mentioning here that a correct account of the environment effects, as done by Jurinovich *et al.* [186], predicts an effective screening ranging between 0.9 and 2.3 for the different BChls in FMO. This poses questions on the validity of average screening factors commonly used to account for the protein environment.

The long-lasting coherences ( $\sim 100$  fs) observed in the FMO system could be explained by taking into account the thermal effects. This can be achieved in practice by computing the spectral density, which describes the



**Figure 4:** Spectral density calculated with and without the effect of external charges (solid and dashed lines, respectively). Figure from Ref. [214] with permission from the American Chemical Society.

frequency-dependent coupling between the system and the thermal environment. The spectral density can be obtained in an approximate way from the Fourier transform of the classical autocorrelation function of the site energy fluctuation:  $C_m^{cl}(t_i) \propto \sum_k \Delta E_m(t_i + t_k) \Delta E_m(t_k)$ , where  $m$  labels the BChl units. For a complete discussion on the spectral density see Ref. [215]. Olbrich *et al.* [214] computed the energies at ZINDO level on snapshots extracted from a classical MD (see Figure 4). They observed that the spectral density at low-energy is due to environment fluctuations; on the other hand, the peaks appearing in the whole energy range can be directly related to internal modes, and their strength is influenced by the fluctuating environment. In general, they found that the spectral density is similar for all BChls. A later work at TDDFT/MM level [216] again on snapshots from a MD found coherent beatings of the populations, consistently with the experimental observations.

**4.1.2. LH2** Quite differently to the FMO system, whose spectra are governed by the direct chromophore–protein interaction, the optical properties of LH2 are dominated by the chromophore–chromophore coupling. It is therefore an ideal system to test the models for the coupling, and to analyse how the environment affects it.

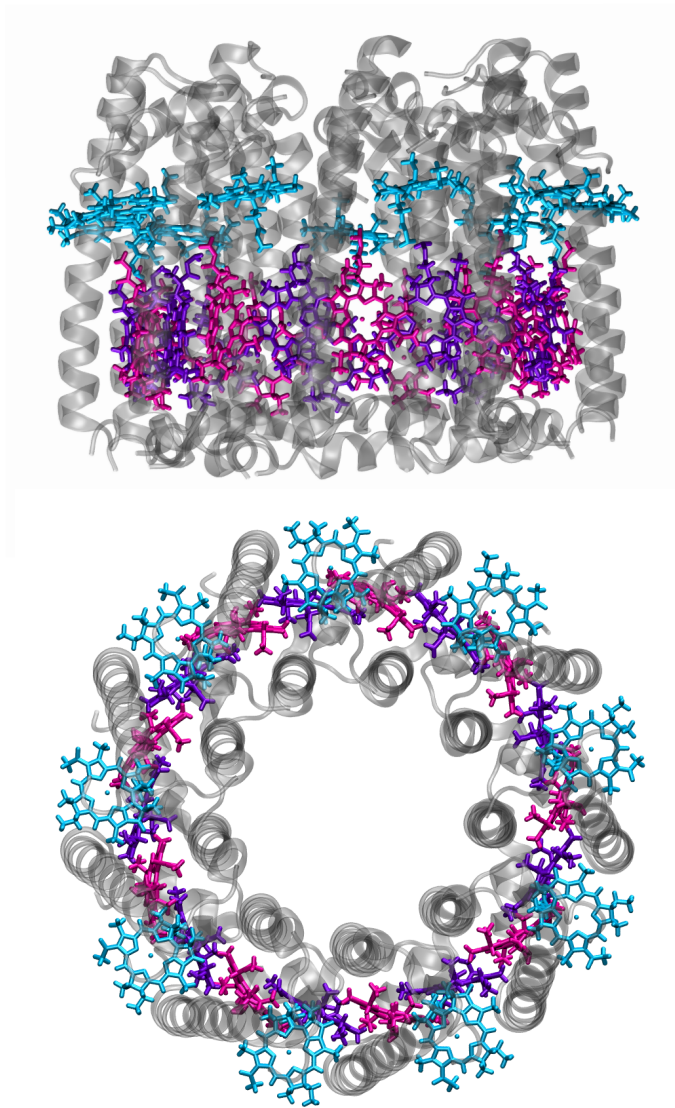
LH2 is a trans-membrane light-harvesting complex whose structure is well known [217]. It acts as an antenna complex in the photosystem of purple bacteria, where it is found together with LH1, which is instead associated with a reaction centre. It is characterised by a very symmetric structure ( $C_8$  or  $C_9$  symmetry) and comprises 24 or 27 bacteriochlorophylls  $a$  arranged in two rings (see Figure 5). Indeed, the absorption spectrum of LH2 shows a clear excitonic signature, as it is characterised by two bands at 800 and 850 nm (red-shifted with respect to the absorption

wavelength of isolated BChl *a*, which is 773 nm in diethyl ether). The BChls shown in blue are responsible for the peak at 800 nm, and will be referred to as B800 BChls. They are only weakly interacting, and the red-shift in their absorption is due to direct effects of the local protein environment. The closely packed purple/violet BChls (B850, absorbing at 850 nm) are instead characterised by a strong electronic coupling between their electronic states, which is responsible for the large red shift observed. The excitonic states resulting from the coupling can be delocalised over the whole ring [218]. The degree of delocalisation is reduced by the static disorder, which also contributes to redistribute the oscillator strength among the exciton states [219, 220]. The energy transfer from the B800 to the B850 states is extremely efficient, and the fluorescence emission of LH2 is due to the B850 ring only. The

One of the first accurate calculations of the electronic couplings in the system were performed using the transition density cube (TDC) method [178], where the transition densities of the chromophores are projected on a grid. The study showed that it is indeed necessary to account for the full shape of the transition density, when computing electronic couplings between chromophores at short range. At distances smaller than 15 Å (between BChls), or 20 Å (with the more elongated carotenoids), the point-dipole approximation fails [178]. This was confirmed by Linnanto *et al.* [221] who showed that couplings computed with a supermolecule approach (by describing the neighbouring B850 BChls as a dimer, estimating the coupling from the energy splitting of the resulting  $Q_y$  excitations) yielded a larger band splitting, more in line with experimental values. The authors also proved that the inclusion of the effect of close aminoacids, through a semiempirical CI approach, is of fundamental importance to recover the correct site energy of B800 BChls. For the B850 BChls, the inter-chromophore coupling is dominant over the chromophore-protein one.

The environment effect on the site energies of the B800 BChls has been analysed by He *et al.* [223] at TDDFT level, who studied both *Rhodospseudomonas acidophila* and *Rhodospseudomonas molischianum*, which display similar energies despite having quite different microenvironment structures. The authors found that, in the former, the observed red shift on the  $Q_x$  and  $Q_y$  excitations with respect to the isolated case is mostly induced by hydrogen bonding, and partly by axial ligation for the  $Q_x$ ; in *Rps. molischianum*, by contrast, no hydrogen bond can be established, but the red shift is due to axial ligation.

Studies combining MD simulations with electronic structure calculations (either semiempirical or ad HF level), and including the environment as a charge distribution, used a polaron analysis to investigate the dynamic effect of the environment. They found that the spectral broadening of the B800 peak is due to the fluctuations of the surrounding polar environment. The environment of B850 BChls, on the other hand, is non-polar, and the peak broadening is explained in terms of

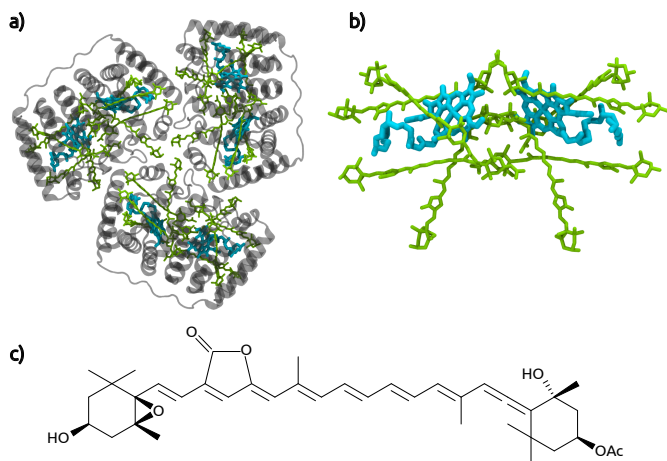


**Figure 5:** Crystal structure of LH2 of *Rhodospseudomonas acidophila* (PDB ID 1NKZ [222]).

the excitonic coupling among BChls [224, 225].

Subsystem DFT calculations were performed by Neugebauer [226], including the effect of the protein on the excitation energies using an effective embedding potential built from the density of the environment residues. They found that the major environment tuning is due to hydrogen bonding on the B800 BChls, causing a red shift of 0.10 and 0.06 eV on the  $Q_x$  and  $Q_y$  transitions, respectively; the B850 energies were instead rather unaffected by ligation or hydrogen bonding. The electronic couplings were also calculated, although lacking an explicit effect of the environment. The site energies and couplings however allowed to reproduce the experimental absorption spectrum.

The efficient energy transfer from the B800 to the B850 rings has been studied by van der Vegte *et al.* [227], using an atomistic description combined with a MD simulation. They showed that the energy transfer process is mediated by intermediate excitonic states, delocalised



**Figure 6:** a): structure of the trimeric PCP complex. The protein is shown as a grey ribbon. Chlorophylls are represented with blue and **PID** with green sticks. b): each monomer contains two Chl molecules and eight **PID** molecules. c): Lewis representation of **PID**.

over the whole complex.

**4.1.3. PCP** The PCP complex (Figure 6A) is a water-soluble protein trimer present in the marine dinoflagellate *Amphidinium carterae*. PCP contains the highest peridinin (**PID**) to chlorophyll-a ratio in nature, namely 4:1 for each domain [4, 6, 228, 229] (Figure 6B).

After a single-photon singlet excitation  $S_0 \rightarrow S_2$  ( $B_u^+$ -like, one-photon allowed), the **PID** chromophore (Figure 6C) couples to an adjacent chlorophyll-a according to a resonant energy-transfer mechanism. Two energy-transfer channels have been experimentally detected [230–234]: the first, accounting for about 25% of the energy, is the direct transfer from  $S_2$  to the  $Q_x$  state of the Chl; the second very efficient route starts from a fast internal conversion from  $S_2$  to  $S_1$  ( $A_g^-$ -like, one-photon forbidden), eventually interacting with an intramolecular charge transfer state and coupling with  $Q_y$  of the Chl.

The determination of the vertical absorption spectrum of **PID** is an essential ingredient in the study of the electronic coupling among the chromophores in PCP, as described in Section 3.4 (see Eq. 38).

Due to the size of **PID**, the theoretical investigation has been limited to semiempirical [202, 231, 235] or DFT [201] calculations. Only recently, high-level approaches have been applied to the gas phase model of **PID**, namely DFT/MRCI [236] and VMC/MBPT [89] methods.

A key structural parameter in the spectral tuning is the average bond length alternation (BLA), defined as the difference between the average length of single and double bonds. Differences of a few hundredths of an Å in average BLA may significantly alter vertical excitation energies of carotenoids, since the molecular orbitals involved in the low-lying region of the absorption spectrum are delocalised along the polyenic chain (Figure 6C), corresponding to a  $\pi \rightarrow \pi^*$  transition. A systematic blue shift in the gas

phase excitation energies is observed when increasing the average BLA, i.e., moving from overcorrelated structures (typically by GGA functionals) to the Hartree-Fock geometry, lacking of electronic correlation [89].

Another crucial aspect is the energy ordering of excited states in **PID** [89, 236]. Only with the inclusion of electronic transitions with double character, the pseudo-dark (one-photon transition forbidden)  $S_1$  state of  $A_g^-$ -like symmetry is found to be the lowest excited state, while the bright  $S_2$  state of  $B_u^+$ -like symmetry is above in energy [236]. The combined use of the VMC ground-state structure and MBPT for the vertical absorption represents a fully *ab initio* approach to obtain excitation energies (2.62 eV for the isolated molecule [89]) in meaningful agreement with the experimental findings, at least for the  $B_u^+$ -state, overcoming the difficulties in the choice of the proper functional in the TDDFT framework [89].

The intrinsic properties of the intramolecular charge-transfer state, supposed to be directly involved in the energy transfer processes occurring in PCP, have been investigated in several polar media [201, 237]: this state is characterised by an enhanced dipole moment, and by a mixing between  $S_1$  and  $S_2$  properties induced the polarity of the environment.

In Ref. [195] environment effects have been included by using the Mulliken charges of the entire PCP monomer complex, obtained on the crystallographic structure. MNDO semiempirical excited-state calculations for  $S_2$  have been performed in presence of this electrostatic background, highlighting the role of the surroundings in the spectral shifts (in the range 0.01-0.27 eV) of the various **PID**. Beyond the monomer calculations, excitonic effects in PCP may be important over the entire protein complex [195, 231, 238, 239]. They are estimated under different approximations (dipole-dipole, transition density) for the electronic coupling within the Förster energy-transfer model. Explicit quantum chemical calculations on coupled dimers of **PIDs** [195] reveal how the excitations are delocalised over the number of pigments, resulting in absorption energies deviating from the single-**PID** values.

#### 4.2. The retinal chromophore in rhodopsin

Rhodopsin is a light-detecting protein (Figure 7) belonging to the family of G-proteins located in the rod cells of the retina of the vertebrates, responsible for dim light vision of animals and humans [240–245]. Absorbing one photon the chromophore, the retinal protonated Schiff base (**RPSB**, Figure 8A, covalently bound to the Lys296), undergoes a very fast ( $\sim 200$  fs) and efficient (quantum yield of  $\sim 0.65$ ) isomerisation from the 11-*cis* to the all-*trans* form [240]. Femtosecond spectroscopy [246] and QM/MM (i. e., electrostatic embedding) molecular dynamics calculations [246–249] underline the importance of the protein environment and of the hydrogen-out-of-plane motion [250] in the isomerisation mechanism (Figure 8B), involving a  $S_1/S_0$  conical intersection [122, 251] along the torsion of the central double bond  $C_{11}-C_{12}$  from the 11-*cis* isomer to the all-*trans* form. The all-*trans* **RPSB**

is the precursor for the signal cascade in the mechanism of vision [240].

Bovine rhodopsin is the most extensively studied G-protein. Many theoretical works were published over the years to elucidate the role of the protein environment in tuning the absorption of **RPSB** [21, 22, 88, 90, 122, 123, 148, 246, 249–283]. This section will focus on bovine rhodopsin.

Experimental absorption on the bare chromophore [284–286] and on the full protein system [287–289] represents, of course, the reference for the theoretical predictions for the vertical maximum  $\lambda_{max}$ . The first excited state  $S_1$  is bright and characterised by a  $\pi \rightarrow \pi^*$  transition, while the second excited state  $S_2$  is one-photon forbidden with a partial double-transition character [285].

An accurate theoretical description of the optical absorption in this complex biological system depends on several aspects, as shown in Sections 2 and 3: the choice of the level of theory for geometry optimisation of the **RPSB**, the possible inclusion of thermal effects through MD and the choice of the method to compute excitation energies, the embedding scheme employed and the definition of the active site and environment subsystems, the protonation state of important residues inducing shifts in the absorption.

Color tuning is usually rationalised in terms of two main points: (i) the structural features of chromophore, and (ii) the modifications of the electronic properties (such as ground and excited state densities) due to the interaction with the polar and/or charged residues of the surrounding protein. In particular, many chromophores (as **PID** [89] and **RPSB**) are characterised by conjugated carbon chains with a large BLA which can be significantly influenced by geometrical distortion and field effects. Since the BLA descriptor was shown to correlate with the  $\lambda_{max}$  value in the optical absorption [89, 90], the accurate evaluation of the ground-state geometry becomes a crucial task. In the case of linear polyenes, DFT functionals (LDA, GGA, hybrids, long-range corrected hybrids, etc.) offer a wide range of values of the BLA, usually overestimating the correlation along the polyenic chain using GGA, whereas the use of coupled cluster or truncated-CI methods is limited to medium-size systems, due to the prohibitive computational cost.

**4.2.1. Gas phase properties of RPSB** A precise determination of the ground state structure of gas phase **RPSB** is a fundamental prerequisite to get a reliable starting point in the theoretical study of the isomerisation mechanism and for the understanding of the role of the protein environment on the spectral tuning. Several key geometrical parameters such as the already mentioned BLA of the conjugated carbon chain and the dihedral angle  $\phi(C_5-C_6-C_7-C_8)$  (Figure 8) involving the torsion of the  $\beta$ -ionone ring were shown to be crucial in the tuning of the optical absorption of the 11-*cis*-**RPSB**, corresponding to the so-called dark state of rhodopsin [88, 90, 240].

Many theoretical works [76, 290–311] were dedicated to the study of the penta-2,4-dieniminium cation (PSB3),

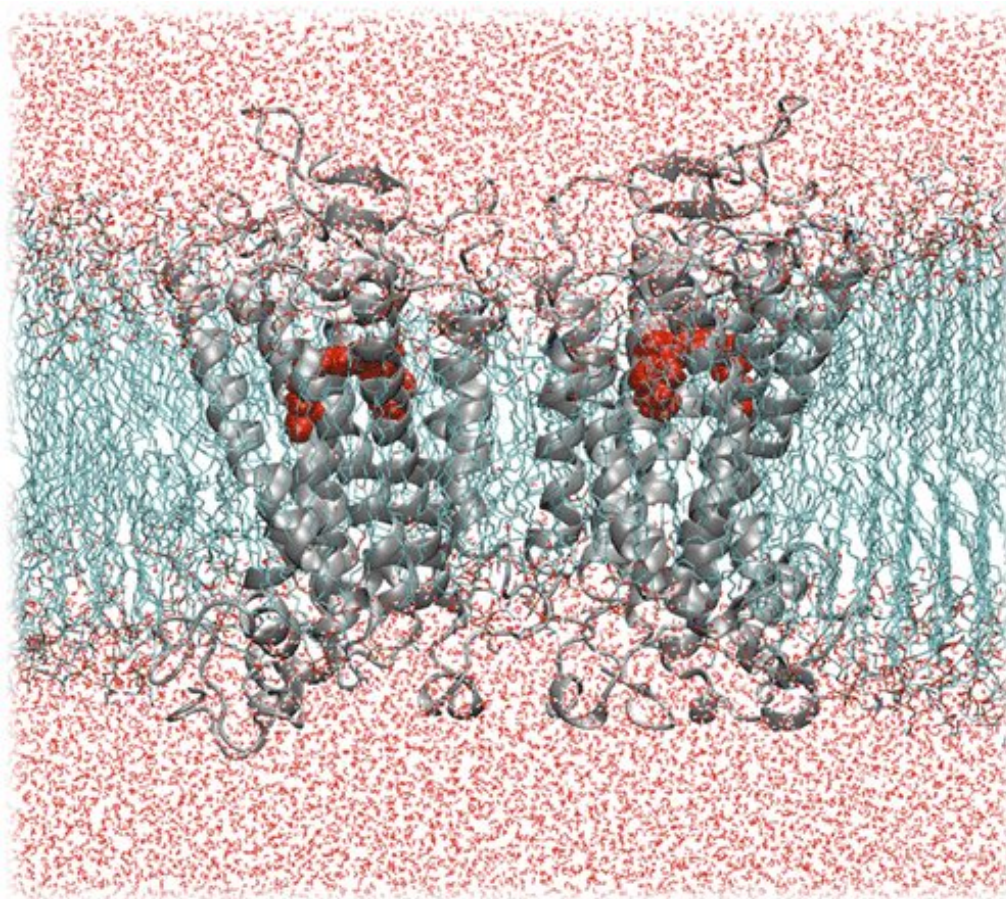
a small system used as reduced computational model of the full **RPSB**, since PSB3 and **RPSB** show similar ground and excited-state properties [305–309]. First, both PSB3 and **RPSB** have equilibrium structures for the *cis* and *trans* isomers where the positive charge is localised on the nitrogen. Second, the transition from the ground state ( $S_0$ ) to the first excited state ( $S_1$ ) exhibits a partial transfer of the positive charge from the nitrogen region towards the opposite end of the conjugated chain [290–304]. Third, twisting the conjugated chain along the central double bond leads to a conical intersection structure between the  $S_0$  and  $S_1$ . Such a structure plays a fundamental role in the photoisomerisation mechanism of **RPSB** since it mediates population transfer from  $S_1$  to  $S_0$ .

The study of the competition between the photoactivated and the thermal path for the isomerisation of the full **RPSB** represents a fundamental step for rationalising the vision mechanism [279]. For this reason, minimum energy path (MEP) calculations for PSB3 are not only essential to compare the intrinsic properties and performances of different theoretical approaches, but they are seen as a necessary step to understand the (photo)physics of rhodopsin systems.

As already observed for **PID**, increasing the average BLA (i. e. reducing the correlation along the carbon chain) produces a blue shift in the excitation [312]. Within DFT, hybrid functionals like PBE0, B3LYP, and M06-2X [95, 264, 312] or long-range-corrected versions (CAM-B3LYP) [313] provide a ground-state gas-phase BLA ranging from 0.033 Å (B3LYP) to 0.053 Å (CAM-B3LYP). CASSCF calculations, employing the full  $\pi$  space, give a larger average BLA of 0.101 Å [314], whereas the MP2 estimate [95, 315] shows a decreased BLA, with a value of 0.044 Å. QMC estimation of the average BLA obtained by VMC/JAGP calculations, is equal to 0.059(3) Å [88]. Taking the VMC/JAGP structure as the gas-phase reference, thanks to a balanced description of the electronic correlation, CAM-B3LYP and M062X bond-length patterns are close to the VMC one, and can be considered an accurate ground-state geometry for the 11-*cis* **RPSB**, as pointed out in Ref. [312].

Moreover, the torsion involving the  $\beta$ -ionone ring is strictly related to the extension of the conjugated chain: CASSCF shows a large absolute value of  $|\phi|$  (57°) [314], due to the large difference between single and double bonds because of the localisation of  $\pi$  electrons, on the other hand widely used functionals determine a smaller  $|\phi|$  angle (33.5° for B3LYP) and an effective larger conjugation length [88]. The VMC/JAGP structure exhibits a dihedral angle in between the two different classes of approaches ( $|\phi| = 42(1)^\circ$ ), as already reported for the average BLA.

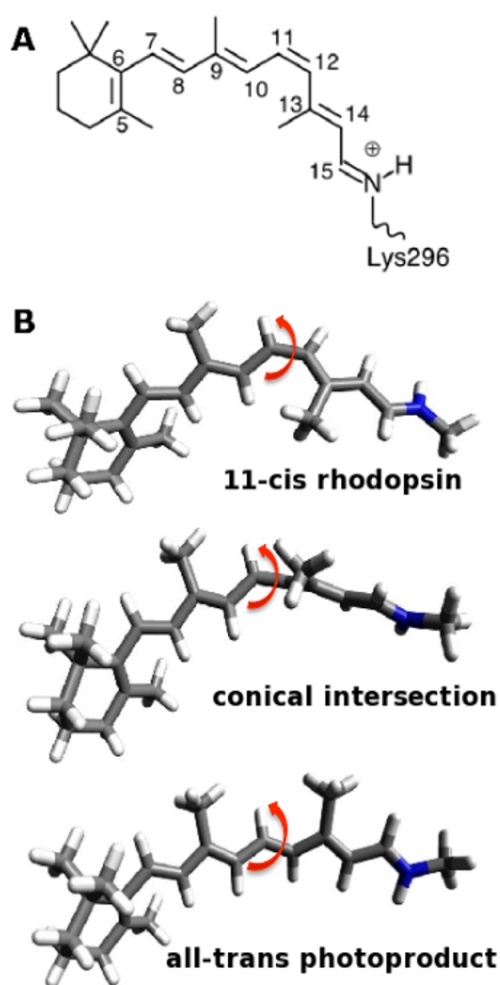
The important work by Valsson *et al.* [312] assigns a value of about 2.3 eV to the vertical excitation of 11-*cis* **RPSB** gas-phase model, thanks to a critical study of the correct balance between the effects of ground-state and excited-state methodologies in the computation of the  $S_1$  energy. Experimentally, a broad peak is observed for the  $S_1$  excitation [284–286] in the 2.03–2.34 eV range



**Figure 7:** Rhodopsin model at atomistic scale. The rhodopsin dimer, with the **RPSB** chromophores depicted in dark red, is embedded into the hydrophobic membrane surrounded by water (light red). Adapted from Ref. [123] with permission of the American Chemical Society.

(Table 2, column "Gas" corresponding to the isolated chromophore). This finding depends on the fact that the torsional degree of freedom related to the  $\beta$ -ionone ring is fully spanned at the experimental temperatures. In order to probe the absorption dependence on the value of the dihedral angle, experiments and calculations were carried out using one unlocked (i. e. no steric hindrance in the torsion of the  $\beta$ -ionone ring) and one locked (i. e. the torsion is rigidly forbidden) model of **RPSB** [286]. The values for the  $S_1$  energy calculated on the VMC/JAGP gas phase **RPSB** model structure at TDDFT and MBPT level [88, 90] (2.26 and 2.19 eV respectively, Tables 1 and 2) are found in the "high-energy" region of the experimental band [286] due to the partial break of conjugation on the  $\beta$ -ionone ring, in agreement with the findings reported by Rajput *et al.* [286] and with the prediction by Ref. [312]. Summarising, the correlated description of both ground-state structure and  $S_1$  excitation, eliminating cancellation of errors and reducing the effects of over- and under-correlation of the most used DFT and wave-function methods, as possible using the JAGP ansatz and the MBPT technique [90], produces a reliable estimation of the vertical absorption for the gas phase **RPSB** of around 2.20-2.30 eV, as collected in Table 2.

*4.2.2. Modelling the protein environment* A large part of theoretical works on bovine rhodopsin is based on the general accepted electrostatic embedding scheme, where only the chromophore is described at the quantum level, and the protein environment is modelled by using fixed and partial point charges [21, 88, 90, 122, 246, 249–251, 254, 258–263, 267–270, 272, 276, 277, 279–281, 316–331]. The charges polarise the electron density of the chromophore, while the environment does not experience the presence of the chromophore. Improvements to the simple embedding scheme with classical charges include the polarisation of the environment itself [148, 262, 270], as explained in Section 3, and the use of QM/QM models, at least for a small region of the protein (<500 atoms), as the FDET approach [14, 22]. Early attempts for a full QM description of rhodopsin were based on the use of the restricted Hartree-Fock [252] and DFT [264] on a system composed of the **RPSB** and few surrounding residues. Moreover, the force matching technique is used to improve the classical force field adopted in a QM/MM calculation [282]. Vertical absorption can be computed within a static approach, i. e., a geometry optimisation of the quantum subsystem affected by the environment followed by excited-state calculations, e. g. as in Refs [21, 88, 90, 249], or ex-



**Figure 8:** **A:** Lewis representation of **RPSB**, covalently bound to the Lys296. **B:** isomerisation mechanism of **RPSB**.

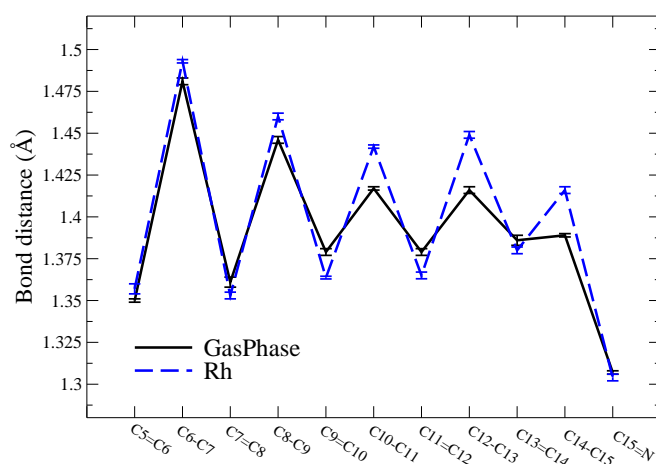
explicitly including thermal and dynamic effects by means of MD trajectories [23, 123, 248, 332–334].

Furthermore, the photoisomerisation of **RPSB** was investigated employing MEP and surface hopping calculations [320, 326, 335–340].

The X-ray structure of bovine rhodopsin from crystallographic data [341–345] represents the starting point for modelling the environment. The highest resolved structure (2.2 Å, named 1U19 [344]) places three water molecules in the active site, while the other structures [341–343] find one or two water molecules. These water molecules define a hydrogen-bond network on the extracellular side of the chromophore. All the X-ray structures substantially give the same conformations of the residues in the active site except for Ser186 [275].

Besides the characterisation of the optical properties of the dark state of rhodopsin, i. e., in presence of the 11-*cis* **RPSB**, attention was also focused to the study of photointermediates of the vision cycle, using CASSCF and CASPT2 [325, 346, 347], the density functional tight

binding approach [347] and MD coupled to semi-empirical methods [23] for estimating the vertical absorption.



**Figure 9:** Bond length pattern (in Å) optimised at VMC level for the gas phase **RPSB** and in presence of the protein environment. Figure adapted from Ref. [88].

The net effect of the protein environment on the **RPSB** bond length pattern is an increase of the average BLA value, from 0.059(3) Å to 0.088(3) Å [88], as seen in Figure 9, with a large difference between single and double bonds, due to a stronger localisation of the  $\pi$  electrons along the carbon chain. In Ref. [88] the ground state  $S_0$  geometry of **RPSB** was optimised at VMC level, both in the gas phase and embedded in the rhodopsin environment ("Rh" in Table 2) exploiting the electrostatic embedding QMC/MM reviewed above.

In spite of all the differences observed for the gas-phase model, DFT, wave-function and QMC methods agree in asserting that the overall electrostatic and geometrical protein effect is to induce an evident reduction of the  $\pi$ -conjugation [88, 90].

This finding is mainly due to the electrostatic coupling between the electronic density of the chromophore and the external field given by the surrounding residues, inducing a distortion from planarity, which instead characterizes the optimised ground-state geometry of the gas-phase **RPSB** [21, 88, 260, 264, 269]. The partial loss of  $\pi$ -conjugation is strictly connected to the value of the dihedral angle around the central bond C<sub>11</sub>-C<sub>12</sub>, equal to  $-15.4^\circ$  for the VMC/JAGP geometry optimisation in presence of the opsin and the membrane [88].

The VMC/MM calculations [88] are based on the 1HZX crystallographic structure [342]. The model comprises one chain of the full protein, water, and the cell membrane, simulated by a layer of *n*-octane molecules [88, 90]. The relaxed coordinates resulting from a DFT/MM annealing of the full system, using the BLYP functional for the chromophore, are the starting point for the VMC/MM calculations.

The position of polar and charged groups in the surrounding of **RPSB** plays a fundamental role in simulating absorption of the dark state of the bovine rhodopsin [21, 275]. The protonation state of the residues

	DFT gap	GW gap	$\Delta E$ (BSE)
Gas Phase	1.30	4.45	2.19 , 2.98
Rhodopsin	1.55	4.85	2.58, 3.47
Distorted	1.16	4.30	2.03 , 2.82

**Table 1:** DFT (at LDA level),  $G^0W^0$  gaps and vertical excitation energies (BSE equation) for the  $S_1$  and  $S_2$  states of the **RPSB** in gas-phase, distorted geometry and in rhodopsin environment. All energies in eV.

can largely influence the absorption. In particular, no general consensus (from both experimental and theoretical sides) is found for the protonation of Glu181 in rhodopsin, neutral or negatively charged, that can be involved in a hydrogen bond network [21, 266, 275, 278, 283, 348]. In Ref [88, 90] Glu181 is assumed to be negatively charged [278, 348, 349], while His211, Asp83, and Glu122 are taken neutral as suggested by FTIR experiments [348, 350].

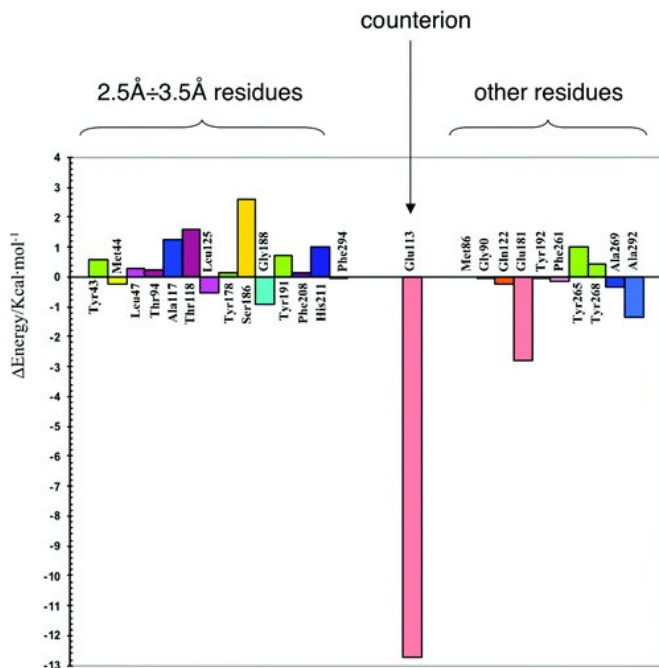
The rhodopsin environment induces a blue shift for the  $S_1$  low-lying state [21, 88, 90, 123, 249, 251, 260, 264, 268, 269, 287–289]. The experimental vertical absorption is indeed 2.48 eV [287–289]. A not fully comprehensive list of theoretical results with different levels of sophistication for the  $S_1$  bright excitation of 11-*cis* **RPSB** is reported in the synoptic Table 2, together with the experimental references.

The role of computational simulations became fundamental in dissecting the various contributions producing the blue shift in the absorption of rhodopsin [242–245]. Three important mutually interacting aspects will be briefly reviewed here: the role of the counterion Glu113, the quenching effect of the other residues, and, more in detail, the interplay between the **RPSB** geometry and the opsin environment in the color tuning.

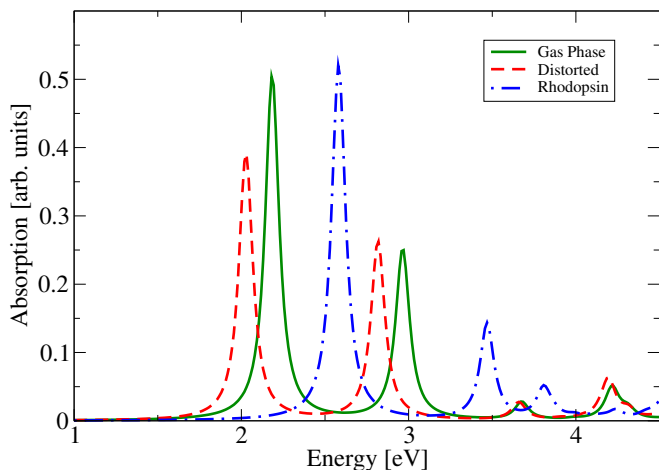
A strong interaction between the counterion Glu113 (a negatively charged glutamate residue) and the terminal positively charged nitrogen atom of the embedded **RPSB** is present in rhodopsin. The addition of the counterion forms a ionic pair inducing a blue shift in the  $S_1$  energy of **RPSB** with respect to the isolated chromophore, as predicted by DFT and CASSCF/CASPT2 calculations [21, 253, 255–257, 271, 351]. An accurate and quantitative estimation of the shift only from Glu113 has been obtained by a the reverse fingerprint analysis at the CASSCF/CASPT2 level of theory, for rhodopsin models including residues within a 2.5–3.5 Å radius around the (distorted) **RPSB**. As clearly shown in Figure 10, the blue shift of Glu113 is around 13 kcal/mol, corresponding to  $\sim 0.6$  eV [21].

The role of the rest of the protein environment in quenching the effect of the counterion is less evident. Tomasello *et al.* reports only a small red-shift of around 0.3 eV [21]. Theoretical investigations [263, 268, 269, 352, 353] reveal a negligible effect (red shift of 0.01–0.1 eV), while other calculations [122, 254, 260, 264, 354, 355] predict a larger red shift of 1.2 eV.

The use of highly correlated methods, as VMC/JAGP for the  $S_0$  geometry and MBPT for the  $S_1$  and  $S_2$  energies



**Figure 10:** Reverse Fingerprint analysis for a CASSCF/CASPT2 rhodopsin model comprising the **RPSB** and residues within a 2.5–3.5 Å radius around the chromophore. Adapted from Ref. [21] with permission of the American Chemical Society.



**Figure 11:** Absorption spectra calculated solving the BS equation for **RPSB** in gas phase (green solid line), the distorted geometry (red dashed-line) and in the protein environment (blue dashed-dot line). Calculations include resonant-antiresonant coupling. Figure from Ref. [90] with permission of Elsevier.

S <sub>0</sub> Geometry	S <sub>1</sub> excitation (eV)	Gas	Dist	Rh	Reference
VMC	TD-B3LYP	2.26	2.05	2.54	[88]
VMC	MBPT	2.19	2.03	2.58	[90]
DFT/B3LYP	TD-B3LYP	2.25	2.18	2.53	[280]
DFT/B3LYP	TD-B3LYP	-	-	2.58	[259]
DFT/PBE0	MCQDPT2	2.07	2.01	2.41	[264]
DFT/B3LYP	DMC	2.41(3)	-	-	[95]
DFT/B3LYP	TD-B3LYP	-	-	2.46	[268]
DFT/B3LYP	MRCISD+Q	2.06	-	2.48	[269]
DFT/B3LYP	SAC-CI	-	-	2.45	[263]
DFT/M06-2X	CASPT2/S-IPEA	2.30	-	-	[312]
DFT/M06-2X	NEVPT2/SC	2.33	-	-	[312]
DFT/BP86	FDET/TD-B3LYP	2.27	2.09	2.54	[22]
MP2	CASPT2/S-IPEA	2.24	-	-	[312]
MP2	NEVPT2/SC	2.27	-	-	[312]
MP2	CASPT2	2.05	-	-	[315]
CASSCF	CASPT2/0-IPEA	-	-	2.59	[122]
CASSCF	CASPT2/0-IPEA	2.28	-	-	[314]
CASSCF	CASPT2/0-IPEA	-	2.18	2.50	[21]
CASSCF	CASPT2/0-IPEA	-	-	2.59	[260]
SCC-DFTB	CASPT2	-	1.93	2.47	[352]
MD	ZINDO	-	-	2.40	[123]
	Exp	2.03-2.34	-	-	[285, 286]
	Exp	-	-	2.48	[287-289]

**Table 2:** Representative collection of theoretical and experimental S<sub>1</sub> vertical absorption energies of **RPSB**. Gs, Dist and Rh correspond to the gas-phase optimised chromophore, to the gas-phase molecule optimised (i. e. disorted) in the protein and to the rhodopsin model, respectively.

[90], for the chromophore properties within a QM/MM allows one to extract accurate values for the vertical absorption. S<sub>2</sub> is the second excited state with a small oscillator strength, at variance with the bright character, i. e. a large oscillator strength, of S<sub>1</sub>. The presence of the protein environment produces a widening of the electronic gap already in the LDA calculation with respect to the gas-phase model (1.55 against 1.30 eV, Table 1) and the effect is enhanced when looking at the quasi-particle gap calculated within the GW approximation (difference of 0.4 eV). The inclusion of the geometrical effects induced by the rhodopsin has the opposite effect of gap reduction (0.14 eV at LDA and 0.16 eV at GW level) [90]. Excitation energies for the distorted chromophore and for the Rh model are 2.03 and 2.58 eV, respectively [90], and they are reported in Table 1. The absorption spectra calculated at GW/BSE level, for the gas phase, distorted structure and in rhodopsin environment are reported in Figure 11: the highest peaks for the three structures correspond to the S<sub>1</sub> excitations. The blue shift in the absorption spectrum is found to be equal to 0.39 eV for S<sub>1</sub>. The absorption spectrum of the distorted geometry is red-shifted with respect to the gas-phase geometry (see Table 2) confirming the role played by the electrostatic coupling with the environment in reproducing the experimentally observed blue shift.

The BSE analysis of the S<sub>1</sub> excitation shows that the vertical transition is characterised by a predominant HOMO → LUMO character (89%) for all the three models [90]. The S<sub>S</sub> excitation has a partial double-transition character, that can not be properly described by the adiabatic approximation [90, 285].

A good agreement is found between the VMC/MBPT excitation energies [90] and the results from large-scale DFT calculations coupled to the FDET approach [22] (Table 2).

The vertical energies in Ref. [22] have been computed within the TDDFT framework taking into account 329 and 370 atoms in two different protein models. The overall FDET/TDDFT blue shift induced by the environment (0.27 eV, with respect to the gas-phase optimised model) well matches with the VMC/MBPT value of 0.39 eV: similarly, the effect of the electrostatic coupling with the surrounding residues corresponds to 0.45 eV for FDET/TDDFT and to 0.55 eV for VMC/MBPT. Within the same QM/MM approach, the VMC/TDDFT (using B3LYP) results [88] are in remarkable agreement with the FDET/TDDFT findings.

To conclude a rapid overview on other rhodopsin systems and on mutations follows here.

Bacterhodopsin acts as a proton pump and is not therefore involved in the vision, even though the protein binding pocket contains the **RPSB**, and significant examples of theoretical works on it are present in literature [327, 356–364].

Mutations in the bovine rhodopsin and engineering rhodopsin mimics [365] are essential to understand the modifications in the absorption spectrum of the spectrum induced by punctual modifications of the environment, i. e. substitutions of specific residues playing a role in spectral tuning [366, 367].

Several theoretical works have also been dedicated to the study of the structural and optical properties of rhodopsin in different biological systems: human and mouse visual pigments [365, 368], red, green and blue cone pigments [369], halorhodopsin [370], salamander [371] and squid rhodopsin [372, 373], proteorhodopsin [374], small white butterflies [375], sensory rhodopsin [376] and the study of the violet vision in fishes [377].



### 4.3. The Green Fluorescent Protein

As outlined in the Introduction, among photoactive proteins, GFP plays a fundamental role in bioimaging. GFP absorbs UV light and subsequently emits green light by fluorescence, and the GFP gene can be fused to other genes without altering their function. Thanks to these two unique properties GFP is routinely applied to the visualisation of protein dynamics and to monitor gene expression in living organisms by exposing them to UV light. In the last years it has been demonstrated that mutations of residues close to the chromophore of the GFP significantly alter the spectral properties of the protein, so that many variants of GFP have been engineered, and nowadays fluorescent markers spanning all the visible spectrum are available [11]. In this Section we focus on recent computational studies on the wild-type GFP (wt-GFP), which is the most studied from a theoretical point of view; many efforts have been devoted to understand the effect of the protein environment on its spectroscopic properties. wt-GFP is composed of 238 amino acids and has a  $\beta$ -barrel tertiary structure with the chromophore contained in its interior (left panel of Fig. 12). The absorption spectrum of wt-GFP presents two main peaks at 2.63 eV and 3.05 eV measured at 1.6 K [379] that, within the three-state model [380], have been ascribed to two stable neutral and anionic forms (A and B, respectively) of the bare chromophore (p-HBDI or p-HBDI<sup>-</sup>). The excitation energy is then released by fluorescence emission with a single main peak at 2.44 eV which is assigned to a different anionic form (I form) [381]. The equilibrium between A and B forms can be altered by external factors such as pH and mutations affecting the protein environment [379,380]. The structures of the A and B forms are sketched in the right panel of Fig. 12. They differ for the protonation state of p-HBDI and Glu-222 and mainly in the hydroxyl group of Thr-203 that is directed toward the chromophore in the B-form. In the recent years many computational *ab-initio* studies have addressed the question on how the protein environment tunes the excitation energies of the A and B forms of the wt-GFP and GFP mutants [96, 143, 146, 153, 378, 381–393]. Calculated excitation energies of some recent works are summarised in Tab. 3.

In particular, many authors approached the task using QM/MM methods [96, 385, 388, 390]. In Ref. [385] Bravaya *et al.* studied the singlet and triplet excitations of the anionic form B, and the effect of the environment to the vertical detachment energy. The model structure was obtained via the flexible effective fragment potential (EFP) QM/MM method [394]. Besides the QM/MM calculations the authors performed a series of fully quantum calculations including several amino acids (Glu222, Arg96, Ser205, His148 and two water molecules) observing only a small variation in the excitation energy due to the interaction of the chromophores with the nearby residues, while an increase of 100% was found in the vertical detachment energy. A very small influence of the protein on the excitation of the A form was previously reported by Hasegawa *et al.* in a SAC-CI/Amber calculation [384].

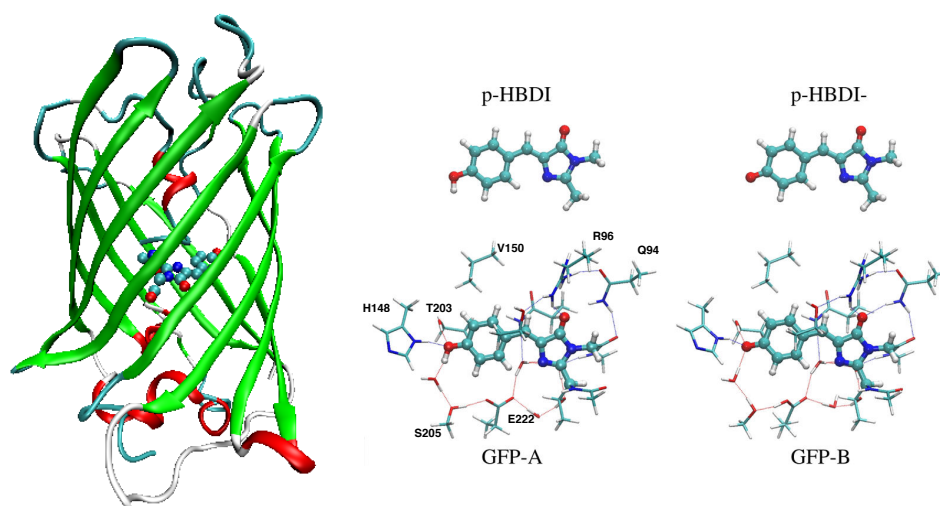
The ability of QM/MM in describing the effect of the protein field on the spectra of GFP was lately questioned by Filippi *et al.* [96]. A small effect of the protein field on the first excitation of the B form was also reported [96], performing an extensive study on the bathochromic shift induced by the environment on both A and B forms with advanced theoretical methods for excited states (TDDFT, CASPT2 and QMC). In this work, models for the A and B form were obtained starting from X-ray structures (the mutant S65T for the

anionic B form) and subsequently equilibrated at MM level and finally refined in a simulated annealing run within QM/MM at PBE/Amber03 level. The chromophore bond lengths in vacuo and in protein from the two forms are shown in Fig. 13, where it is evident that the environment does not dramatically change the structure of the two forms. The main differences were found in the BLA of the central bridge for the anionic form, on the other hand the two forms are more similar in protein than in vacuum. This large series of calculations showed that, while the adopted excited state theories were able to reproduce with reasonable agreement the experiments for the isolated chromophore, the inclusion of the protein environment at QM/MM level, although inducing a shift between the anionic and neutral forms, was not able to correctly describe the bathochromic shift experimentally observed, when using the values extrapolated from experiments [395] as the gas phase excitations. The calculated shifts were larger for the A form than the B form, and the excitation energies seen to be blue-shifted with respect to the experiments. Moreover the effect of the protein on the anionic chromophore as described by QM/MM results was in the opposite direction with respect to that experimentally observed (i. e., a redshift). These discrepancies were ascribed either to the effect of the polarisation of the environment, neglected in the present calculations, or to the need of investigating different protonation states in the chromophore and environment. In the same work the authors explored the possible presence of a solvated hydronium in the proximity of the chromophore, as suggested in Ref [396], but QM/MM calculations showed that such configuration was not stable.

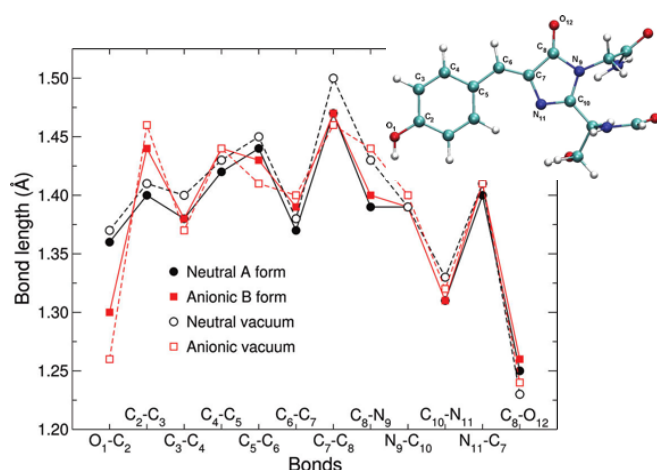
The fact that the protein induces a red-shift to the gas-phase excitation also for the B form was supported in a study based on PCM and ONIOM strategies by Petrone *et al.* [388]. The authors calculated excitation energies at CAM-B3LYP level for the anionic form in different solvents and in protein. Besides a systematic error, the experimental trend in solution was reproduced, indicating the extrapolated value of 2.84 eV [395] as the reference value for the gas-phase absorption. As a consequence, they found that the protein environment induces a red shift of about 0.2 eV with respect to the gas-phase excitation.

Recently, the A, B and I forms were investigated by Grikorenko *et al.* [390] using a QM/MM approach, including the side chains of Arg96, Glu222, Ser205, His148, and Thr203, and two water molecules in the QM portion. The authors calculated the total and excitation energies for the different structures and the anionic chromophore (B and I forms), which were found to be about 1 kcal/mol larger than those of the A form. In this study the authors analysed the role of the surrounding amino acids, pointing out that a major role in stabilising the anionic B form was played by the orientation of the Glu222 protonated side chain (*syn* in the I form and *anti* in the B form) rather than the Thr203, as originally suggested in the three-state model [380].

The effect of the polarisation of the environment on the optical absorption of GFP was recently approached with different strategies, either by cluster methods, i.e. adding more and more residues surrounding the chromophore in the quantum region, or by considering polarisable embedding techniques (QM/MMPol). Kaila *et al.* in Ref. [387] performed reduced virtual space CC2 (RCS-CC2) calculations on a model composed by the chromophore and 13 nearby residues (161 atoms), finding a quantitative agreement with experiments. Calculations on the isolated chromophore, either in the protein



**Figure 12:** Left panel:  $\beta$ -sheet barrel tertiary structure of GFP. The chromophore in the interior is shown in the ball-and-stick representation. Right panel: A (neutral) and B (anionic) forms of the wt-GFP chromophores responsible for the light absorption and adjacent residues as obtained by relaxing the structures at B3LYP/SVP level. The hydrogen bonding network surrounding the chromophore in the two form is also sketched. Reproduced from Ref [378] with permission from the American Chemical Society.

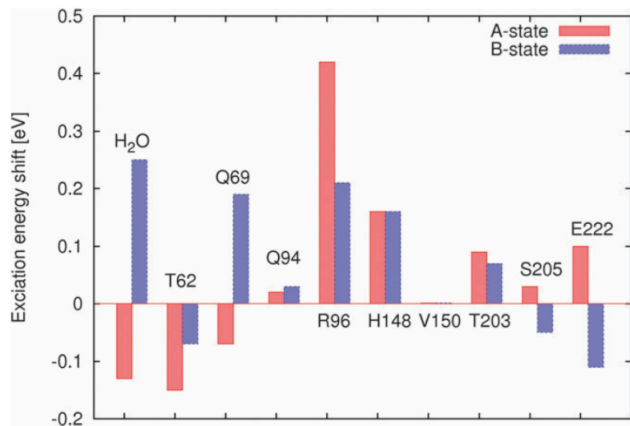


**Figure 13:** Bond lengths of the chromophore of the neutral A and anionic B forms as obtained in PBE/Amber CPMD simulations. The results for the chromophore models optimised in vacuo with BLYP/cc-pVTZ are also shown. Adapted from Ref. [96] with permission from the American Chemical Society

geometry, or in the relaxed one, permitted to divide the spectral tuning into electrostatic (70-80%) and steric effects (20-30%). A reasonable agreement (with a blueshift of  $\sim 0.2\text{eV}$ ) was also found when considering a QM/MM scheme where the surrounding environment was treated using the CHARMM force field with point charges extracted from DFT calculations, pointing out the importance of including the protein residues within the first “solvation shell”. A further analysis obtained by excluding the nearby amino acids one by one in QM/MM calculations (reverse fingerprint analysis, as performed for rhodopsin [21]) permitted to reveal the individual shift induced

by each of the residues. As shown in Figure 14), some of the components (e.g. water) induce spectral shifts of different sign on the two forms. The same authors in a following paper [378] compared the cluster approaches with QM/MM methods at different levels of theory and for different models. In this work the good performance of the RVS-CC2 method in describing the excitations of the chromophore and the effect of the environment, also at QM/MM level, is confirmed, provided that a large quantum region is used, highlighting the important role played by the surrounding residues and a proper description of the hydrogen bonding network surrounding the chromophore. In the same study the tendency of the CAM-B3LYP functional in overestimating excitation energies of GFP was also noticed. Cluster calculations on different models, differing for the orientation of the Thr203 residue, for the location of one of the water molecules and for the orientation of the molecular plane of His148, performed at CC2 level, differed by 0.1 eV only, and showed that the orientation of the Thr203 residue has a very small effect on the excitation energy. Moreover, QM/MM calculations including the nearest residues in the QM part indicated that most of the excitation energy shift is tuned by the first shell of amino acids.

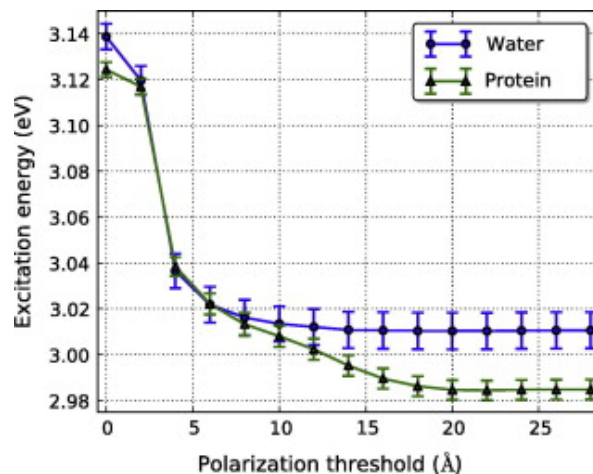
The effect of the mutual polarisation between the chromophore and the surrounding environment by using polarisable embedding TDDFT (PE-TDDFT) was first studied by Steindal *et al.* [386]. Starting from crystallographic structures, the authors found a good agreement with the experimental findings. In particular they tested with success the ability of PE-TDDFT in reproducing the effect of the interaction of each single amino acid surrounding the chromophore, by comparing the PE-TDDFT results with full-QM calculations. The impact of the inclusion of the polarisation effects in the environment was estimated to provide a red-shift of  $\sim 0.1\text{eV}$ . As in CC2/MM calculations of ref [387], a shift of  $\sim 0.3\text{eV}$  was found, induced by the crystallographic waters in the excitation of the anionic form. The same group in Ref. [146] reported a PE-TDDFT study on different fluorescent proteins. Contrary to the



**Figure 14:** Shift in excitation energy (in eV) upon removal of individual protein residues at the QM(CC2/def2-TZVP)/MM level of theory. Residues with positive/negative shifts have a red-/blue-shifting effect on the chromophore spectra. Reproduced from Ref. [387] with permission from the PCCP Owner Societies.

previous study, they considered a relaxed QM/MM structure and averaged structures from MD simulations without finding important differences between the two models. For the wt-GFP, the environment induces a red-shift of 0.18 and 0.13 eV respectively for the A and B forms, while a shift of 0.1 eV was due to polarisation effects. In line with the work of Kaila *et al.* [387], 70% of the environment effect was found to be due to electrostatic interaction for the neutral form, while no shift was predicted to the gas-phase excitation energy for the anionic form, when neglecting polarisation effects, as previously observed in Ref. [96]. In a following publication [153], Beereport *et al.* carried out PE-TDDFT calculations on snapshots extracted from MD simulations, where the extent of the polarisation effect on the determination of excitation energies was studied. The electrostatic interaction between the charge density of the chromophore and all the residues was taken into account, while the induced dipoles were included only up to a threshold distance. It was found that it is not sufficient to include only the polarisation of nearby residues, as it is necessary to account for the polarisation interaction of sites up to 20 Å away from the chromophore. Such a long-range polarisation effect was not found when considering the chromophore in a solvent, as shown in Fig 15. This discrepancy was explained by the presence of partially charged side groups in the protein matrix.

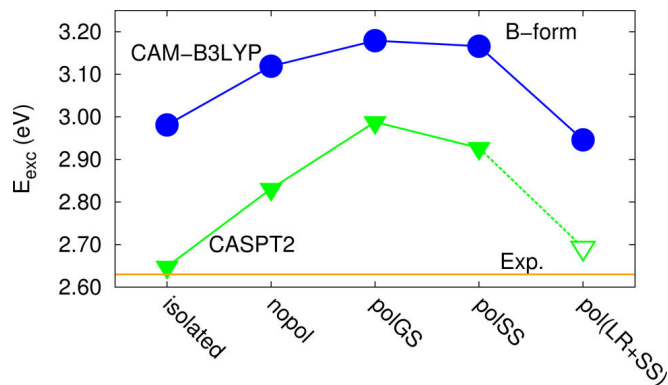
The long-range polarisation effects in wt-GFP was also confirmed very recently by Schwabe *et al.* [143]. The authors first reported that a RI-CC2 approach accounting for the environment with a polarisable embedding model reached a similar accuracy of the full-QM cluster approach. PE-RVSCC2 calculations were then carried out taking into account more and more residues, up to the entire protein, showing a slow convergence of the excited state energies with the size of the environment. This result also showed that quantum chemical calculations on a limited region around the chromophore could predict excitations in fortuitous agreement with the experiments. Such a slow convergence with the system size



**Figure 15:** Excitation energies at PE-CAM-B3LYP level of the chromophore of GFP in both the protein environment and solvated in water, using different polarisation cut-off thresholds. The results are averaged over 50 snapshots extracted from a MD simulation. Standard errors are shown. Electrostatic interactions were included for all atoms in the protein and for all water molecules within a sphere with radius 30Å around the chromophore. Adapted from Ref. [153] with permission from Elsevier.

in cluster methods was previously reported by Isborn *et al.* [397], showing that more than 700 atoms in the quantum region were needed to converge excitations in the photoactive yellow protein. Surprisingly, the need to include explicit crystallographic water in the QM region was not reported, as previously done in RVS-CC/MM and PE-DFT studies [146, 387], and a later work by Pikulska *et al.* [391], where different embedding models were tested to reproduce circular dichroism signals for wt-GFP and other fluorescent proteins.

Very recently Daday *et al.* [392] studied the chromophore-protein interaction using a large variety of excited-state techniques, ranging from TDDFT, wave-function methods and QMC, and with different methods to couple the chromophore with the environment (QM/MM, DFT embedding, polarisable embedding and cluster methods). The inclusion of dynamical thermal effects was also considered through QM/MM MD simulations. The stability of the hydrogen-bond network was studied for both A and B form and while the A form displays a very stable hydrogen-bond network, differently large deviations from the average structure were found for the B form. As reported in previous calculations [96], the QM/MM approach without polarisation effects was not able to retrieve the correct behaviour of excitation energies, as a description of the environment only in terms of fixed point charges causes a blue shift with respect to the experiments, and the account of thermal fluctuations did not improve the agreement (this had been also previously observed in Refs [146, 389]). The polarisation effects were then explored using three different approaches: by considering the environment response to the ground state density only, with no relaxation upon excitation (polGS); in a state-specific scheme where the polarisation relaxation depends on the density of the new state (polSS); in a linear-response scheme which includes the dynamical response of the environment to the transition density of



**Figure 16:** Excitation energies computed at CASPT2 and CAM-B3LYP level on the B-form chromophore with different descriptions of the protein environment: static point charges (nopol), point charges and induced dipoles according to the polGS, polSS and polLR schemes. The CASPT2 pol(LR+SS) value is an estimate. Adapted from Ref. [392] with permission from the American Chemical Society

the chromophore (polLR). It was found that the first two approaches did not much affect the energies, compared to the non-polarisable MM approach, while the linear-response approach caused a substantial red shift with respect to the polGS results for both the A ( $-0.17$  eV) and B ( $-0.22$  eV) forms (See Fig.16). It was observed that the polSS and polLR approaches describe the electrostatic and dispersion responses of the environment, respectively, and that in general both effects should be included, although in GFP the latter is dominant. Moreover, large cluster TDDFT calculations (up to 529 atoms) showed that the polLR method was able to capture the chromophore-protein interaction in the right way and that a cluster containing 300 atoms was large enough to describe the excitations for the two forms.

In the last years, several theoretical studies have been carried out to better describe and understand how the protein environment affects and tunes the photoexcited GFP chromophore. The approximations of the quantum chemical approaches, the uncertainties on the molecular structure around the chromophore region and on the protonation states and orientations of amino acid residues that are not fully resolved add to the necessary approximation performed to describe the large chromophore-protein environment. Nevertheless, despite these difficulties, multiscale approaches have once again proven to be a powerful instrument to obtain satisfactory results, and particularly suggesting the relevance of polarisation effects on the optical properties, including them either by classical methods or by considering large regions around the chromophore.

## 5. Conclusions and outlook

In this review we have reported the state-of-art of the *in silico* electronic absorption of biochromophores embedded in photoactive proteins.

Multiscale methods represent the most affordable computational scheme to properly describe the interaction between the chromophore and the complex environment, since a full quantum description of the system is impossible. Large effort has

been dedicated to the definition of accurate theoretical approaches able to include high-level methods for the description of structural and optical properties of the active site, *i.e.*, the protein region containing the chromophore, and efficient models for the perturbation produced by the protein system.

Quantum chemistry and DFT have been widely employed together with classical and *ab initio* molecular dynamics, and with continuum and discrete models for the electrostatic coupling. Computational schemes accounting for polarization effects of the environment that turn to play an important role on the excitations of chromophore have been also successfully coupled with DFT and wave function based methods. QMC and MBPT methods are in the last years emerging as suitable and promising tools to treat optical excitations of biological systems in complex environments. Moreover, embedding schemes, as the frozen density embedding theory, have been seen to well describe small portions of the protein, of course including the chromophore, thanks to the use of two quantum levels of theory, the higher for the chromophore and the lower for the rest of the environment.

Thermal and anharmonic effects in simulating the absorption can also be added, by applying approximate excited-state methods along a molecular dynamics trajectory.

A variety of photoactive biological systems has been theoretically investigated over the years. Here we have briefly reviewed the most recent results for the absorption in light-harvesting systems, the bovine rhodopsin and the green fluorescent protein, for which a large number of experimental data is available. For these systems, theory can be extremely helpful in understanding the fine features responsible for the colour tuning and can, in turn, play a predictive role.

The most noticeable limitation in this type of calculations is strictly connected to the scaling of quantum methods with respect to the number of electrons of the system. The application of high-level methods for computing absorption must be done carefully, and the introduced approximations should be always verified in order to get reliable results. In order to gain deeper insight into the description of the optical properties of biochromophores and of related processes occurring in the living organisms, improvements for the excited-state description would be desirable. Possible developments would be the overcoming of the adiabatic approximation in TDDFT and MBPT, which would allow to correctly describe excitations with double-transition nature, which are one-photon forbidden but are involved in energy-transfer mechanisms in light harvesting complexes, or defining a robust theoretical procedure for excited states in the QMC framework.

Concerning the coupling of the system with the environment, the extension of QM/MM embedding beyond the electrostatic coupling also for GW/BSE calculations and QMC would be surely a step further in modelling electronic absorption of biochromophores in proteins. The overcome of the traditional separation between the fields of electronic structure calculations and molecular dynamics should be encouraged. Several implementations allowing mixed QM/MM-MD simulations are becoming available, and the inclusion of the environment polarisation, through the development of better polarisable force-fields, consistent with the QM description of the active site, are an important goal for the future.

A completely novel approach, which is alternative to QM/MM models and just moving its first steps, and which has not been discussed in this review, is the use of machine learning techniques to further reduce the computational requirements

Geometry	Excited-state method	Gas-phase		Protein		Reference
		Neutral	Anionic	A-form	B-form	
QM/MM Electrostatic Embedding						
B3LYP/MM <sup>a</sup>	SAC-CI/MM <sup>c</sup>	3.23	-	3.21 <sup>g</sup>	-	[384]
CASSCF/MM <sup>b</sup>	CASPT2-0-IPEA/MM <sup>b</sup>	-	2.66	-	2.81	[381]
PBE/MM <sup>c</sup>	CAM-B3LYP/MM <sup>d</sup>	3.56	3.05	3.42	3.10	[96]
PBE/MM <sup>c</sup>	LC-BLYP/MM <sup>d</sup>	3.79	3.10	3.61	3.17	[96]
PBE/MM <sup>c</sup>	CASPT2/MM <sup>d</sup>	3.82	2.76	3.53	2.82	[96]
PBE/MM <sup>c</sup>	QMC/MM <sup>c</sup>	-	3.04(4)	-	3.1(1) <sup>e</sup>	[96]
PBE/MM <sup>c</sup>	CAM-B3LYP/MM <sup>d</sup>	-	-	3.38	3.11	[392]
PBE/MM <sup>c</sup>	LC-BLYP/MM <sup>d</sup>	-	-	3.53	3.18	[392]
PBE/MM <sup>c</sup>	CASPT2/MM <sup>d</sup>	-	-	3.24	2.82	[392]
PBE/MM <sup>c</sup>	QMC/MM <sup>c</sup>	-	-	3.55(2)	3.10(2) <sup>e</sup>	[392]
PBE/MM <sup>c</sup>	NEVPT/MM <sup>c</sup>	-	-	-	3.06	[392]
PBE0/MM <sup>c</sup>	SOS-CIS(D)/MM <sup>b</sup>	-	2.62	-	2.70	[385]
PBE0/MM <sup>c</sup>	SOS-CIS(D)/MM <sup>c</sup>	-	-	3.18	2.53	[390]
PBE0/MM <sup>c</sup>	XMCQDPT2/MM <sup>c</sup>	-	-	3.20	2.56	[390]
B3LYP/ONIOM <sup>f</sup>	CAM-B3LY/MM <sup>c</sup>	-	3.12	-	2.96	[388]
B3LYP <sup>h</sup>	RVS-CC2/MM <sup>b</sup>	3.69	2.91	3.38	2.75	[387]
B3LYP <sup>n</sup>	RVS-CC2/MM <sup>b</sup>	-	-	3.43	2.87	[378]
Cluster representation of the protein						
B3LYP <sup>h</sup>	RVS-CC2	3.69	2.91	3.13	2.68	[378, 387]
B3LYP/MM <sup>i</sup>	RVS-CC2	-	-	3.13	2.72	[143]
B3LYP <sup>h</sup>	CAM-B3LYP	-	-	3.37	3.00	[378]
B3LYP <sup>h</sup>	B3LYP	-	-	2.98	2.91	[378]
B3LYP <sup>n</sup>	RVS-CC2	-	-	3.25	2.77	[378]
B3LYP <sup>n</sup>	RVS-CC2/MM	-	-	3.43	2.87	[378]
PBE/MM <sup>l</sup>	CAM-B3LYP	-	-	3.28	3.03	[392]
PBE/MM <sup>l</sup>	LC-BLYP	-	-	3.48	3.14	[392]
PBE/MM <sup>m</sup>	CAM-B3LYP/MM	-	-	3.28	3.05	[392]
QM/MM Polarisable Embedding						
B3LYP/MM <sup>j</sup>	PE-RVS-CC2	-	-	3.33	2.72	[143]
PDB ID:1EMB	PE-CAM-B3LYP	3.75	3.17	2.93	2.65	[386]
B3LYP/MM <sup>j</sup>	PE-CAM-B3LYP	3.69	3.13	3.45	2.99	[146]
PBE/MM <sup>c</sup>	PE-CAM-B3LYP	-	-	3.19	2.96	[392]
PBE/MM <sup>c</sup>	PE-LC-BLYP	-	-	3.35	3.04	[392]
B3LYP/MM <sup>j</sup>	PE-CAM-B3LYP	3.61	3.06	3.42	3.05	[391]
FDET						
PDB ID: 1GFL <sup>k</sup>	PW91/DFT	-	-	3.33	3.09	[378]
PBE/MM <sup>c</sup>	CAM-B3LYP/DFT	-	-	3.43	3.23	[392]
PBE/MM <sup>c</sup>	CASPT2/DFT	-	-	-	3.02	[392]
Exp.						
		3.51*	2.84*	3.05	2.63	[379, 395]

**Table 3:** Representative collection of recent theoretical vertical absorption energies of A and B forms of wt-GFP. a) Amber94. b) CHARMM. c) Amber03. d) Amber99. e) The statistical error is indicated in parenthesis. f) AMBER+PCM for solvation. g) 3.27eV was found when Water 22, S205, E222, and S65 were included in the QM part. h) 161 atoms: CRO+T62,Q69,Q94,R96,H148,V150,T203,S205,E222,4H2O. i)OPLS and same residue of h. j) OPLS. k) 158 atoms, only the chromophore in the active region. l) 345 atoms:CRO+T62,Q69,Q94,R96,H148,V150,I167,T203,S205,E222,8H2O. m) 279 atoms: CRO+T62,Q94,R96,Y145,H148,T203,S205,E222,4H2O. n) Structure from [96] reoptimised at B3LYP level. \* Gas-phase values extrapolated from solution experiments. Photodestruction spectroscopy experiments [398–401] assign 2.99 eV for the neutral form and 2.6–2.7 eV for the anionic form. In protein absorption maxima at 1.6 K are reported, experiments at 295 K show maxima absorption peaks at 3.12 eV and 2.59 eV for the A and B form respectively.

while keeping an *ad lib.* accuracy. See for instance the work of Ref. [402], where a multilayer perceptron has been built to predict BChl excitation energies from TDDFT

calculations, including the effect of the classical non-polarisable environment. With the appropriate choice of variables and training sets, the calculation was much faster than the

QM/MM counterpart and proved to be equally accurate.

Finally, but not less important, development of algorithms for excited-state properties permitting to fully exploit the computer power nowadays available, and constantly growing in high-performing computing centres, would allow to easily explore dynamical and temperature effects by performing extensive MD calculations coupled to high-level quantum chemistry methods.

## Acknowledgements

We thank B. F. Milne, L. Guidoni, J. Toulouse and B. Mennucci for useful discussions. DV acknowledges partial support from the EU Centre of Excellence "MaX - Materials Design at the Exascale" (Grant No. 676598). EC thanks E. Luppi for support. SC acknowledges the European Research Council (ERC) for financial support in the framework of the starting grant EnLight-277755.

## References

- [1] Björn L O 2015 *Photobiology: The Science of Life and Light* (Springer)
- [2] Furukawa T, Hurley J B and Kawamura S 2014 *Vertebrate Photoreceptors: Functional Molecular Bases* (Springer)
- [3] Blankenship R E 2014 *Molecular Mechanisms of Photosynthesis* 2nd ed (Wiley-Blackwell)
- [4] Hofmann E, Wrench P M, Sharples F P, Hiller R G, Welte W and Diederichs K 1996 *Science* **272** 1788–1791
- [5] Di Valentin M, Tait C E, Salvadori E, Orian L, Polimeno A and Carbonera D 2014 *Biochim. Biophys. Acta* **1837** 85–97
- [6] Schulte T, Niedzwiedzki D M, Birge R R, Hiller R G, Polivtka T, Hofmann E and Frank H A 2009 *Proc. Natl. Acad. Sci. U. S. A.* **106** 20764–20769
- [7] Bonetti C, Alexandre M T A, van Stokkum I H M, Hiller R G, Groot M L, van Grondelle R and Kennis J T M 2010 *Phys. Chem. Chem. Phys.* **12** 9256
- [8] Applebury M L and Hargrave P A 1986 *Vision research* **26** 1881–1895
- [9] Sullivan K F and Kay S A 1999 *Green fluorescent proteins* vol 58 (Gulf Professional Publishing)
- [10] Newman R H, Fosbrink M D and Zhang J 2011 *Chem. Rev.* **111** 3614–3666
- [11] Day R N and Davidson M W 2009 *Chem. Soc. Rev.* **38** 2887–2921
- [12] Mukamel S 1999 *Principles of nonlinear optical spectroscopy* 6 (Oxford University Press on Demand)
- [13] Tomasi J, Mennucci B and Cammi R 2005 *Chem. Rev.* **105** 2999–3093
- [14] Wesolowski T A, Shedge S and Zhou X 2015 *Chem. Rev.* **115** 5891–5928
- [15] Warshel A and Levitt M 1976 *J. Mol. Biol.* **103** 227–249
- [16] Senn H M and Thiel W 2009 *Angew. Chem. Int. Ed.* **48** 1198–1229
- [17] Brunk E and Rothlisberger U 2015 *Chem. Rev.* **115** 6217–6263
- [18] Karplus M 2014 *Angew. Chem. Int. Ed.* **53** 9992–10005
- [19] Levitt M 2014 *Angew. Chem. Int. Ed.* **53** 10006–10018
- [20] Warshel A 2014 *Angew. Chem. Int. Ed.* **53** 10020–10031
- [21] Tomasello G, Olaso-González G, Altoé P, Serrano-Andrés M S L, Merchán M, Orlandi G, Bottoni A, and Garavelli M 2009 *J. Am. Chem. Soc.* **131** 5172–5186
- [22] Zhou X, Sundholm D, Wesolowski T A and Kaila V R I 2014 *J. Am. Chem. Soc.* **136** 2723–2726
- [23] Campomanes P, Neri M, Horta B A C, Röhrig U F, Vanni S, Tavernelli I, and Rothlisberger U 2014 *J. Am. Chem. Soc.* **136** 3842–3851
- [24] Cole D J and Hine N D M 2016 *J. Phys.: Condens. Matter* **28** 393001
- [25] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
- [26] Marques M A, Ullrich C A, Nogueira F, Rubio A, Burke K and Gross E K U 2006 *Time-dependent density functional theory* vol 706 (Springer)
- [27] Tapavicza E, Tavernelli I and Rothlisberger U 2007 *Phys. Rev. Lett.* **98** 023001
- [28] Alonso J L, Andrade X, Echenique P, Falceto F, Prada-Gracia D and Rubio A 2008 *Phys. Rev. Lett.* **101** 096403
- [29] Cammi R, Mennucci B and Tomasi J 2000 *J. Phys. Chem. A* **104** 5631–5637
- [30] Scalmani G, Frisch M J, Mennucci B, Tomasi J, Cammi R and Barone V 2006 *J. Chem. Phys.* **124** 094107
- [31] Marenich A V, Cramer C J, Truhlar D G, Guido C A, Mennucci B, Scalmani G and Frisch M J 2011 *Chem. Sci.* **2** 2143–2161
- [32] Röhrig U F, Frank I, Hutter J, Laio A, VandeVondele J and Rothlisberger U 2003 *ChemPhysChem* **4** 1177–1182
- [33] Daday C, König C, Valsson O, Neugebauer J and Filippi C 2013 *J. Chem. Theory Comput.* **9** 2355–2367
- [34] Autschbach J 2009 *ChemPhysChem* **10** 1757–1760
- [35] Maitra N T, Zhang F, Cave R J and Burke K 2004 *J. Chem. Phys.* **120** 5932–5937
- [36] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73** 33–83
- [37] Austin B M, Zubarev D Y and Lester W A J 2012 *Chem. Rev.*

- 112 263–288
- [38] Onida G, Reining L and Rubio A 2002 *Rev. Mod. Phys.* **74** 601
- [39] Helgaker T, Jørgensen P and Olsen J 2000 *Molecular Electronic-Structure Theory* (John Wiley & Sons, LTD)
- [40] Szabo A and Ostlund N S 1996 *Modern Quantum Chemistry* (Dover publications, Inc.)
- [41] Parr R G and Yang W 1989 *Density-Functional Theory of Atoms and Molecules* (Oxford University Press)
- [42] Ullrich C A 2012 *Time-Dependent Density-Functional Theory: Concepts and Applications* (Oxford University Press)
- [43] Andersson K, Malmqvist P A, Roos B, Sadlej A and Wolinski K 1990 *J. Phys. Chem.* **94** 5483–5488
- [44] Andersson K, Malmqvist P A and Roos B 1992 *J. Chem. Phys.* **96** 1218–1226
- [45] Angeli C, Cimiraglia R and Malrieu J P 2001 *Chem. Phys. Lett.* **350** 297–305
- [46] Angeli C, Cimiraglia R, Evangelisti S, Leininger T and Malrieu J P 2001 *J. Chem. Phys.* **114** 10252–10264
- [47] Angeli C, Cimiraglia R, and Malrieu J P 2002 *J. Chem. Phys.* **117** 9138–9153
- [48] Send R, Kaila V R I and Sundholm D 2011 *J. Chem. Phys.* **134** 214114–9
- [49] Send R, Kaila V R I and Sundholm D 2011 *J. Chem. Theory Comput.* **7** 2473–2484
- [50] Burke K 2012 *J. Chem. Phys.* **136** 150901
- [51] Kümmel S and Kronik L 2008 *Reviews of Modern Physics* **80** 3
- [52] Dreuw A and Head-Gordon M 2005 *Chem. Rev.* **105** 4009–4037
- [53] Pople J A, Seeger R and Krishnan R 1997 *Int. J. Quantum Chem. Suppl.* **Y-11** 149–163
- [54] Roos B O 1987 *Advances in Chemical Physics; Ab Initio Methods in Quantum Chemistry-II* (John Wiley & Sons, LTD)
- [55] Segarra-Martí J, Garavelli M and Aquilante F 2015 *J. Chem. Theory Comput.* **11** 3772–3784
- [56] Grimme S and Waletzke M 1999 *J. Chem. Phys.* **111** 5645–5655
- [57] Kleinschmidt M, Marian C M, Waletzke and Grimme S 209 *J. Chem. Phys.* **130** 044708–044711.
- [58] König C and Neugebauer A 2012 *ChemPhysChem* **13** 386–425
- [59] Ridley J and Zerner M 1973 *Theor. Chim. Acta* **32** 111–134
- [60] Dewar M J S and Thiel W 1977 *J. Am. Chem. Soc.* **99** 4899
- [61] Magdaong N M, Niedzwiedzki D M, Greco J A, Liu H, Yano K, Kajikawa T, Sakaguchi K, Katsumura S, Birge R R and Frank H A 2014 *Chem. Phys. Lett.* **593** 132–139
- [62] Dreuw A and Head-Gordon M 2004 *J. Am. Chem. Soc.* **126** 4007–4016
- [63] Castro A, Appel H, Oliveira M, Rozzi C A, Andrade X, Lorenzen F, Marques M A L, Gross E K U and Rubio A 2006 *Phys. Status Solidi (B)* **243** 2465–2488
- [64] Casida M E 1995 *Recent Advances in Density Functional Methods: (Part I)* **1** 155
- [65] Stein T, Kronik L and Baer R 2009 *J. Chem. Phys.* **131** 244119
- [66] Coccia E and Guidoni L 2012 *J. Comput. Chem.* **33** 2332–2339
- [67] Reynolds P J, Ceperley D M, Alder B J and Lester W A 1982 *J. Chem. Phys.* **77** 5593
- [68] Casula M, Moroni S, Sorella S and Filippi C 2010 *J. Chem. Phys.* **132** 154113
- [69] Zen A, Coccia E, Luo Y, Sorella S and Guidoni L 2014 *J. Chem. Theory Comput.* **10** 1048–1061
- [70] Casula M and Sorella S 2003 *J. Chem. Phys.* **119** 6500–6511
- [71] Casula M, Attaccalite C and Sorella S 2004 *J. Chem. Phys.* **121** 7110–7126
- [72] Sorella S *TurboRVB* quantum monte carlo package (accessed date may 2016) URL <http://people.sissa.it/~sorella/web/index.html>
- [73] Coleman A J 1963 *Rev. Mod. Phys.* **35** 668–687
- [74] Coleman A J 1965 *J. Math. Phys* **6** 1425–1431
- [75] Barborini M and Coccia E 2015 *J. Chem. Theory Comput.* **11** 5696–5704
- [76] Zen A, Coccia E, Gozem S, Olivucci M and Guidoni L 2015 *J. Chem. Theory. Comput.* **11** 992–1005
- [77] Bytautas L, Henderson T M, Jimenez-Hoyos C A, Ellis J K and Scuseria G E 2011 *J. Chem. Phys.* **135** 044119
- [78] Bytautas L, Scuseria G E and Ruedenberg K 2015 *J. Chem. Phys.* **143** 094105
- [79] Barborini M and Guidoni L 2015 *J. Chem. Theory Comput.* **11** 508–517
- [80] Barborini M and Guidoni L 2015 *J. Chem. Theory Comput.* **11** 4109–4118
- [81] Drummond N D, Towler M D and Needs R J 2004 *J. Phys. Rev. B* **70** 234119
- [82] Bressanini D and Reynolds P J 1998 *Advances in Chemical Physics, Monte Carlo Methods in Chemical Physics* **105** 5345–5350
- [83] Umrigar C J, Toulouse J, Filippi C, Sorella S and Hennig R G 2007 *Phys. Rev. Lett.* **98** 110201
- [84] Toulouse J and Umrigar C J 2007 *J. Chem. Phys.* **126** 084102
- [85] Toulouse J and Umrigar C J 2008 *J. Chem. Phys.* **128** 174101
- [86] Sorella S and Capriotti L 2010 *J. Chem. Phys.* **133** 234111
- [87] Barborini M, Sorella S and Guidoni L 2012 *J. Chem. Theory Comput.* **8** 1260–1269
- [88] Coccia E, Varsano D and Guidoni L 2013 *J. Chem. Theory Comput.* **9** 8–12
- [89] Coccia E, Varsano D and Guidoni L 2014 *J. Chem. Theory Comput.* **10** 501–506
- [90] Varsano D, Coccia E, Pulci O, Conte A M and Guidoni L 2014 *Comp. Theor. Chem.* **1040-1041** 338–346
- [91] Zen A, Luo Y, Mazzola G, Guidoni L and Sorella S 2015 *J. Chem. Phys.* **142** 144111
- [92] Mazzola G, Yunoki S and Sorella S 2014 *Nat. Comm.* **5** 3487
- [93] Mazzola G and Sorella S 2015 *Phys. Rev. Lett.* **114** 105701
- [94] Zimmerman P M, Toulouse J, Zhang Z, Musgrave C B and Umrigar C J 2009 *J. Chem. Phys.* **131** 124103
- [95] Valsson O and Filippi C 2010 *J. Chem. Theory Comput.* **6** 1275–1292
- [96] Filippi C, Buda F, Guidoni L and Sinicropi A 2011 *J. Chem. Theory Comput.* **8** 112–124
- [97] Fetter A L and Walecka J D 1971 *Quantum theory of many-body systems* (New York: McGraw-Hill)
- [98] Hedin L 1965 *Phys. Rev.* **139** A796
- [99] Strinati G 1982 *Phys. Rev. Lett.* **49** 1519
- [100] Varsano D, Marini A and Rubio A 2008 *Phys. Rev. Lett.* **101** 133002
- [101] Blase X and Attaccalite C 2011 *Appl. Phys. Lett.* **99** 171909
- [102] Baumeier B, Andrienko D and Rohlfing M 2012 *J. Chem. Theory Comput.* **8** 2790–2795
- [103] Duchemin I and Blase X 2013 *Phys. Rev. B* **87** 245412
- [104] Körbel S, Boulanger P, Duchemin I, Blase X, Marques M A and Botti S 2014 *J. Chem. Theory Comput.* **10** 3934–3943
- [105] Baumeier B, Rohlfing M and Andrienko D 2014 *J. Chem. Theory Comput.* **10** 3104–3110
- [106] Jacquemin D, Duchemin I and Blase X 2015 *Mol. Phys.* **114** 957–967
- [107] Jacquemin D, Duchemin I and Blase X 2015 *J. Chem. Theory Comput.* **11** 5340–5359
- [108] Bruneval F, Hamed S M and Neaton J B 2015 *J. Chem. Phys.* **142** 244101
- [109] Jacquemin D, Duchemin I and Blase X 2015 *J. Chem. Theory Comput.* **11** 3290–3304
- [110] Duchemin I, Deutsch T and Blase X 2012 *Phys. Rev. Lett.* **109** 167801
- [111] Faber C, Boulanger P, Duchemin I, Attaccalite C and Blase X 2013 *J. Chem. Phys.* **139** 194308
- [112] Ma Y, Rohlfing M and Molteni C 2009 *Phys. Rev. B* **80** 241405
- [113] Ma Y, Rohlfing M and Molteni C 2009 *J. Chem. Theory Comput.* **6** 257–265
- [114] Yin H, Ma Y, Mu J, Liu C and Rohlfing M 2014 *Phys. Rev. Lett.* **112** 228301
- [115] Marini A, Hogan C, Grüning M and Varsano D 2009 *Comput. Phys. Commun.* **180** 1392–1403
- [116] Hybertsen M S and Louie S G 1986 *Phys. Rev. B* **34** 5390
- [117] Bruneval F and Marques M A 2012 *J. Chem. Theory Comput.* **9** 324–329
- [118] Lischner J, Sharifzadeh S, Deslippe J, Neaton J B and Louie S G 2014 *Phys. Rev. B* **90** 115130

- [119] Marom N, Caruso F, Ren X, Hofmann O T, Körzdörfer T, Chelikowsky J R, Rubio A, Scheffler M and Rinke P 2012 *Phys. Rev. B* **86** 245127
- [120] Pinheiro Jr M, Caldas M J, Rinke P, Blum V and Scheffler M 2015 *Phys. Rev. B* **92** 195134
- [121] Körzdörfer T and Marom N 2012 *Phys. Rev. B* **86** 041110
- [122] Andruniow T, Ferré N and Olivucci M 2004 *Proc. Natl. Acad. Sci. U. S. A.* **101** 17908–17913
- [123] Valsson O, Campomanes P, Tavernelli I, Rothlisberger U and Flippin C 2013 *J. Chem. Theory Comput.* **9** 2441–2454
- [124] Malcıoğlu O B, Calzolari A, Gebauer R, Varsano D and Baroni S 2011 *J. Am. Chem. Soc.* **133** 15425–15433
- [125] Zhao G J and Han K L 2012 *Acc. Chem. Res.* **45** 404–413
- [126] Zhao G J and Han K L 2008 *Biophys. J.* **94** 38–46
- [127] Zhao G J, Liu J Y, Zhou L C and Han K L 2007 *J. Phys. Chem. B* **111** 8940–8945
- [128] Mennucci B 2012 *WIREs Comput. Mol. Sci.* **2** 386–404
- [129] Chipman D M 1997 *J. Chem. Phys.* **106** 10194–10206
- [130] Klamt A 2011 *WIREs Comput. Mol. Sci.* **1** 699–709
- [131] Gao J 1996 *Acc. Chem. Res.* **29** 298–305
- [132] Lin H and Truhlar D G 2006 *Theor. Chem. Acc.* **117** 185–199
- [133] Warshel A, Kato M and Pislakov A V 2007 *J. Chem. Theory Comput.* **3** 2034–2045
- [134] Ponder J W, Wu C, Ren P, Pande V S, Chodera J D, Schnieders M J, Haque I, Mobley D L, Lambrecht D S, Robert A DiStasio J, Head-Gordon M, Clark G N I, Johnson M E and Head-Gordon T 2010 *J. Phys. Chem. B* **114** 2549–2564
- [135] Wallrapp F H and Guallar V 2011 *WIREs Comput. Mol. Sci.* **1** 315–322
- [136] Lopes P E M, Roux B and MacKerell A D 2009 *Theor. Chem. Acc.* **124** 11–28
- [137] Luque F J, Dehez F, Chipot C and Orozco M 2011 *WRESs Comput. Mol. Sci.* **1** 844–854
- [138] Thompson M A 1996 *J. Phys. Chem.* **100** 14492–14507
- [139] Curutchet C, Muñoz-Losa A, Monti S, Kongsted J, Scholes G D and Mennucci B 2009 *J. Chem. Theory Comput.* **5** 1838–1848
- [140] Olsen J M, Aidas K and Kongsted J 2010 *J. Chem. Theory Comput.* **6** 3721–3734
- [141] Lipparini F and Barone V 2011 *J. Chem. Theory Comput.* **7** 3711–3724
- [142] Boulanger E and Thiel W 2012 *J. Chem. Theory Comput.* **8** 4527–4538
- [143] Schwabe T, Beerepoot M T P, Olsen J M H and Kongsted J 2015 *Phys. Chem. Chem. Phys.* **17** 2582–2588
- [144] Li Q, Mennucci B, Robb M A, Blancafort L and Curutchet C 2015 *J. Chem. Theory Comput.* **11** 1674–1682
- [145] Jensen L, van Duijnen P T and Snijders J G 2003 *J. Chem. Phys.* **119** 3800
- [146] Beerepoot M T P, Steindal A H, Kongsted J, Brandsdal B O, Frediani L, Ruud K and Olsen J M H 2013 *Phys. Chem. Chem. Phys.* **15** 4735
- [147] Guareschi R, Zulfikri H, Daday C, Floris F M, Amovilli C, Mennucci B and Filippi C 2016 *J. Chem. Theory Comput.* **12** 1674–1683
- [148] Caprasecca S, Jurinovich S, Viani L, Curutchet C and Mennucci B 2014 *J. Chem. Theory Comput.* **10** 1588–1598
- [149] Olsen J M H, Steinmann C, Ruud K and Kongsted J 2015 *J. Phys. Chem. A* **119** 5344–5355
- [150] van Duijnen P T and Swart M 1998 *J. Phys. Chem. A* **102** 2399–2407
- [151] Arcisauskaitė V, Kongsted J, Hansen T and Mikkelsen K V 2009 *Chem. Phys. Lett.* **470** 285–288
- [152] Sala J, Guàrdia E and Masia M 2010 *J. Chem. Phys.* **133** 234101
- [153] Beerepoot M T, Steindal A H, Ruud K, Olsen J M H and Kongsted J 2014 *Comput. Theor. Chem.* **1040–1041** 304–311
- [154] Sneskov K, Schwabe T, Christiansen O and Kongsted J 2011 *Phys. Chem. Chem. Phys.* **13** 18551
- [155] Mennucci B and Curutchet C 2011 *Phys. Chem. Chem. Phys.* **13** 11538
- [156] Guarneri Prandi I, Viani L, Andreussi O and Mennucci B 2016 *J. Comput. Chem.* **37** 981–991
- [157] Steindal A H, Ruud K, Frediani L, Aidas K and Kongsted J 2011 *J. Phys. Chem. B* **115** 3027–3037
- [158] Caprasecca S, Curutchet C and Mennucci B 2012 *J. Chem. Theory Comput.* **8** 4462–4473
- [159] Cammi R, Corni S, Mennucci B and Tomasi J 2005 *J. Chem. Phys.* **122** 104513
- [160] Caricato M, Mennucci B, Tomasi J, Ingrosso F, Cammi R, Corni S and Scalmani G 2006 *J. Chem. Phys.* **124** 124520
- [161] Corni S, Cammi R, Mennucci B and Tomasi J 2005 *J. Chem. Phys.* **123** 134512
- [162] Guido C A, Jacquemin D, Adamo C and Mennucci B 2015 *J. Chem. Theory Comput.* **11** 5782–5790
- [163] Laio A, VandeVondele J and Röthlisberger U 2002 *J. Chem. Phys.* **116** 6941
- [164] CPMD copyright IBM Corp 1990–2008, Copyright MPI für Festkörperforschung Stuttgart 1997–2001 URL <http://www.cpmd.org/>
- [165] Duchemin I, Jacquemin D and Blase X 2016 *J. Chem. Phys.* **144** 164106
- [166] Chung L W, Sameera W M C, Ramozzi R, Page A J, Hatanaka M, Petrova G P, Harris T V, Li X, Ke Z, Liu F, Li H B, Ding L and Morokuma K 2015 *Chem. Rev.* **115** 5678–5796
- [167] Wesolowski T A 2004 *J. Am. Chem. Soc.* **126** 11444–11445
- [168] Neugebauer J 2007 *J. Chem. Phys.* **126** 134116
- [169] Neugebauer J 2010 *Phys. Rep.* **489** 1–87
- [170] Neugebauer J 2009 *Chem Phys Chem* **10** 3148–3173
- [171] Jacob C R and Neugebauer J 2014 *WIREs Comput Mol Sci* **4** 325–362
- [172] Elliot P, Cohen M H, Burke K and Wasserman A 2009 *J. Chem. Theory Comput.* **5** 827–833
- [173] Huang C, Pavone M and Carter E A 2011 *J. Chem. Phys.* **134** 154110
- [174] Chenu A and Scholes G D 2015 *Annu. Rev. Phys. Chem.* **66** 69–96
- [175] Chang J C 1999 *J. Chem. Phys.* **67** 3901
- [176] Kistler K A, Spano F C and Matsika S 2013 *J. Phys. Chem. B* **117** 2032–2044
- [177] Fujimoto K J 2014 *J. Chem. Phys.* **141** 214105
- [178] Krueger B P, Scholes G D and Fleming G R 1998 *J. Phys. Chem. B* **102** 5378–5386
- [179] Iozzi M F, Mennucci B, Tomasi J and Cammi R 2004 *J. Chem. Phys.* **120** 7029–7040
- [180] Muñoz-Losa A, Curutchet C, Galván I F and Mennucci B 2008 *J. Chem. Phys.* **129** 034104
- [181] Knox R S and Amerongen H V 2002 *J. Phys. Chem. B* **106** 5289–5293
- [182] Hsu C P, Fleming G R, Head-Gordon M and Head-Gordon T 2001 *J. Chem. Phys.* **114** 3065–3072
- [183] Hsu C P 2009 *Acc. Chem. Res.* **42** 509–18
- [184] Neugebauer J, Curutchet C, Muñoz-Losa A and Mennucci B 2010 *J. Chem. Theory Comput.* **6** 1843–1851
- [185] Curutchet C, Kongsted J, Muñoz-Losa A, Hossein-Nejad H, Scholes G D and Mennucci B 2011 *J. Am. Chem. Soc.* **133** 3078–84
- [186] Jurinovich S, Curutchet C and Mennucci B 2014 *ChemPhysChem* **15** 3194–3204
- [187] Curutchet C, Mennucci B, Scholes G D and Beljonne D 2008 *J. Phys. Chem. B* **112** 3759–3766
- [188] Caprasecca S, Curutchet C and Mennucci B 2013 *J. Phys. Chem. C* **117** 12423–12431
- [189] van Amerongen H, van Grondelle R and Valkunas L 2000 *Photosynthetic Excitons* (World Scientific Publishing Co. Pte. Ltd.) ISBN 9789812813664
- [190] Panitchayangkoon G, Hayes D, Fransted K A, Caram J R, Harel E, Wen J, Blankenship R E and Engel G S 2010 *Proc. Natl. Acad. Sci.* **107** 12766–12770
- [191] Cheng Y C and Fleming G R 2009 *Annu. Rev. Phys. Chem.* **60** 241–262
- [192] Engel G S, Calhoun T R, Read E L, Ahn T K, Mančal T, Cheng Y C, Blankenship R E and Fleming G R 2007 *Nature* **446** 782–786
- [193] Lee H, Cheng Y C and Fleming G R 2007 *Science* **316** 1462–1465



- [194] Linnanto J and Korppi-Tommola J 2006 *Phys. Chem. Chem. Phys.* **8** 663–687
- [195] Bricker W P and Lo C S 2015 *J. Phys. Chem. B* **119** 5755–5764
- [196] List N H, Curutchet C, Knecht S, Mennucci B and Kongsted J 2013 *J. Chem. Theory Comput.* **9** 4928–4938
- [197] Zerner M C 1991 *Rev. Comput. Chem.* **2** 313–365
- [198] Adamo C and Barone V 1999 *J. Chem. Phys.* **110** 6158
- [199] Becke A D 1993 *J. Chem. Phys.* **98** 5648
- [200] Yanai T, Tew D P and Handy N C 2004 *Chem. Phys. Lett.* **393** 51–57
- [201] Vaswani H, Hsu C, Head-Gordon M and Fleming G 2003 *J. Phys. Chem. B* **107** 7940–7946
- [202] Kusumoto T, Horibe T, Kajikawa T, Hasegawa S, Iwashita T, Cogdell R J, Birge R R, Frank H A, Katsumura S and Hashimoto H 2010 *Chem. Phys.* **373** 71–79
- [203] König C and Neugebauer J 2013 *J. Chem. Theory Comput.* **9** 1808–1820
- [204] Curutchet C and Mennucci B 2016 *Chem. Rev.* DOI:10.1021/acs.chemrev.5b00700
- [205] Jornet-Somoza J, Alberdi-Rodríguez J, Milne B F, Andrade X, Marques M A, Nogueira F, Oliveira M J, Stewart J J and Rubio A 2015 *Physical Chemistry Chemical Physics* **17** 26599–26606
- [206] Matthews B W, Fenna R E, Bolognesi M C, Schmid M F and Olson J M 1979 *J. Mol. Biol.* **131** 259–285
- [207] Tronrud D E, Wen J, Gay L and Blankenship R E 2009 *Photosynth. Res.* **100** 79–87
- [208] Adolphs J and Renger T 2006 *Biophys. J.* **91** 2778–97
- [209] Müh F, Madjet M E A, Adolphs J, Abdurahman A, Rabenstein B, Ishikita H, Knapp E W and Renger T 2007 *Proc. Natl. Acad. Sci. U.S.A.* **104** 16862–16867
- [210] Renger T and Müh F 2013 *Phys. Chem. Chem. Phys.* **15** 3348–71
- [211] Gao J, Shi W j, Ye J, Wang X, Hirao H and Zhao Y 2013 *J. Phys. Chem. B* **117** 3488–95
- [212] Wen J, Zhang H, Gross M L and Blankenship R E 2009 *Proc. Natl. Acad. Sci.* **106** 6134–6139
- [213] Olbrich C, Jansen T L C, Liebers J, Aghtar M, Strümpfer J, Schulten K, Knoester J and Kleinekathöfer U 2011 *J. Phys. Chem. B* **115** 8609–8621
- [214] Olbrich C, Strümpfer J, Schulten K and Kleinekathöfer U 2011 *J. Phys. Chem. Lett.* **2011** 1771–1776
- [215] Lee M K, Huo P and Coker D F 2016 *Annu. Rev. Phys. Chem.* **67** 639–668
- [216] Shim S, Rebentrost P, Valleau S and Aspuru-Guzik A 2012 *Biophys. J.* **102** 649–660
- [217] Cogdell R J, Isaacs N W, Freer A A, Howard T D, Gardiner A T, Prince S M and Papiz M Z 2003 *{FEBS} Letters* **555** 35 – 39
- [218] Smyth C, Fassiolli F and Scholes G D 2012 *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **370** 3728–3749
- [219] Alden R G, Johnson E, Nagarajan V, Parson W W, Law C J and Cogdell R G 1997 *J. Phys. Chem. B* **101** 4667–4680
- [220] Hu X, Ritz T, Damjanović A and Schulten K 1997 *J. Phys. Chem. B* **101** 3854–3871
- [221] Linnanto J, Korppi-Tommola J E I and Helenius V M 1999 *J. Phys. Chem. B* **103** 8739–8750
- [222] Papiz M Z, Prince S M, Howard T, Cogdell R J and Isaacs N W 2003 *J. Mol. Biol.* **326** 1523 – 1538
- [223] He Z, Sundström V and Pullerits T 2002 *J. Phys. Chem. B* **106** 11606–11612
- [224] Damjanović A, Kosztin I, Kleinekathöfer U and Schulten K 2002 *Phys. Rev. E* **65** 031919
- [225] Janosi L, Kosztin I and Damjanović A 2006 *J. Chem. Phys.* **125**
- [226] Neugebauer J 2008 *J. Phys. Chem. B* **112** 2207–2217
- [227] van der Vegte C P, Prajapati J D, Kleinekathöfer U, Knoester J and Jansen T L C 2015 *J. Phys. Chem. B* **119** 1302–13
- [228] Wörmke S, Mackowski S, Brotsudarmo T and Bräuchle C 2007 *Biochim. Biophys. Acta* **1767** 956–964
- [229] Di Valentin M, Ceola S, Salvadori E and Carbonera D 2008 *Biochim. Biophys. Acta* **1777** 186–195
- [230] Bautista J A, Hiller R G, Frank P, Sharples F P, Gosztola D, Wasielewski M and Frank H A 1999 *J. Phys. Chem. A* **103** 2267–2273
- [231] Damjanović A, Ritz T and Schulten K 2000 *Biophys. J.* **79** 1695–1705
- [232] Krueger B P, Lampoura S S, van Stokkum I H M, Papagiannakis E, Salverda J M, Gradinaru C C, Rutkauskas D, Hiller R G and van Grondelle R 2001 *Biophys. J.* **80** 2843–2855
- [233] Zigmantas D, Hiller R G, Sundström V and Polivka T 2002 *Proc. Natl. Acad. Sci. U. S. A.* **99** 16760–16765
- [234] Polivka T, Hiller R G and Frank H A 2007 *Arch. Biochem. Biophys.* **458** 111–120
- [235] Shima S, Ilagan R, Gillespie N, Sommer B, Hiller R, Sharples F, Frank H and Birge R 2003 *J. Phys. Chem. A* **107** 8052–8066
- [236] Knecht S R, Marian C M, Kongsted J and Mennucci B 2013 *J. Phys. Chem. B* **117** 13808–13815
- [237] Wagner N L, Greco J A, Enriquez M M, Frank H A and Birge R R 2013 *Biophys. J.* **104** 1314–1325
- [238] Kleima F J, Gobets E H B, van Stokkum I H M, van Grondelle R, Diederichs K and van Amerongen H 2000 *Biophys. J.* **78** 344–353
- [239] Polivka T, Pascher T, Sundström V and Hiller R G 2005 *Photosynth. Res.* **86** 217–227
- [240] Palczewski K 2006 *Annu. Rev. Biochem.* **75** 743–767
- [241] Ernst O P, Lodowski D T, Elstner M, Hegemann P, Brown L S and Kandor H 2014 *Chem. Rev.* **114** 126–163
- [242] Wanko M, Hoffmann M, Frauenheim T and Elstner M 2006 *J. Comput. Aided Mol. Des.* **20** 511–518
- [243] Nielsen M B 2009 *Chem. Soc. Rev.* **38** 913–924
- [244] Rivalta I, Nenov A and Garavelli M 2014 *Phys. Chem. Chem. Phys.* **16** 16865–16879
- [245] Gozem S, Melaccio F, Luk H L, Rinaldi S and Olivucci M 2014 *Chem. Soc. Rev.* **43** 4019–4036
- [246] Polli D, Altoé P, Weingart O, Spillane K M, Manzoni C, Brida D, Tomasello G, Orlandi G, Kukura P, Mathies R A, Garavelli M and Cerullo G 2010 *Nature* **467** 440–443
- [247] Röhrig U F, Guidoni L, Laio A, Frank I and Rothlisberger U 2004 *J. Am. Chem. Soc.* **126** 15328–15329
- [248] Röhrig U F, Guidoni L and Rothlisberger U 2005 *ChemPhysChem* **6** 1836–1847
- [249] Frutos L M, Andruniow T, Santoro F, Ferré N and Olivucci M 2007 *Proc. Natl. Acad. Sci. U. S. A.* **104** 7764–7769
- [250] Schapiro I, Ryazantsev M N, Frutos L M, Ferré N and Olivucci M 2011 *J. Am. Chem. Soc.* **133** 3354–3364
- [251] Coto P B, Sinicropi A, Vico L D, Ferré N and Olivucci M 2006 *Mol. Phys.* **104** 983–991
- [252] Yamada A, Kakitani T, Yamamoto S and Yamato T 2002 *Chem. Phys. Lett.* **366** 670–675
- [253] Schreiber M, Buss V and Sugihara M 2003 *J. Chem. Phys.* **119** 12045–12048
- [254] Ferré N and Olivucci M 2003 *J. Am. Chem. Soc.* **125** 6868–6869
- [255] Hufen J, Sugihara M and Buss V 2004 *J. Phys. Chem. B* **108** 20419–20426
- [256] Sugihara M, Buss V, Entel P and Hafner J 2004 *J. Phys. Chem. B* **108** 3673–3680
- [257] Ferré N, Cembran A, Garavelli M and Olivucci M 2004 *Theor. Chem. Acc.* **112** 1–7
- [258] Fujimoto K, Hasegawa J Y, Hayashi S, Kato S and Nakatsuji H 2005 *Chem. Phys. Lett.* **414** 239–242
- [259] Gascon J A, Sproviero E M and Batista V S 2006 *Acc. Chem. Res.* **39** 184–193
- [260] Coto P B, Strambi A, Ferré N and Olivucci M 2006 *Proc. Natl. Acad. Sci. U. S. A.* **103** 17154–17159
- [261] Trabanino R J, Vaidehi N and William A Goddard I 2006 *J. Phys. Chem. B* **110** 17230–17239
- [262] Matsuura A, Sato H, Houjou H, Saito S, Hayashi T and Sakurai M 2006 *J. Comput. Chem.* **27** 1623–1630
- [263] Fujimoto K, Hayashi S, Hasegawa J Y and Nakatsuji H 2007 *J. Chem. Theory Comput.* **3** 605–618
- [264] Bravaya K, Bochenkova A, Granovsky A and Nemukhin A 2007 *J. Am. Chem. Soc.* **129** 13035–13042
- [265] Hall K F, Vreven T, Frisch M J and Bearpark M J 2008 *J.*

- Mol. Biol.* **383** 106–121
- [266] Sekharan S and Buss V 2008 *J. Am. Chem. Soc.* **130** 17220–17221
- [267] Altun A, Yokoyama S and Morokuma K 2008 *Photochem. Photobiol.* **84** 845–854
- [268] Altun A, Yokoyama S and Morokuma K 2008 *J. Phys. Chem. B* **112** 6814–6827
- [269] Altun A, Yokoyama S and Morokuma K 2008 *J. Phys. Chem. B* **112** 16883–16890
- [270] Wanko M, Hoffmann M, Frähmcke J, Frauenheim T, and Elstner M 2008 *J. Phys. Chem. B* **112** 11468–11478
- [271] Standfuss J, Zaitseva E, Mahalingam M, and Vogel R 2008 *J. Mol. Biol.* **380** 145–157
- [272] Schapiro I, Weingart O and Buss V 2009 *J. Am. Chem. Soc.* **131** 16–17
- [273] Hayashi S, Tajkhorshid E and Schulten K 2009 *Biophys. J.* **96** 403–416
- [274] Rajamani R, Lin Y L and Gao J 2010 *J. Comput. Chem.* **32** 854–865
- [275] Frähmcke J S, Wanko M, Phatak P, Mroginski M A and Elstner M 2010 *J. Phys. Chem. B* **114** 11338–11352
- [276] Sekharan S and Morokuma K 2011 *J. Am. Chem. Soc.* **133** 19052–19055
- [277] Sekharan S and Morokuma K 2011 *J. Am. Chem. Soc.* **133** 4734–4737
- [278] Sandberg M N, Amora T L, Ramos L S, Chen M H, Knox B E and Birge R R 2011 *J. Am. Chem. Soc.* **133** 2808–2811
- [279] Gozem S, Schapiro I, Ferré N and Olivucci M 2012 *Science* **337** 1225–1228
- [280] Ville R I Kaila and Robert Send and Dage Sundholm 2012 *J. Phys. Chem. B* **116** 2249–2258
- [281] Hernández-Rodríguez E W, Montero-Alejo A L, López R, Sánchez-García E, Montero-Cabrera L A and de la Vega J M G 2013 *J. Comput. Chem.* **34** 2460–2471
- [282] Doemer M, Maurer P, Campomanes P, Tavernelli I and Rothlisberger U 2014 *J. Chem. Theory Comput.* **10** 412–422
- [283] Sandberg M N, Greco J A, Wagner N L, Amora T L, Ramos L A, Chen M H, Knox B E and Birge R R 2014 *SOJ Biochem.* **1** 1–12
- [284] Andersen L H, Nielsen I B, Kristensen M B, El Ghazaly M O A, Haacke S, Nielsen M B and Petersen M A 2005 *J. Am. Chem. Soc.* **127** 12347–12350
- [285] Nielsen I B, Lammich L and Andersen L H 2006 *Phys. Rev. Lett.* **96** 018304–4
- [286] Rajput J, Rahbek D B, Andersen L H, Hirshfeld A, Sheves M, Altoé P, Orlandi G, and Garavelli M 2010 *Angew. Chem. Int. Ed.* **49** 1790–1793
- [287] Rao V R and Oprian D D 1996 *Annu. Rev. Biophys. Biomol. Struct.* **25** 287–314
- [288] Sakmar T P, Menon S T, Marin E P and Award E S 2002 *Annu. Rev. Biophys. Biomol. Struct.* **31** 443–484
- [289] Filipek S, Stenkamp R E, Teller D C and Palczewski K 2003 *Annu. Rev. Physiol.* **65** 851–879
- [290] Garavelli M, Celani P, Bernardi F, Robb M A and Olivucci M 1997 *J. Am. Chem. Soc.* **119** 6891–6901
- [291] Vreven T, Bernardi F, Garavelli M, Olivucci M, Robb M A and Schlegel H B 1997 *J. Am. Chem. Soc.* **119** 12687–12688
- [292] Garavelli M, Vreven T, Celani P, Bernardi F, Robb M A and Olivucci M 1998 *J. Am. Chem. Soc.* **120** 1285–1288
- [293] Gonzalez-Luque R, Garavelli M, Bernardi F, Merchan M, Robb M A and Olivucci M 2000 *Proc. Natl. Acad. Sci. U. S. A.* **97** 9379–9384
- [294] Martin C H and Birge R R 1998 *J. Phys. Chem. A* **102** 852–860
- [295] Migani A, Robb M A and Olivucci M 2003 *J. Am. Chem. Soc.* **125** 2804–2808
- [296] Fantacci S, Migani A and Olivucci M 2004 *J. Phys. Chem. A* **108** 1208–1213
- [297] Weingart O, Migani A, Olivucci M, Robb M A, Buss V and Hunt P 2004 *J. Phys. Chem. A* **108** 4685–4693
- [298] Wanko M, Garavelli M, Bernardi F, Niehaus T A, Frauenheim T and Elstner M 2004 *J. Chem. Phys.* **120** 1674–1692
- [299] Weingart O, Buss V and Robb M A 2005 *Phase Trans.* **78** 17–24
- [300] Wanko M, Hoffmann M, Strodel P, Koslowski A, Thiel W, Neese F, Frauenheim T and Elstner M 2005 *J. Phys. Chem. B* **109** 3606–3615
- [301] Weingart O, Schapiro I and Buss V 2007 *J. Phys. Chem. B* **111** 3782–3788
- [302] Zaari R R and Wong S Y 2009 *Chem. Phys. Lett.* **469** 224–228
- [303] Keal T W, Wanko M and Thiel W 2009 *Theor. Chem. Acc.* **123** 145–156
- [304] Send R, Sundholm D, Johansson M P and Pawlowski F 2009 *J. Chem. Theory. Comput.* **5** 2401–2414
- [305] Gozem S, Huntress M, Schapiro I, Lindh R, Granovsky A A, Angeli C and Olivucci M 2012 *J. Chem. Theory Comput.* **8** 4069–4080
- [306] Gozem S, Krylov A I and Olivucci M 2013 *J. Chem. Theory Comput.* **9** 284–292
- [307] Xu X, Gozem S, Olivucci M and Truhlar D G 2013 *J. Phys. Chem. Lett.* **4** 253–258
- [308] Huix-Rotllant M, Filatov M, Gozem S, Schapiro I, Olivucci M and Ferré N 2013 *J. Chem. Theory Comput.* **9** 3917–3932
- [309] Gozem S, Melaccio F, Lindh R, Krylov A I, Granovsky A A, Angeli C and Olivucci M 2013 *J. Chem. Theory Comput.* **9** 4495–4506
- [310] Schapiro I and Neese F 2014 *Comp. Theor. Chem.* **1040-1041** 84–98
- [311] Gozem S, Melaccio F, Valentini A, Filatov M, Huix-Rotllant M, Ferré N, Frutos L M, Angeli C, Krylov A I, Granovsky A A, Lindh R and Olivucci M 2014 *J. Chem. Theory. Comput.* **10** 3074–3084
- [312] Valsson O, Angeli C and Filippi C 2012 *Phys. Chem. Chem. Phys.* **14** 11015–11020
- [313] Rostov I V, Amos R D, Kobayashi R, Scalmani G, and Frisch M J 2010 *J. Phys. Chem. B* **114** 5547–5555
- [314] Cembran A, Gonzalez-Luque R, Altoé P, Merchan M, Bernardi F, Olivucci M and Garavelli M 2005 *J. Phys. Chem. A* **109** 6597–6605
- [315] Sekharan S, Weingart O and Buss V 2006 *Biophys. J.* **91** L07–L09
- [316] Schreiber M and Buss V 2003 *Int. J. Quantum Chem.* **95** 882–889
- [317] Sugihara M, Buss V, Entel P, Elstner M and Frauenheim T 2003 *Biochem.* **41** 15259–15266
- [318] Vreven T and Morokuma K 2003 *Theor. Chem. Acc.* **109** 125–132
- [319] Blomgren F and Larsson S 2005 *J. Comput. Chem.* **26** 738–742
- [320] Blomgren F and Larsson S 2005 *J. Phys. Chem. B* **109** 9104–9110
- [321] Sugihara M, Hufen J and Buss V 2006 *Biochem.* **45** 801–810
- [322] Altoé P, Bernardi F, Conti I, Garavelli M, Negri F and Orlandi G 2007 *Theor. Chem. Acc.* **117** 1041–1059
- [323] Altoé P, Stenta M, Bottoni A and Garavelli M 2007 *Theor. Chem. Acc.* **118** 219–240
- [324] Sekharan S, Sugihara M, Weingart O, Okada T and Buss V 2007 *J. Am. Chem. Soc.* **129** 1052–1054
- [325] Sugihara M and Buss V 2008 *Biochem.* **47** 13733–13735
- [326] Strambi A, Coto P B, Frutos L M, Ferré N and Olivucci M 2008 *J. Am. Chem. Soc.* **130** 3382–3388
- [327] Altoé P, Cembran A, Olivucci M and Garavelli M 2011 *Proc. Natl. Acad. Sci. U. S. A.* **107** 20172–20177
- [328] Hasegawa J, Fujimoto K J and Kawatsu T 2012 *J. Chem. Theory Comput.* **8** 4452–4461
- [329] Chen M H, Kuemmel C, Birge R R and Knox B E 2012 *Biochem.* **51** 4117–4125
- [330] Huntress M M, Gozem S, Malley K R, Jailaubekov A E, Vasileiou C, Vengris M, Geiger J H, Borhan B, Schapiro I, Larsen D S and Olivucci M 2013 *J. Phys. Chem. B* **117** 10053–10070
- [331] Polli D, Weingart O, Brida D, Poli E, Maiuri M, Spillane K M, Bottoni A, Kukura P, Mathies R A, Cerullo G and Garavelli M 2014 *Angew. Chem. Int. Ed.* **53** 2504–2507
- [332] Buss V, Weingart O and Sugihara M 2000 *Angew. Chem. Int. Ed.* **39** 2784–2786
- [333] Schlegel B, Sippl W and Höltje H D 2005 *J. Mol. Model.* **12** 49–64

- [334] Klaffki N, Weingart O, Garavelli M and Spohr E 2012 *Chem. Phys. Phys. Chem.* **14** 14299–14305
- [335] Vico L D, Garavelli M, Bernardi F, and Olivucci M 2005 *J. Am. Chem. Soc.* **127** 2433–2442
- [336] Weingart O, Altoé P, Stenta M, Bottoni A, Orlandi G and Garavelli M 2011 *Phys. Chem. Chem. Phys.* **13** 3645–3648
- [337] Chung W C, Nanbu S and Ishida T 2012 *J. Phys. Chem. B* **116** 8009–8023
- [338] Rivalta I, Nenov A, Weingart O, Cerullo G, Garavelli M and Mukamel S 2014 *J. Phys. Chem. B* **118** 8396–8405
- [339] Schapiro I, Roca-Sanjuán D, Lindh R and Olivucci M 2015 *J. Comput. Chem.* **36** 312–320
- [340] Polli D, Rivalta I, Nenov A, Weingart O, Garavelli M and Cerullo G 2015 *Photochem. Photobiol. Sci.* **14** 213–228
- [341] Palczewski K, Kumasaka T, Hori T, Behnke C A, Motoshima H, Fox B A, Le Trong I, Teller D C, Okada T, Stenkamp R E, Yamamoto M and Miyano M 2000 *Science* **289** 739–745
- [342] Teller D C, Okada T, Behnke C A, Palczewski K, and Stenkamp R E 2001 *Biochem.* **40** 7761–7772
- [343] Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau E M and Shichida Y 2002 *Proc. Natl. Acad. Sci. U. S. A.* **99** 5982–5987
- [344] Okada T, Sugihara M, Bondar A N, Elstner M, Entel P and Buss V 2004 *J. Mol. Biol.* **342** 571–583
- [345] Nakamichi H, Buss V and Okada T 2007 *Biophys. J.* **92** L106–L108
- [346] Schreiber M and Buss V 2003 *Int. J. Quantum Chem.* **95** 882–889
- [347] Schreiber M, Sugihara M, Okada T and Buss V 2006 *Angew. Chem. Int. Ed.* **45** 4274–4277
- [348] Lüdeke S, Beck M, Yan E C Y, Sakmar T P, Siebert F and Vogel R 2005 *J. Mol. Biol.* **353** 345–356
- [349] Röhrig U F, Guidoni L and Rothlisberger U 2002 *Biochemistry* **41** 10799–10809
- [350] Fahmy K, Jager F, Beck M, Zvyaga T A, Sakmar T P and Siebert F 1993 *Proc. Natl. Acad. Sci. U. S. A.* **90** 10206–10210
- [351] Martínez-Mayorga K, Pitman M C, Grossfield A, Feller S E, and Brown M F 2006 *J. Am. Chem. Soc.* **128** 16502–16503
- [352] Sekharan S, Sugihara M and Buss V 2007 *Angew. Chem. Int. Ed.* **46** 269–271
- [353] Fujimoto K, Hasegawa J Y and Nakatsuji H 2009 *Bull. Chem. Soc. Jpn.* **82** 1140–1148
- [354] Strambi A, Coto P B, Ferré N and Olivucci M 2007 *Theor. Chem. Acc.* **118** 185–191
- [355] Andruniow T and Olivucci M 2009 *J. Chem. Theory. Comput.* **5** 3096–3104
- [356] Hayashi S and Ohmine I 2000 *J. Phys. Chem. B* **104** 10678–10691
- [357] Houjou H, Inoue Y, and Sakurai M 2001 *J. Phys. Chem. B* **105** 867–879
- [358] Warshel A and Chu Z T 2001 *J. Phys. Chem. B* **105** 9857–9871
- [359] Hayashi S, Tajkhorshid E, Pebay-Peyroula E, Royant A, Landau E M, Navarro J and Schulten K 2001 *J. Phys. Chem. B* **105** 10124–10131
- [360] Hayashi S, Tajkhorshid E and Schulten K 2002 *Biophys. J.* **83** 1281–1297
- [361] Rajamani R and Gao J 2002 *J. Comput. Chem.* **23** 96–105
- [362] Hoffmann M, Wanko M, Strodel P, König P H, Frauenheim T, Schulten K, Thiel W, Tajkhorshid E and Elstner M 2006 *J. Am. Chem. Soc.* **128** 10808–10818
- [363] Prokhorenko V I, Nagy A M, Waschuk S A, Brown L S, Birge R R and Miller R J D 2006 *Science* **313** 1257–1261
- [364] Li X, Chung L W and Morokuma K 2011 *J. Chem. Theory Comput.* **7** 2694–2698
- [365] Suomivuori C M, Lang L, Sundholm D, Gamiz-Hernandez A P and Kaila V R I 2016 *Chem. Eur. J.* **22** 1–9
- [366] Melaccio F, Ferré N and Olivucci M 2012 *Phys. Chem. Chem. Phys.* **14** 12485–12495
- [367] Hernández-Rodríguez E W, Sánchez-García E, Crespo-Otero R, Montero-Alejo A L, Montero L A and Thiel W 2012 *J. Phys. Chem. B* **116** 1060–1076
- [368] Altun A, Yokoyama S and Morokuma K 2009 *J. Phys. Chem. A* **113** 11685–11692
- [369] Fujimoto K, Hasegawa J Y and Nakatsuji H 2008 *Chem. Phys. Lett.* **318** 318–320
- [370] Ryazantsev M N, Altun A and Morokuma K 2012 *J. Am. Chem. Soc.* **134** 5520–5523
- [371] Sekharan S, Altun A and Morokuma K 2010 *J. Am. Chem. Soc.* **132** 15856–15859
- [372] Sekharan S, Altun A and Morokuma K 2010 *Chem. Eur. J* **16** 1744–1749
- [373] Sekhara S and Morokuma K 2010 *J. Phys. Chem. Lett.* **1** 668–672
- [374] Rangarajan R, Galan J F, Whited G and Birge R R 2007 *Biochem.* **46** 12679–12686
- [375] Sekharan S, Yokoyama S and Morokuma K 2011 *J. Phys. Chem. B* **115** 15380–15388
- [376] Strambi A, Durbeej B, Ferré N and Olivucci M 2011 *Proc. Natl. Acad. Sci. U. S. A.* **107** 21322–21326
- [377] Tada T, Altun A and Yokoyama S 2009 *Proc. Natl. Acad. Sci. U. S. A.* **106** 17457–17462
- [378] Send R, Suomivuori C M, Kaila V R I and Sundholm D 2015 *J. Phys. Chem. B* **119** 2933–2945
- [379] Creemers T, Lock A, Subramaniam V, Jovin T and Völker S 1999 *Nat. Struct. Mol. Biol.* **6** 557–560
- [380] Brejc K, Sixma T K, Kitts P A, Kain S R, Tsien R Y, Ormö M and Remington S J 1997 *Proc. Natl. Acad. Sci. U. S. A.* **94** 2306–2311
- [381] Sinicropi A, Andruniow T, Ferré N, Basosi R and Olivucci M 2005 *J. Am. Chem. Soc.* **127** 11534–11535
- [382] Marques M A, López X, Varsano D, Castro A and Rubio A 2003 *Phys. Rev. Lett.* **90** 258101
- [383] Laino T, Nifosí R and Tozzini V 2004 *Chem. Phys.* **298** 17–28
- [384] Hasegawa J Y, Fujimoto K, Swerts B, Miyahara T and Nakatsuji H 2007 *J. Comput. Chem.* **28** 2443–2452
- [385] Bravaya K B, Khrenova M G, Grigorenko B L, Nemukhin A V and Krylov A I 2011 *J. Phys. Chem. B* **115** 8296–8303
- [386] Steindal A H, Olsen J M H, Ruud K, Frediani L and Kongsted J 2012 *Phys. Chem. Chem. Phys.* **14** 5440
- [387] Kaila V R I, Send R and Sundholm D 2013 *Phys. Chem. Chem. Phys.* **15** 4491
- [388] Petrone A, Caruso P, Tenuta S and Rega N 2013 *Phys. Chem. Chem. Phys.* **15** 20536
- [389] Amat P and Nifosí R 2013 *J. Chem. Theory Comput.* **9** 497–508
- [390] Grigorenko B L, Nemukhin A V, Polyakov I V, Morozov D I and Krylov A I 2013 *J. Am. Chem. Soc.* **135** 11541–11549
- [391] Pikulska A, Steindal A H, Beerepoot M T P and Pecul M 2015 *J. Phys. Chem. B* **119** 3377–3386
- [392] Daday C, Curutchet C, Sinicropi A, Mennucci B and Filippi C 2015 *J. Chem. Theory Comput.* **11** 4825–4839
- [393] Jacchetti E, Gabellieri E, Cioni P, Bizzarri R and Nifosí R 2016 *Phys. Chem. Chem. Phys.* **18** 12828–12838
- [394] Gordon M S, Freitag M A, Bandyopadhyay P, Jensen J H, Kairys V and Stevens W J 2001 *J. Phys. Chem. A* **105** 293–307
- [395] Dong J, Solntsev K M and Tolbert L M 2006 *J. Am. Chem. Soc.* **128** 12038–12039
- [396] Shinobu A, Palm G J, Schierbeek A J and Agmon N 2010 *J. Am. Chem. Soc.* **132** 11093–11102
- [397] Isborn C M, Gotz A W, Clark M A, Walker R C and Martínez T J 2012 *Journal of chemical theory and computation* **8** 5092–5106
- [398] Nielsen S B, Lapierre A, Andersen J U, Pedersen U, Tomita S and Andersen L 2001 *Phys. Rev. Lett.* **87** 228102
- [399] Lammich L, Petersen M Å, Nielsen M B and Andersen L H 2007 *Biophys. J.* **92** 201–207
- [400] Forbes M W and Jockusch R A 2009 *J. Am. Chem. Soc.* **131** 17038–17039
- [401] Chingin K, Balabin R M, Frankevich V, Barylyuk K, Nieckarz R, Sagulenko P and Zenobi R 2011 *Int. J. Mass Spectrom.* **306** 241–245
- [402] Häse F, Valleau S, Pyzer-Knapp E and Aspuru-Guzik A 2016 *Chem. Sci.* **7** 5139–5147