The Journal of Chemical Physics

The atomistic modeling of light-harvesting complexes from the physical models to the computational protocol

Cite as: J. Chem. Phys. **156**, 120901 (2022); https://doi.org/10.1063/5.0086275
Submitted: 24 January 2022 • Accepted: 01 March 2022 • Accepted Manuscript Online: 02 March 2022 • Published Online: 22 March 2022

Edoardo Cignoni, Vladislav Slama, Lorenzo Cupellini, et al.

openaccess
View Online
Export Citation
CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Excited states of chlorophyll a and b in solution by time-dependent density functional theory
The Journal of Chemical Physics (2022); https://doi.org/10.1063/5.0083395

Computational spectroscopy of complex systems
The Journal of Chemical Physics **155**, 170901 (2021); https://doi.org/10.1063/5.0064092

Chlorophyll a de-excitation pathways in the LHCII antenna
The Journal of Chemical Physics **156**, 070902 (2022); https://doi.org/10.1063/5.0073825

© 2022 Author(s).

AIP Publishing

The Journal of Chemical Physics PERSPECTIVE scitation.org/journal/jcp

The atomistic modeling of light-harvesting complexes from the physical models to the computational protocol

Cite as: J. Chem. Phys. 156, 120901 (2022); doi: 10.1063/5.0086275
Submitted: 24 January 2022 • Accepted: 1 March 2022 •
Published Online: 22 March 2022

View Online Export Citation CrossMark

Edoardo Cignoni, Vladislav Slama, Lorenzo Cupellini, and Benedetta Mennucci[a)]

AFFILIATIONS

Dipartimento di Chimica e Chimica Industriale, University of Pisa, via G. Moruzzi 13, 56124 Pisa, Italy

Note: This paper is part of the JCP Special Topic on Photosynthetic Light-Harvesting and Energy Conversion.
[a)]**Author to whom correspondence should be addressed:** benedetta.mennucci@unipi.it

ABSTRACT

The function of light-harvesting complexes is determined by a complex network of dynamic interactions among all the different components: the aggregate of pigments, the protein, and the surrounding environment. Complete and reliable predictions on these types of composite systems can be only achieved with an atomistic description. In the last few decades, there have been important advances in the atomistic modeling of light-harvesting complexes. These advances have involved both the completeness of the physical models and the accuracy and effectiveness of the computational protocols. In this Perspective, we present an overview of the main theoretical and computational breakthroughs attained so far in the field, with particular focus on the important role played by the protein and its dynamics. We then discuss the open problems in their accurate modeling that still need to be addressed. To illustrate an effective computational workflow for the modeling of light harvesting complexes, we take as an example the plant antenna complex CP29 and its H111N mutant.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0086275

I. INTRODUCTION

In photosynthesis, the initial step is accomplished by specialized supercomplexes, photosystems, which comprise two components, light harvesting (LH) pigment–protein complexes, and the reaction center. LH complexes are unique biological systems, which integrate a protein matrix with a generally large aggregate of pigments. The number and type of pigments and the structure of the protein matrix are specific for each different photosynthetic organism.[1] Through the multichromophoric aggregate, the LH complex absorbs sunlight and rapidly funnels the energy toward neighboring complexes and ultimately to the reaction center.[2–4] The energy delivery by LH complexes is explained in terms of multichromophoric excited electronic states called excitons. Each exciton generally involves more than one pigment and is characterized by an energy slightly different from the others even if the aggregate is composed of the same types of molecules. The resulting energy ladder is not only determined by the strength of the interactions between excitations localized on the single pigments but also significantly tuned by the interactions of each pigment with the surrounding residues and the electric fields created by the whole protein and the external environment. Moreover, the excitons are coupled with the nuclear degrees of freedom of the multichromophoric aggregate and the protein.

Due to the complexity outlined above, an accurate atomistic modeling of LH complexes is extremely challenging, as it requires the integration of very different physical models, which span from quantum mechanical (QM) methods to classical models, dynamic descriptions, and statistical analyses. Moreover, for that integration to work, the different models need to be all coupled together, as the functioning of the LH complex is a real multiscale process in space and time.

In the last few decades, there have been important advances in the atomistic modeling of LH complexes,[5–10] which have been largely pushed by the increasing amount of experimental data about their structures, the spectroscopic properties, and the energy

© Author(s) 2022

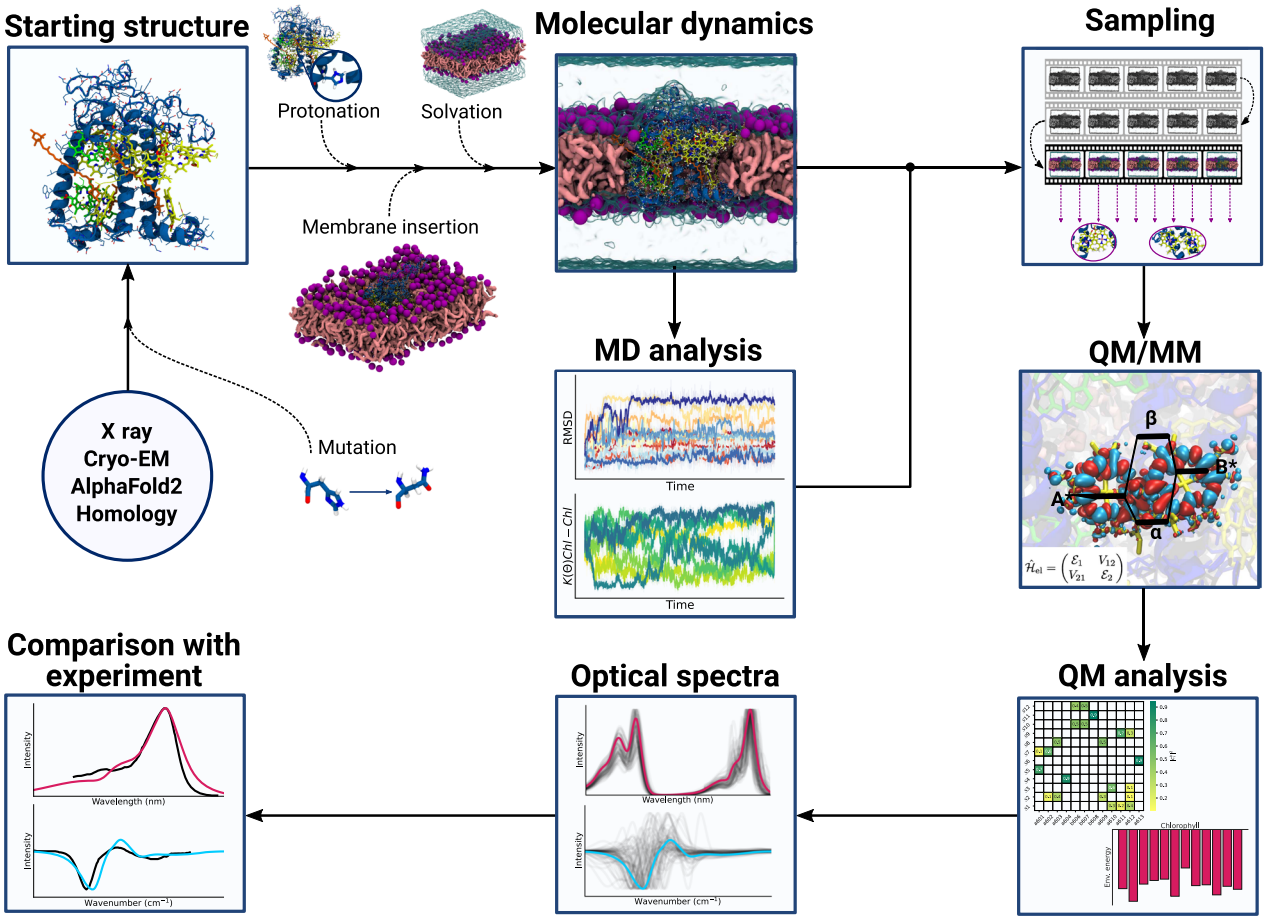

FIG. 1. General workflow for the atomistic modeling of light harvesting complexes. Atomistic simulations of LH complexes start either from experimentally determined structures (determined from, for example, x ray or cryo-EM) or from computational guesses, the latter being either traditional homology modeling or newer neural-network based algorithms, such as AlphaFold 2. Amino acidic mutations can be introduced in order to exploit atomistic simulations for complementing mutagenesis experiments. The starting structure is then protonated, inserted in a membrane, and solvated in order to simulate an environment that mimics the *in vivo* conditions. Molecular dynamics (MD) simulations are then run to sample LH conformations at the desired external conditions of temperature and pressure. The analysis of the MD simulations (both in terms of global convergence tests and of a more focused analysis of the embedded pigments) is employed to guide the sampling of the pigment conformations to be used in QM/MM calculations. The first part of the trajectory is usually discarded from the sampling (gray film in the "sampling" box) so as to avoid sampling transient, non-equilibrated structures (see Sec. II). QM/MM calculations are run over the sampled pigment conformations in order to extract the exciton parameters for simulating the spectra (see Secs. III and IV). The obtained results are the subject of a careful analysis and are post-processed in order to deal with, e.g., an inaccurate treatment of the vibrational degrees of freedom of the pigments (Sec. IV B). The post-processed data are finally employed to compute the optical spectra, the quality of which is assessed by a comparison with experimental data. An application of this workflow is provided in Sec. V.

transfer processes.[11–17] In parallel, there has been an important activity in the development of computational methods that have allowed for simulating properties and processes of LH complexes embedded in different environments.[18–21]

This Perspective reviews these advances using the typical workflow followed in the computational investigation of a LH complex, which starts from a predicted or experimentally characterized structure, adds the environment, simulates the coupled dynamics of all the parts, calculates the excitonic states, and finally reproduces the optical spectra to be directly compared with experiments. A graphical representation of the workflow with the indication of all the different components is given in Fig. 1. In the presentation and discussion of the different parts of the workflow, we use a minor LH complex present in PSII of plants (CP29) as an example of application. Finally, the optical spectra of the complex will be compared with a recently proposed mutant[22] to show the accuracy and completeness that the atomistic modeling can achieve in predicting even small effects as those induced by a single residue mutation.

II. SETUP OF THE SYSTEM AND CONFIGURATIONAL SAMPLING

At the present time, LH complexes of different photosynthetic organisms have been structurally characterized

© Author(s) 2022

through x-ray diffraction or cryo-EM techniques. These structures represent a fundamental ingredient for the atomistic models; however, they also present important limitations. First of all, they do not include any information on the external environment, which for most of the LH complexes is a membrane. Moreover, fluctuations due to the temperature are not accounted for. Finally, the standard resolution of crystal or cryo-EM structures of LH complexes is not high enough to describe the internal geometry of the pigments with the atomic accuracy needed for safely applying QM calculations on them.

For all these reasons, an atomistic modeling of LH complexes based on only the crystal structure can give a qualitative description of the system at very low temperatures, but, in general, it cannot be representative of the room-temperature situation. To go beyond this too simplistic static picture, we can introduce molecular dynamics (MD) simulations. Molecular dynamics has been shown to be a very effective approach for sampling configurations of the LH complex and its environment at the temperature of interest.[21]

Molecular dynamics simulations of light-harvesting complexes usually started from the available x ray or cryo-EM structures deposited in public repositories.[23] Another means of obtaining starting structures for MD simulations when the experimental structures are not available is to employ homology modeling techniques, which have also been employed for LH complexes.[24–26] These techniques can be employed if the target protein sequence is sufficiently similar to the sequences of known structures. In addition, recent algorithms based on neural networks, such as AlphaFold 2,[27,28] have proven able to determine protein structures as accurately as their experimental counterparts, providing another means of obtaining starting structures for molecular simulations. Especially when the starting structure is chosen based on a computational guess, such as in homology modeling or by employing neural networks, MD simulations and their spectroscopic analysis can be employed to validate the structures and gain insights on the architecture of systems that are still to be structurally resolved experimentally. Finally, the starting point being either an experimental structure or a computational guess, it is easy to slightly alter the initial structure so as to introduce, e.g., point mutations. Especially in the context of investigating LH complexes, the simulation of mutants can be significantly valuable as it complements site-directed mutagenesis experiments. We provide such an example in Sec. V.

Due to the complexity and the large dimension of the whole system (comprising the complex and the surrounding environment), a simplified description is commonly used for the potential energy (and the corresponding forces) in terms of Molecular Mechanics (MM) Force Fields (FFs). At the present times, accurate FFs are available to describe protein structures and their thermal fluctuations and for describing water and lipid membranes. However, the application of the available FFs to the pigments is by far less accurate. This is the reason why a reparameterization of the FF for describing the nuclear degrees of freedom of the pigments is generally introduced. In this reparameterization of the FF, QM calculations of internal geometry and vibrational properties are commonly used as reference data.[29–31]

Once the potential energy model of all the components has been defined, the system can be assembled and its trajectory can be determined following the same protocols used in MD simulations of biomacromolecules. In order to achieve a meaningful sampling of the configurational space, ns timescales are often not sufficient. We have observed fluctuations on the 100 ns timescale in LH systems, and we, therefore, recommend μs timescales whenever possible.[32] We also recommend running several replicas of the same system obtained using different initial velocities. Different replicas are then subject to standard convergence assessment tests, such as monitoring the kinetic energy of the system, the root-mean-square deviation (RMSD) of different parts of the protein, and the average area available to each lipid in the plane of the membrane. Other quantities, such as selected coordinates or order parameters, may be used to assess equilibration. Averages of these quantities may systematically drift during a transient time, indicating that the system has yet to sample the equilibrium ensemble. When the averages reach a plateau or oscillate around a mean value, the trajectory can be considered locally equilibrated. The transient part should be discarded before further analysis and especially before computing excitation properties.

The distinct feature of LH complexes is the dense packing of pigments in highly specific arrangements that maximize energy transfer while avoiding concentration quenching.[3] At variance with other biomacromolecular systems, then, the analysis of LH complex trajectories should pay particular attention to the relative arrangement of the embedded pigments, i.e., their relative distances and orientations. Pigment–pigment orientation measures that are directly linked to their interaction strength can be based on the Förster orientational factor κ [see Eq. (5) in Sec. III B]. In addition, analyzing the hydrogen bond interactions between the protein and the bound pigments can provide useful insights on interpreting the spectroscopic properties of the complex.[25] This kind of analysis provides information on the interaction network within the LH complex and on how it is influenced by the protein dynamics.

Further effective analysis instruments aiming at disentangling the complex conformational landscapes in LH complexes can be borrowed from the ever-growing field of unsupervised learning.[33] In particular, dimensionality reduction methods [such as principal component analysis (PCA)[34] and Isomap[35]] can be employed to consistently reduce the redundancy in the underlying conformational ensemble starting from properly featured protein conformations,[36] thereby improving the ease of characterization of protein conformations. In a similar spirit, kinetic-based independent component analysis algorithms, such as time-lagged independent component analysis (tICA),[37] can also be used to characterize LH complexes conformations while conserving some degree of information lying in the temporal structure of the trajectory.[38] Additionally, clustering techniques can be adopted to group similar conformations together, effectively coarse-graining the explored structures and facilitating the identification of key structural differences between the observed conformers. When running subsequent QM/MM calculations, an *a priori* division of the explored ensemble into clusters may be useful to reduce the number of calculations while ensuring an accurate sampling of the LH complex conformations. In addition, clustering can be used to connect structural features to spectroscopic observables.[26,36]

© Author(s) 2022

III. THE ELECTRONIC DESCRIPTION OF THE MULTICHROMOPHORIC AGGREGATE

The electronic properties of the multichromophoric aggregate of LH complexes are commonly described through the Frenkel exciton model. In this model, the electronic excited states of the entire aggregate are expanded on a basis of excitations localized on single pigments. In this basis, the excited states of the aggregate are described by the following exciton Hamiltonian:

$$\hat{\mathscr{H}}_{\mathrm{ex}} = \sum_i \mathscr{E}_i |i\rangle\langle i| + \sum_{i\neq j} V_{ij} |i\rangle\langle j|, \tag{1}$$

where $\mathscr{E}_i$ are denoted as *site energies* and represent the excitation energies of the isolated pigments and V_{ij} are the *excitonic couplings* between different pigments. Here, we assume for simplicity that each pigment contributes with one excitation. The wavefunction of the exciton state K can be expressed in terms of the exciton coefficients c_i^K,

$$|K\rangle = \sum_i c_i^K |i\rangle. \tag{2}$$

From a strictly computational point of view, the exciton model can be seen as a "divide and conquer" method, where the complexity of the multichromophoric aggregate is reduced by separating the problem into smaller elements. In fact, the matrix elements of the Hamiltonian (1) can be easily computed from quantum chemical calculations on single pigments. However, the exciton model has enormous advantages that go beyond the mere computational efficiency. Strikingly, the exciton coefficients allow for a straightforward interpretation of the aggregate's excited states and of their energy variations. As we will see in Sec. IV, this model also enables the description and calculation of optical line shapes.

Physically, the exciton model corresponds to neglecting all charge-transfer contributions to the aggregate excited states. This is generally a reasonable assumption as the inter-pigment distances in most of the LH complexes are small enough to have non-negligible excitonic couplings, but at the same time, they are large enough to avoid significant charge transfer. However, when needed, the Hamiltonian (1) can be augmented with charge-transfer states by including their energies and couplings with localized excitations.[39,40]

A. Site energies

Site energies represent the first main ingredient in the exciton Hamiltonian. Each site energy can be computed as the excitation energy of a pigment embedded in its environment. As such, calculation of these quantities requires an appropriate QM method and a description for the environment. Among the QM methods, a fair compromise between computational cost and accuracy is represented by linear-response time-dependent density functional theory (TD-DFT). Despite its well-known deficiencies, TD-DFT represents a valid and extensively tested method to compute excited states. More accurate QM methods have been employed for light-harvesting complexes, but their use has been mostly limited to benchmark calculations. Multireference methods, such as multireference configuration interaction-DFT (MRCI-DFT) and complete-active-space SCF (CASSCF), possibly in combination with perturbation theory (CASPT2), have been used to complement or assess TD-DFT methods.[41–44] More recently, single-reference accurate methods have been employed. Approximate coupled cluster methods, such as second-order coupled cluster (CC2) and pair natural orbital coupled cluster [domain-based local pair natural orbital similarity transformed equation of motion-coupled cluster singles and doubles (DLPNO-STEOM-CCSD)], have been applied to photosynthetic pigments.[45–47] Other alternatives are post-SCF methods based on equations of motion or response approaches, such as the algebraic-diagrammatic construction (ADC) and the Bethe–Salpeter equation in the GW approximation (GW/BSE). These methods have found applications for photosynthetic pigments.[45,48] Semiempirical methods represent a much less expensive alternative to obtain excitation properties with reasonable accuracy. Among these, time-dependent density functional tight binding (TD-DFTB) and its long-range corrected (LC) version showed promise for describing light-harvesting pigments.[49]

Pigments in LH complexes are surrounded by a highly anisotropic environment, and they are bound in a protein pocket by establishing specific interactions, such as hydrogen bonds. Only a computational model that describes this anisotropy can reproduce excitation energy variations among the different sites of the complex. The simplest such model is quantum mechanics/molecular mechanics (QM/MM) with electrostatic embedding. In electrostatic embedding QM/MM, the chromophore is treated with the QM level of choice, whereas the environment is described as a set of fixed point charges. Irrespective of the QM description, this model introduces an external one-electron operator analogous to the electron–nucleus Coulomb interaction. The conceptual simplicity and ease of implementation coupled with any QM method have made this model extremely popular for excited-state calculations, also in light-harvesting complexes.[47,50,51] While electrostatic embedding QM/MM can describe the effect of the MM environment on the QM part, it cannot describe how the environment is affected by the presence of the QM electron density or by its changes in response to excitation.

A more refined QM/MM flavor is the polarizable embedding QM/MM.[52] Polarization has been included in QM/MM using various formulations based, e.g., on Drude oscillators,[53–55] fluctuating charges,[56–58] or induced point dipoles (IPDs).[59–66]

In any of these formulations, the environment and the QM density can mutually polarize each other. From the point of view of the QM method, this formulation introduces a nonlinear term in the electronic Hamiltonian through effective two-electron contributions. These terms render much less straightforward coupling a polarizable embedding to a new QM method. Moreover, the response of the MM part also becomes more complicated. Calculating the MM polarization requires to account for the mutual polarization between MM atoms. Within the IPD formulation, the induced dipoles on the MM part can be computed by solving the following system:[66,67]

$$A\boldsymbol{\mu}^{\mathrm{ind}} = \mathbf{E}^{\mathrm{ext}}, \tag{3}$$

where $\mathbf{E}^{\mathrm{ext}}$ is the electric field resulting from the QM density and the MM charges and A is a matrix containing the atomic polarizabilities and the interactions between all induced dipoles. To avoid the so-called "polarization catastrophe," e.g., the divergence of the Coulomb interaction between two point dipoles when they get too

© Author(s) 2022

close, distance dependent damping functions originally proposed by Thole[68] are commonly introduced.

All these specificities make polarizable QM/MM methods computationally more expensive than standard electrostatic embedding analogs. New linear-scaling implementations are required to reduce the computational overhead of the formulation.[69] The IPD formulation of polarizable QM/MM methods (from now on QM/MMPol) has been extensively used for calculating site energies of pigments within LH complexes coupled with linear-response TD-DFT methods.[19,25,70–72]

It is important to disentangle the different effects that the environment has on the site energies. The excitation energy of a chromophore can be modulated by the environment directly, through electrochromic effects, and indirectly by changing the internal geometry of the chromophore. These effects can be separated naturally by computing the excitation energy of the chromophore embedded in the MM(Pol) environment and isolated *in vacuo*.[19] The difference between the two calculations represents the direct electrochromic effect due to the environment. We stress that the geometry of the pigment must be the same in both calculations in order to obtain a pure electrochromic effect when taking the difference.

The indirect geometrical effect on the site energies can be defined by comparing the *in vacuo* site energies of each distinct chromophore with a reference geometry. The environment contribution is very sensitive to the fluctuations of the environment. As such, an adequate sampling of environment degrees of freedom via, e.g., MD simulations, is needed to obtain a realistic description of the chromophores embedded in LH proteins. Indeed, as shown in Fig. 2, an analysis based on the experimental structure only, ignoring thermal effects included with the MD, provides a misleading picture of the influence of the environment on the excitation energy of the embedded chromophores. This kind of analysis allows us to separate "geometrical" and "environment" effects on the site energies and to analyze the origin of the specific tuning of the site energy of each pigment.[72]

Despite their power, QM/MM calculations of excitation energies suffer from systematic errors stemming mainly from the quantum chemical method and, in part, on the method used for obtaining chromophore geometries. In fact, the expected accuracy of QM methods for excitation energies is at least around 0.2 eV for TD-DFT methods and ~0.1 eV for more accurate methods, such as ADC(2) and CC2.[73] These errors are larger than the typical width of site energies in LH complexes, but they do not prevent a quantitative application of QM/MM methods to computing site energy ladders. In fact, while these errors are large, they are usually systematic, and they do not affect excitation energy differences among chemically identical pigments. For this reason, it has become customary to rigidly shift the calculated spectra, or equivalently the site energies, by a certain amount that takes into account the intrinsic error of the QM level of theory.[25,51,72,74] This correction is justified by assuming that the error is systematic. In this respect, the same shift should be applied to different systems if they contain the same pigments and are treated at the same level of theory.

The situation is more complicated when a LH complex contains different types of pigments, such as Chl a and Chl b, or when multiple excited states need to be considered. There, it is not guaranteed that the QM method will give the same error for all excited states. For example, even the DLPNO-CCSD method gives different errors for the different transitions of Chl a.[46] A possible solution is to use a different parametric shift for different pigment

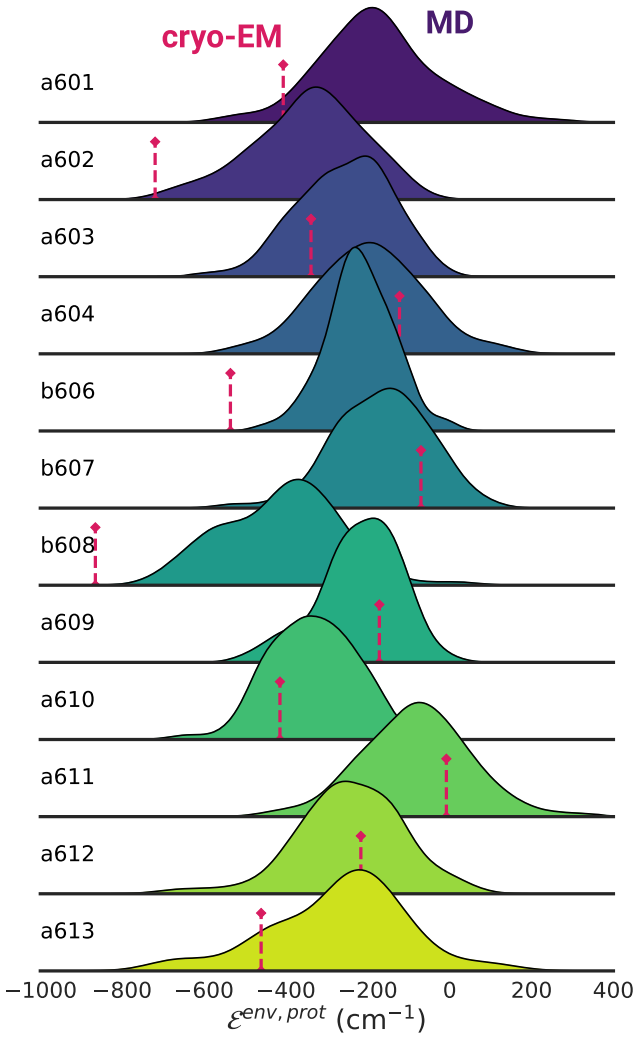

FIG. 2. Environment effect on chlorophylls in CP29. The distributions show the environment contribution to the site energy $\mathscr{E}^{\text{env,prot}}$ due to the protein environment only (i.e., without the membrane and waters) in the MD simulation of CP29. This contribution is obtained as $\mathscr{E}^{\text{env,prot}} = \mathscr{E}^{\text{QM/MMPol,prot}} - \mathscr{E}^{\text{geom}}$, where $\mathscr{E}^{\text{QM/MM,prot}}$ is the site energy obtained through a QM/MM calculation in the protein environment and $\mathscr{E}^{\text{geom}}$ is the site energy obtained from a vacuum QM calculation on the pigment alone (see also Sec. IV B). A distribution is shown for each chlorophyll present in the CP29 system. The label of each chlorophyll is shown on the left. The red dashed line represents the same contribution computed on the starting (cryo-EM) structure (PDB: 3JCU). Hydrogen atoms are optimized prior to the calculation in the cryo-EM structure. All values have been computed using a QM/MMpol model using the M062X/6-31G(d) level of calculation.

© Author(s) 2022

types. This approach seems quite straightforward, but it introduces a large number of empirical parameters. In some cases, the relative error for two different pigments is small enough not to influence substantially the spectral shape,[72] and it can be simply ignored. When the relative error cannot be ignored, the safe choice is to determine the shift parameters from a system other than the system of interest, and possibly simpler. For example, the excitation energies of chromophores in solution could be used to determine the relative error that a given QM method makes on different transitions.

A much simpler approach to environment effects on site energies of pigments in LH complexes is the use of continuum models.[75,76] These models lose the fine details of the protein environment, which makes the description necessarily less accurate than an atomistic model. However, continuum models can still represent a useful and inexpensive alternative to QM/MM especially if the QM subsystem is augmented to include the protein residues involved in specific interactions with the pigments.[70,77] A recent alternative is to couple both MMPol and a continuum model to the QM description of a chromophore.[78,79] This strategy has been applied to phycocyanin complexes of red algae.[80]

Other approaches are able to go beyond the classical description of the protein environment. The frozen density embedding (FDE) scheme and its extension to excited states (subsystem TD-DFT) divide the entire system into fragments, each described quantum mechanically and experiencing the embedding potential from the other QM fragments.[81] This method has been applied to computing site energies and couplings in LH complexes.[82]

B. Couplings

Electronic couplings between localized excitations determine exciton delocalization effects, which reflect on optical spectra of LH complexes and on the exciton dynamics.

The electronic coupling can be decomposed into a sum of a long-range Coulomb contribution and a short-range term.[83,84] The latter term is dependent on the overlap between orbitals of the interacting chromophores and is generally neglected. In the common case of bright excitations, the Coulomb term is dominant even when the two chromophores are in contact. Therefore, the coupling can be calculated as

$$V_{ij} \simeq V_{ij}^{\mathrm{Coul}} = \iint \frac{\rho_i^{\mathrm{tr}}(\vec{r})\rho_j^{\mathrm{tr}}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,\mathrm{d}^3r\,\mathrm{d}^3r', \tag{4}$$

where $\rho_{i/j}^{\mathrm{tr}}$ is the transition density of the excitation localized on chromophore i or j.

Equation (4) has a classical interpretation as the electrostatic interaction between two charge densities that represent the excitations of the coupled chromophores. For this reason, the Coulomb coupling is often approximated through a multipole expansion or a distributed monopole expansion. The first multipolar order is the interaction between transition dipoles $\vec{\mu}^{\mathrm{tr}}$,

$$V_{dd} = \frac{|\vec{\mu}_i^{\mathrm{tr}}||\vec{\mu}_j^{\mathrm{tr}}|\kappa_{ij}}{R_{ij}^3}, \tag{5}$$

which can be expressed solely in terms of the distance between chromophores R_{ij} and the orientational factor κ. While this approximation breaks down at the interchromophore distances relevant for LH complexes[84,85] and, thus, it should not be used for computing electronic couplings, it is still useful to uncover and interpret trends or differences in computed couplings or to quickly prescreen inter-pigment interactions.

Contrary to multipole expansions, distributed expansions place a charge on each atom of the chromophore in order to reproduce the transition density and calculate the Coulomb coupling as an interaction between collections of point charges. In the TrEsp method,[86] these atomic transition charges are determined by fitting the electrostatic potential generated by $\rho_{i/j}^{\mathrm{tr}}$. This method has proven very effective in approximating the true Coulomb coupling almost exactly.

Importantly, the electronic couplings can also be affected by the surrounding environment. The environment influences the QM density of the chromophore, and therefore, it also affects the transition properties. This results in an indirect effect on the couplings. This effect can be described with any QM/MM embedding (electrostatic or polarizable). If the couplings are calculated in the context of subsystem TDDFT,[82,85] the effect of the surrounding environment is also included.

A second effect arises from the polarizability of the environment. The environment polarizes in response to the electric field generated by an electronic transition on chromophore j and, in turn, distorts the electric potential experienced by the other chromophore. This effect results in a completely new contribution to the Coulomb coupling. In the IPD formulation of polarizable QM/MM models, this contribution can be calculated as the interaction between the transition density ρ_i^{tr} and all the dipoles $\boldsymbol{\mu}_k^{\mathrm{MMPol}}$ induced on the environment atoms by the other transition density,[62]

$$V_{ij}^{\mathrm{MMPol}} = -\sum_k \int \mathrm{d}\mathbf{r}\rho_i^{\mathrm{tr}}(\vec{r})\frac{(\mathbf{r}-\mathbf{r}_k)}{|\mathbf{r}-\mathbf{r}_k|^3}\cdot\boldsymbol{\mu}_k^{\mathrm{MMPol}}\left(\rho_j^{\mathrm{tr}}\right), \tag{6}$$

where $\boldsymbol{\mu}_k^{\mathrm{MMPol}}\left(\rho_j^{\mathrm{tr}}\right)$ are obtained by solving Eq. (3) with an electric field generated by ρ_j^{tr}. Due to the symmetric bilinear nature of the interaction, this contribution does not change by exchanging i and j.[67] This new explicit contribution from the environment is proportional to the magnitude of both transition densities; therefore, it increases with an increase in the Coulomb coupling. In addition, it often has opposite sign to the Coulomb contribution, thus resulting in a reduction of the total coupling with respect to the Coulomb one.

If one approximates the environment as a dielectric, the sum of V_{ij}^{MMPol} and V_{ij}^{Coul} can be thought as an interaction screened by the dielectric. This model has been extensively used in the literature, for example, coupled to the TrEsp method in the so-called Poisson-TrEsp strategy[87] and to the Polarizable Continuum Model (PCM).[88] However, in the highly inhomogeneous environment of LH complexes, the mediated interaction is not always well represented by a dielectric screening. In fact, a different positioning of the environment molecules around the chromophores can lead to substantially different screenings on the Coulomb coupling.[9,40,89]

© Author(s) 2022

As mentioned above, the Coulomb coupling is often a very good approximation of the total electronic coupling, at least for bright transitions. Things change when we consider optically forbidden transitions and chromophores in van der Waals contact with each other. For optically forbidden transitions, the first multipole term of the Coulomb coupling is zero. This does not guarantee that the Coulomb coupling is negligible, but it usually means that this term is quite small. When two chromophores are in contact, the short-range coupling term is nonzero. Combining the two cases, it is possible that the short-range coupling is of the same magnitude as the Coulomb coupling, if not larger.

Unfortunately, short-range contributions to the coupling are not as easy to compute as the Coulomb term.[90] Some short-range terms arise from the fact that, at close contact, the electron clouds of the two chromophores experience Pauli repulsion, i.e., distort to satisfy the exclusion principle. As a consequence, perturbation expressions based on the isolated chromophores do not yield accurate results for this contribution.[83]

A more complete method to account for all contributions in electronic couplings is to use a QM calculation of the chromophoric dimer.[40,90,91] In this case, the electronic Hamiltonian has to be transformed to the monomeric "site" basis with a localization procedure. Given that the dimer of chromophores is considered in the QM calculation, this approach gives, in principle, all contributions to the coupling. However, this strategy is clearly more computationally expensive. On the other hand, this strategy allows for deriving both exciton and charge-transfer couplings from the same calculations[40,92] and, thus, describe all types of short-range interactions among chromophores.

IV. SIMULATION OF OPTICAL SPECTRA

The exciton model described in Sec. III is the basis for the simulation of the optical spectra of LH complexes. For example, the same excitonic coefficients defining the exciton states in Eq. (2) can be used to reconstruct the excitonic transition dipoles,

$$\mu_K = \sum_i c_i^K \mu_i, \tag{7}$$

which, together with the excitonic energies, determine the intensity and position of the expected spectroscopic signals. For excitonic systems, however, vertical excitations and transition dipoles are not enough to properly reproduce the spectra. In fact, as for single molecules, vibronic and broadening effects have to be accounted for.

From now on, for the sake of simplicity, we focus on absorption spectra, but similar considerations can be given for other optical spectroscopies and, at least in part, for two-dimensional electronic spectroscopy.[20]

Whereas the excitation energies and dipoles describe the position and intensity of the spectra, the band shape is determined by coupling of the electronic transitions to the vibrational degrees of freedom. The most common approach to include effects of the vibrational motion is the second order cumulant expansion in the exciton basis.[93] The cumulant expansion formalism is exact for Gaussian fluctuations of the excitation energy, i.e., for a linear coupling of the excitation energy to the vibrations of the system.[94]

The most commonly used flavor of cumulant expansion yields the absorption spectrum as a sum over exciton states,

$$\alpha(\omega) \propto \omega \sum_K |\boldsymbol{\mu}_K|^2 \int_{-\infty}^{\infty} e^{-i(\omega_K-\omega)t - g_K(t) - t/\tau_K}\, \mathrm{d}t, \tag{8}$$

where each exciton K is characterized by its transition dipole μ_K, excitation energy ω_K, and lifetime τ_K, where the latter depends on the rates of exciton relaxation from state K to the other states. The information on vibronic coupling is contained in the line shape function $g_K(t)$ for each state K. Similarly to the excitonic transition dipoles [Eq. (7)], $g_K(t)$ is reconstructed from the quantities of the individual pigments and their exciton coefficients, namely,

$$g_K(t) = \sum_i |c_i^K|^4 g_i(t), \tag{9}$$

where $g_i(t)$ is the line shape function of the pigment i.

The line shape function can be obtained from the spectral density $J(\omega)$ as

$$g(t) = -\int_0^{\infty} \mathrm{d}\omega \frac{J(\omega)}{\pi\omega^2}\left[\coth\left(\frac{\beta\hbar\omega}{2}\right)(\cos(\omega t) - 1) - i(\sin(\omega t) - \omega t)\right], \tag{10}$$

where $\beta = 1/k_b T$ introduces the temperature dependence of the optical spectra. The intensity of the spectral density $J(\omega)$ describes how much each vibrational mode at frequency ω influences the excitation energy. Different approaches have been proposed in the literature for the calculation of spectral densities: the most used are described in Sec. IV A.

The formalism described above is only one of the possible choices when using the cumulant expansion in an exciton system. In fact, it contains two key approximations. First, the excited-state dynamics is assumed to be Markovian and exciton relaxation is described by an exponential decay function with lifetime τ_K, which is usually derived with the Redfield rate equation. Various approaches at overcoming this approximation have been devised, either by correcting the Redfield rates[95] or by directly considering time-dependent rates in the cumulant expansion.[96] The second approximation is the secular approximation, which neglects off-diagonal elements in the cumulant expansion and allows for writing the spectra as a sum over exciton states [Eq. (8)]. The most general approach, the full second-order cumulant expansion, relaxes both of these approximations.[97,98]

So far, we have introduced only the effects due to the "dynamic" component of excitation energy fluctuations, which defines what is also called "dynamic disorder" and contributes to the homogeneous line shape. However, an additional effect, commonly denoted as "static disorder," should be considered as well. Here, the motions responsible for the disorder are those that lead to energy fluctuations slower than the absorption process, such as conformational changes of the environment and/or the pigments. Static disorder is usually included by introducing a set of exciton Hamiltonians calculated using site energies and couplings obtained for different configurations of the system (e.g., uncorrelated snapshots extracted from the MD simulation).[72,99–101] The set of eigenstates resulting from this ensemble of excitonic Hamiltonians is then used to

© Author(s) 2022

generate the corresponding set of spectra through Eq. (8). These spectra are finally convoluted together to obtain the final ensemble-averaged spectrum.

When applying such a strategy, however, one has to pay particular attention to how the different excitonic parameters respond to changes in the pigment geometry. The electronic couplings are quite stable with respect to changes in the internal geometry of the pigments depending more on their relative position and orientation, particularly when bright transitions are involved. Site energies, on the other hand, are known to vary significantly upon modest changes of the molecular geometry and even more when the pigment presents an extended conjugated chain as in the case of (bacterio)chlorophylls and carotenoids.[25] If a MM-MD trajectory is used for considering static disorder, site energy changes following alterations of the internal geometry of the pigments can be inflated if an inaccurate force field for the pigments is employed. Furthermore, even if an accurate force field is employed, the fluctuations of the site energies with the fast (e.g., molecular vibrations) degrees of freedom must be removed from the ensemble, to avoid double counting of this contribution when computing optical spectra. A viable route for introducing the static disorder in the simulation of the spectra is presented in Sec. IV B.

A. Spectral density

The spectral density describes how the nuclear degrees of freedom couple to a specific electronic transition. In the spirit of the second-order cumulant expansion, the spectral density is defined from the time-dependent fluctuations of the site excitation energy,

$$J(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} \Im \langle \Delta\mathscr{E}(t)\Delta\mathscr{E}(0)\rangle_g \mathrm{d}t, \tag{11}$$

where $\Delta\mathscr{E}(t) = \mathscr{E}(t) - \langle\mathscr{E}\rangle$ is the fluctuation of the site energy and $\langle\Delta\mathscr{E}(t)\Delta\mathscr{E}(0)\rangle_g$ denotes its autocorrelation function in the ground-state nuclear ensemble of the pigment.

The real nuclear quantum ensemble is generally inaccessible to practical simulations, but further simplifications can be made. In the case of linear coupling to a harmonic vibrational bath, the spectral density takes the simple form,

$$J(\omega) = \pi \sum_k S_k \omega_k^2 [\delta(\omega - \omega_k) - \delta(\omega + \omega_k)], \tag{12}$$

where ω_k are the frequencies of the normal modes and S_k are the corresponding Huang–Rhys factors, which quantify the dimensionless displacement of the excited-state potential energy surface with respect to that of the ground state. From the spectral density, one can obtain the reorganization energy as follows:

$$\lambda_i = \int_0^{\infty} d\omega \, \frac{J_i(\omega)}{\pi\omega}, \tag{13}$$

which measures the energy change of the pigment in its excited state when it moves from the ground-state equilibrium geometry to that of the excited state.

In general, the spectral density of an embedded pigment can be separated into two contributions, one due to the intra-pigment vibrational modes $J_{vib}(\omega)$ and one arising from the motions of the environment $J_{env}(\omega)$.

The environmental part of the spectral density is commonly represented as a continuous function with low frequency contributions. This function is obtained either from the fitting of the experimental spectra, i.e., the difference between absorption and fluorescence spectra, or from a normal mode analysis of the whole pigment protein complex.[102] The intra-pigment part of the spectral density is composed of discrete peaks at higher frequencies representing normal modes localized on the pigment. This part gives rise to what is normally called vibronic coupling. High-resolution spectroscopy, such as low-temperature fluorescence line narrowing (FLN) spectra,[103–105] can resolve individual vibronic peaks and give information on their frequencies and corresponding Huang–Rhys factors. Therefore, the $J_{vib}(\omega)$ contribution has been often extracted from experiments.

Normal-mode frequencies can be also computed from standard normal-mode analysis, and Huang–Rhys factors can be computed in the vertical gradient (VG) approximation.[51,106–109] In this approximation, the Huang–Rhys factors are computed from the gradient of the excited PES at the ground-state equilibrium geometry, and the spectral density is constructed as

$$J_{vib}(\omega) = \sum_\xi \frac{1}{\hbar\omega_\xi} \left(\frac{\partial H_{pigment}}{\partial q_\xi}\bigg|_{q_\xi=0} \right)^2 \frac{\omega_\xi^2 \omega \gamma_\xi}{\left(\omega^2 - \omega_\xi^2\right)^2 + \omega^2\gamma_\xi^2}, \tag{14}$$

where q_ξ are pigment normal mode dimensionless coordinates. The damping factors γ_ξ represent the finite correlation time of the normal mode.

Alternatively, the spectral density can be obtained through MD trajectories. Along such trajectories, the excitation energy is then computed at every time step, obtaining the so-called classical autocorrelation function $C_{cl}(t)$,

$$J(\omega) = \beta\omega \int_0^{\infty} dt \; C_{cl}(t) \cos(\omega t), \tag{15}$$

where a quantum correction $\beta\omega$ is employed to satisfy detailed balance,[110] and the classical autocorrelation function is computed as

$$C_{cl}(t_n) = \frac{1}{N-n} \sum_{k=1}^{N-n} \Delta\mathscr{E}(t_n + t_k)\Delta\mathscr{E}(t_k). \tag{16}$$

At variance with the vertical gradient, MD simulations do not rely on the harmonic approximation and, therefore, effectively incorporate some degree of anharmonicity.

In earlier applications, the trajectory of the system was obtained using MM-based MD. The SDs computed in this way, however, were too sensitive to the quality of the underlying force field, which has to match the QM description of the chromophore to give correct results.[31,112–116] More recently, the increase in computer power and algorithms has allowed for the use of Born–Oppenheimer QM/MM(Pol) MDs to generate the trajectory.[101,117,118] Commonly, a DFT description is used for the ground state QM/MM-MD simulations, which is then combined with a TD-DFT modeling of the excitation energies. Recently, semiempirical methods, such as DFTB, have been proposed as a computationally effective alternative to compute spectral densities of pigments in LH complexes.[50,74,119]

© Author(s) 2022

To give an example of how different models for computing the spectral density influence the corresponding line shape, in Figs. 3(a) and 3(b), we compare the experimentally derived spectral density of CP29 with those obtained from vertical gradient calculations or QM/MM MD simulations. The VG calculations were performed for Chls a604 and a610 of CP29, while the QM/MM MD spectral density refers to Chl a604.[74] As one can note, comparing the two VG calculations, it is difficult to distinguish Chl a604 from a610. Moreover, both models agree quite well with the experimental estimate, with the QM/MM MD estimates showing a slightly better fit of the peak positions. This suggests that the inclusion of anharmonicity through QM/MM MD improves the description of intermolecular vibrations.

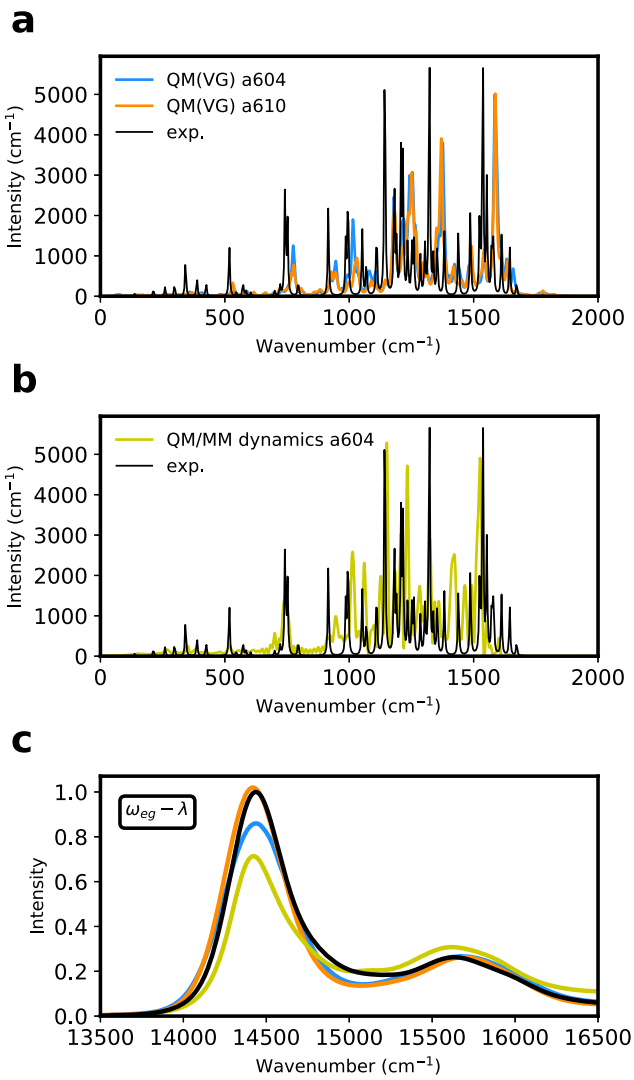

FIG. 3. Spectral densities and spectra of chlorophyll. A comparison of the spectral density of a single Chl *a* obtained from the fitting of FLN experiments (Expt.)[111] with the SD computed within the vertical gradient (VG) approach (a) and from a QM/MM dynamics (b).[74] (c) Optical spectra for a single Chl *a* obtained using different intra-pigment parts of the spectral densities, as illustrated in (a) and (b). The excitation energy is shifted by the corresponding reorganization energy λ. The same environmental contribution J_{env} was added to the intra-pigment part of the spectral density to obtain the complete spectral density of the chlorophyll embedded in the LH complex. This term is modeled as an overdamped Brownian oscillator with the reorganization energy equal to 60 cm^{-1} and a correlation time of 177 fs. No environmental contribution was added to the spectral density obtained from a QM/MM dynamics because in this case, the environmental contribution is inherently included.

The different SD models have also been used to simulate the absorption line shape. The spectra are computed using the same adiabatic excitation energy $\omega_{eg}^{0} = \omega_{eg} - \lambda$ [Fig. 3(c)]. From this comparison, it is clear that the differences between the experimentally derived spectral density (SD exp) and the computed ones have a limited impact on the actual line shape. In fact, despite the apparent differences in Figs. 3(a) and 3(b), the line shapes in Fig. 3(d) are all very similar and differ mainly in the relative intensity of the main 0–0 peak and the vibronic band.

Strikingly, among the vertical gradient calculations, the one on Chl a610 more closely resembles the line shape obtained from the experimental spectral density. This suggests that the low-temperature FLN experiment extracts the vibrational frequencies of Chl a610, which belongs to the terminal emitter state. The line shape calculated for Chl a604 deviates from the experimentally derived one and even more so does the one computed from QM/MM MD, which takes into account temperature effects. This comparison suggests that the experimentally derived spectral density is biased toward the low-energy Chls, and an accurate representation of vibronic coupling requires QM/MM calculations.

On the other hand, the line shapes obtained with the different methods are sufficiently similar to each other. For the purpose of computing exciton spectra, the choice of spectral density is probably not critical.

B. Static disorder

As reported above, MD is an efficient route to sample the configurational space of the solvated LH complex. The resulting trajectories can be finally used to extract different configurations of the system on which different excitonic Hamiltonians are calculated to account for static disorder in the simulation of the spectra.

We consider here the case of an LH complex containing several identical chlorophylls, and we assume to have performed different MD replicas of the LH complex embedded in its environment. The value of the site energy calculated at a given configuration of the whole system (LH complex + environment) extracted from an MD trajectory can be interpreted as a sum of different contributions, namely,

$$\mathscr{E}_{i,r,f}^{\mathrm{MD}} = \mathscr{E}_{i,r,f}^{\mathrm{geom,\,fast}} + \mathscr{E}_{i,r,f}^{\mathrm{geom,\,slow}} + \mathscr{E}_{i,r,f}^{\mathrm{env}}, \tag{17}$$

where i, r, and f label the pigment, the index of the MD replica, and the frame of the replica, respectively. Figure 4(a) shows the high variability of $\mathscr{E}^{MD}$, stemming from the dynamics of both the pigment and environment degrees of freedom.

Here, $\mathscr{E}^{\mathrm{geom,fast}}$ indicates the sum of the site energy calculated on the isolated pigment in its equilibrium geometry and a correction

© Author(s) 2022

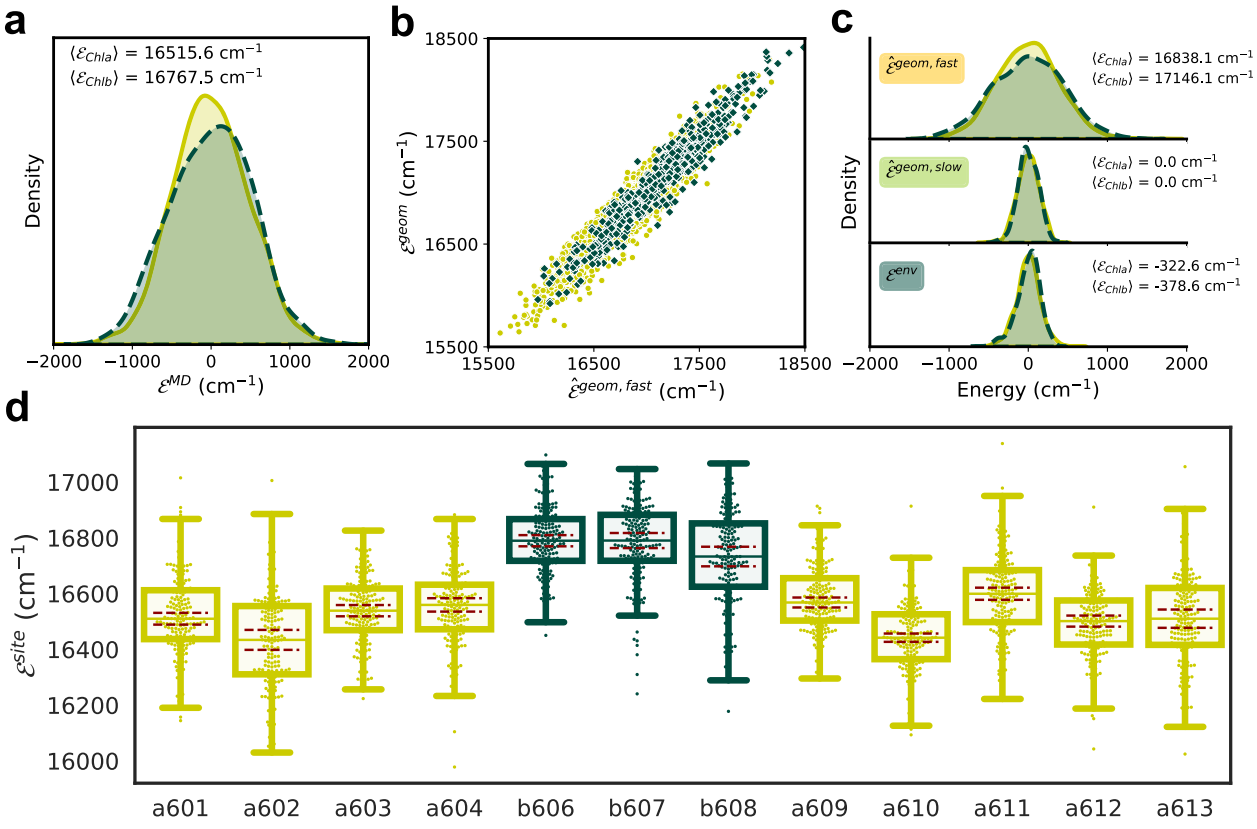

FIG. 4. Chlorophyll site energies in CP29. (a) Distribution of the Chl site energies obtained with a MMPol TD-DFT excited state calculation on geometries extracted from the classical MD ($\mathscr{E}^{MD}$). (b) Estimated fast contribution of the geometric (vacuum) Chl site energy ($\hat{\mathscr{E}}^{\text{geom, fast}}$) vs the geometric contribution ($\mathscr{E}^{\text{geom}}$). (c) Site energy decomposition into fast $\hat{\mathscr{E}}^{\text{geom,fast}}$ (top), geometrically slow $\hat{\mathscr{E}}^{\text{geom,slow}}$ (middle), and environment $\mathscr{E}^{\text{env}}$ (bottom) contributions. The estimated probability density of each contribution is shown. (d) Chlorophyll site energies $\mathscr{E}^{\text{site}}$, Eq. (19), in CP29. Chl *a* are shown in light green, and Chl *b* are shown in dark green. The boxplots show the median (central line), the first and third interquartile (extension of the box), and 1.5 × IQR (extension of the whiskers). The underlying points are shown for clarity. The red dashed horizontal lines indicate the error associated with the median of the distribution. The error is computed as $1.96 \times Std_B[\mathscr{E}^{\text{site}}]$, where $Std_B[\cdot]$ is the standard deviation computed from 1000 bootstrap samples.

stemming from the variation of its fast nuclear degrees of freedom, such as the bond lengths, at that specific configuration of the system. In parallel, $\mathscr{E}^{\text{geom,slow}}$ is the further change of the site energy due to change of the slower nuclear degrees of freedom of the pigment, such as angle distortions due to protein environment constraints. The total geometric contribution $\mathscr{E}^{\text{geom}} = \mathscr{E}^{\text{geom,fast}} + \mathscr{E}^{\text{geom,slow}}$ is directly accessible from a vacuum QM calculation of the pigment using a structure sampled from the MD simulation of the LH complex. Finally, $\mathscr{E}^{\text{env}}$ is the explicit environment contribution, obtained by subtracting the total geometric component from a QM/MM(pol) calculation as $\mathscr{E}^{\text{env}}_{i,r,f} = \mathscr{E}^{\text{QM/MM(Pol)}}_{i,r,f} - \mathscr{E}^{\text{geom}}_{i,r,f}$.

A straightforward separation of the fast and slow components of the geometrical contribution is not available. Still, one can exploit the remarkable linear dependence of the Chl site energy from the bond lengths of the conjugated ring[25] to obtain an estimate of the fast component $\mathscr{E}^{\text{geom,fast}}$. After featurizing the chlorophyll geometry as a set of bond lengths, the estimate of the fast geometrical component of the site energy $\hat{\mathscr{E}}^{\text{geom,fast}}_{i,r,f} = \mathbf{x}_{i,r,f}\mathbf{w}$ can be obtained by minimizing the squared loss within a regularized linear model,

$$\arg\min_{\mathbf{w}} \frac{1}{2}\sum_{i,r,f}\left(\mathscr{E}^{\text{geom}}_{i,r,f} - \mathbf{x}_{i,r,f}\mathbf{w}\right)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2, \tag{18}$$

where $\mathbf{x}_{i,r,f}$ denotes the set of chlorophyll bond lengths (padded with 1 to include the intercept in the fitting), $\mathbf{w}$ is the coefficient vector of the linear model, and λ is a regularization parameter that prevents overfitting. The determination of $\hat{\mathscr{E}}^{\text{geom,fast}}$ through the regularized fitting is shown in Fig. 4(b).

The slow component of the geometrical contribution is then obtained by subtracting the fast component as $\hat{\mathscr{E}}^{\text{geom,slow}}_{i,r,f} = \mathscr{E}^{\text{geom}}_{i,r,f} - \hat{\mathscr{E}}^{\text{geom,fast}}_{i,r,f}$. Figure 4(c) shows the distribution of $\hat{\mathscr{E}}^{\text{geom,fast}}$, $\hat{\mathscr{E}}^{\text{geom,slow}}$, and $\mathscr{E}^{\text{env}}$. The fast geometrical contribution evidently

© Author(s) 2022

accounts for the most variance in $\mathscr{E}^{\mathrm{MD}}$, with the slow geometrical contribution and the environment contribution presenting a comparable variability.

Once all the contributions have been computed for each pigment, replica and configuration, we define the correct site energy to be used in the evaluation of static disorder as

$$\mathscr{E}^{\mathrm{site}}_{i,r,f} = \left\langle \hat{\mathscr{E}}^{\mathrm{geom,fast}}_{i,r,f} \right\rangle_{i,r,f} + \left\langle \hat{\mathscr{E}}^{\mathrm{geom,slow}}_{i,r,f} \right\rangle_{f} + \mathscr{E}^{\mathrm{env}}_{i,r,f}, \tag{19}$$

where $\langle\cdot\rangle_x$ denotes the average with respect to x.

The site energies computed with Eq. (19) are shown in Fig. 4(d). The distribution of each site energy is represented with a box plot, and we also report the confidence interval obtained via bootstrapping (red dashed lines). The statistical uncertainty on the mean site energy is around 50 cm^{-1} for most Chls and ensures that the obtained site-energy ladder is statistically significant.

In Eq. (19), the fast geometrical contribution has been averaged on all the chlorophylls and all the configurations of the different replicas. For the slow contribution instead, the average is performed on all the configurations within each replica separately but not on the pigments, as we have assumed that different chlorophylls will have different distortions due to their specific environment constraints.

With this scheme, we (i) correct the intrinsic error of the MM-MD simulation for the pigment bond lengths, (ii) exclude double-counting of bond length contribution to the energy disorder, and (iii) reduce the "noise" in the excitation energies coming only from the limited sampling of pigment geometries. We also point out that other approaches avoiding double-counting of vibrational motions in the final spectra, relying on geometry optimizations of the pigments, are present in the literature.[120,121]

Another approach to separating static and dynamic disorder in MD simulations is based on the calculation of site energies for a large number of closely spaced frames.[122] If the site energy is mediated over a time window T, the fast fluctuations are reduced by a factor $\sqrt{T}$, while slower fluctuations are unaffected. The standard deviation of the means can be extrapolated for large T, yielding the magnitude of static disorder. This approach requires many calculations along a time window of several ns, and it is thus not feasible with standard QM/MM calculations. Using interpolation schemes for the potential energy surfaces[122] makes this strategy viable. A further strategy involves running separate simulations, both for a single pigment and for an ensemble of pigments. Simulations on single pigments are employed to estimate the fast fluctuations of the site energies. Simulations on the ensemble are then used to extract the static disorder.[123]

V. APPLICATION TO CP29 AND ITS MUTANT

In Secs. III A, IV A, and IV B, we have presented some results obtained for the excitonic parameters of CP29. Here, we extend the analysis by comparing CP29 with a mutant and give a rationale for the observed changes in the absorption spectra of the two complexes.

The mutant that we are considering has been investigated experimentally by Guardini *et al.*[22] It presents an asparagine (Asn) residue replacing the histidine (His) as the axial ligand of Chl *a*603 and, in the following, will be denoted as CP29-H111N for simplicity (we adopt here the same numbering of the referred article). This mutant is particularly interesting as it retains the targeted Chl *a*603, contrary to other mutations at the same H111 site, and it induces an increase in the red edge of the absorption spectrum, which has been connected to a change in the quenching properties of the complex. Authors suggested that this mutation results in a reduced distance/altered orientation between chlorophylls *a*603 and *a*609, induced by the lower steric hindrance of Asn as compared to His, and increased excitonic interactions between the two.[22]

A. Molecular dynamics of CP29-WT and CP29-H111N

MD simulations for both CP29-WT and CP29-H111N were employed for investigating the rearrangements occurring at the level of the Vio-*a*603-*a*609 cluster. The simulation protocol employed for the CP29-WT simulations has been presented elsewhere.[32] For the CP29-H111N simulations, we have employed the very same protocol, with the only exception that the His residue H111 was mutated into the N111 residue before starting the minimization.

The reduced steric hindrance of Asn as compared to His results in Chl *a*603 being closer to its axial ligand in CP29-H111N as compared to CP29-WT. In Fig. 5(a), we show the distance between the Mg atom of Chl *a*603 and the Cα carbon of its axial ligand. The Mg-Cα distance is reduced by ~14% in CP29-H111N as compared to CP29-WT (the mean distance is 6.64 Å in CP29-WT and 5.74 Å in CP29-H111N), confirming the rationalizations of Guardini *et al.*[22] The *a*603 position appears to be very robust to local structural changes induced by the protein dynamics in CP29-WT, with a standard deviation of ~0.13 Å, and less so in CP29-H111N (standard deviation of ~0.28 Å; see also Fig. 5). Thus, the closer Chl *a*603 in CP29-H111N is more sensitive to local protein fluctuations than in the WT.

The same simulations have also been employed to investigate the relative position and orientation of the *a*603-*a*609 pair. When looking at the minimum distance between the two chlorophylls, we see that it is shorter in the mutant than in the WT (the average minimum distance is 3.54 and 3.38 Å in CP29-WT and CP29-H111N, respectively, which translates into a 4% decrease in the mutant compared to the WT). Differences in the *a*603-*a*609 cluster between CP29-WT and CP29-H111N, however, are not limited to their position, but also their relative orientation is affected by the altered position of *a*603 resulting from the mutation. Figure 5(c) shows the squared orientational factor κ_{ij} (i = *a*603, j = *a*609) appearing in Eq. (5), which we analyze here as a qualitative indicator of increased/decreased electronic coupling between the two pigments. The plot in Fig. 5 shows a 22% increase of κ^2_{ij} in CP29-H111N as compared to CP29-WT (the average value of κ^2_{ij} is 1.76 and 2.14 for CP29-WT and CP29-H111N, respectively) and a similar sensitivity to the protein dynamics (standard deviation of 0.25 and 0.22 for CP29-WT and CP29-H111N, respectively). The altered position and orientation of this chlorophyll pair can also be observed when looking at the MD [see Figs. 5(b) and 5(d) for a particular snapshot of the WT and the mutant].

© Author(s) 2022

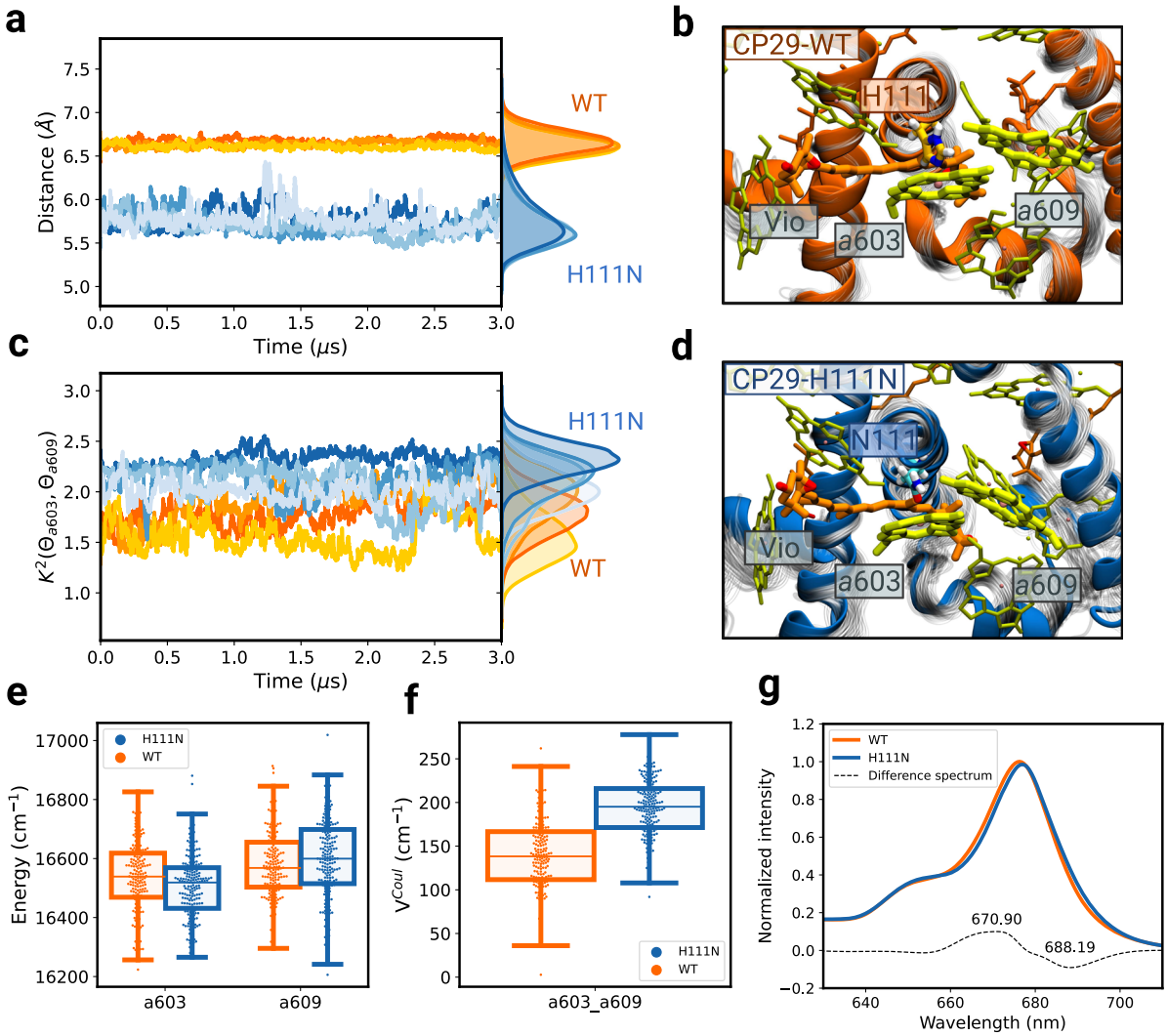

FIG. 5. Differences between CP29-WT and CP29-H111N. (a) Distance between the Mg atom of Chl *a*603 and the Cα carbon of its axial ligand. The axial ligand is H111 for CP29-WT and N111 for CP29-H111N. The estimated probability distribution of the distance is shown in the inset on the right. Different shades of orange (CP29-WT) and blue (CP29-H111N) correspond to different replicas. (b) Arrangement of the L2 pocket in CP29-WT. The histidine residue (H111) coordinating Chl *a*603 is highlighted. (c) Förster orientation factor squared between Chl *a*603 and *a*609 in CP29-WT and CP29-H111N. The estimated probability distribution of the orientation factor is shown in the inset on the right. Different shades of orange (CP29-WT) and blue (CP29-H111N) correspond to different replicas. (d) Arrangement of the L2 pocket in CP29-H111N. The asparagine (N111) residue coordinating Chl *a*603 is highlighted. (e) Site energies [as computed with Eq. (19)] for Chl *a*603 and Chl *a*609 in both CP29-WT and CP29-H111N. (f) Coulomb coupling [as computed with Eqs. (4) and (6)] between the Chl pair *a*603-*a*609 for both CP29-WT and CP29-H111N. (g) Absorption spectrum of CP29-WT and CP29-H111N. The spectrum of CP29-WT is normalized, and the one of CP29-H111N rescaled by the same normalization factor. The dashed line shows the difference spectrum (WT − H111N). The maximum and minimum of the difference spectrum are shown. In all plots, orange lines correspond to CP29-WT and blue lines correspond to CP29-H111N. The boxplots show the median (central line), the first and third interquartile (extension of the box), and 1.5 × IQR (extension of the whiskers). The underlying points are shown for clarity.

B. Site energies, couplings, and optical spectrum

The geometrical analysis of the trajectories strongly suggests that the electronic coupling between the *a*603-*a*609 chlorophyll pair will be enhanced in the mutant with respect to WT. Meanwhile, the altered positions of these chlorophylls in the mutant may lead to substantially modified chlorophyll site energies. Assuming for simplicity that the two chlorophylls only make up the lowest or second-lowest exciton in CP29 (which is a reasonable assumption according to previous quantum chemical calculations and spectroscopic results[77,124]), both their site energies and their coupling are important for determining the resulting excitonic state. Site energies and couplings have been computed at the QM/MMpol level

© Author(s) 2022

using TD-DFT M062X/6-31G(d) for both the CP29-WT and CP29-H111N, selecting uniformly spaced structures from the last 1 μs of trajectory for each replica, resulting in a total of 179 and 200 calculations for both chlorophylls for CP29-WT and CP29-H111N, respectively. The phytil tail of the chlorophyll was removed from the QM part (and considered in the polarizable MM system) by cutting the bond after the first aliphatic carbon of the tail. Additionally, in order to recover the environment contribution to the site energy $\mathscr{E}^{\mathrm{env}}$ and extract the geometrical site energy $\mathscr{E}^{\mathrm{geom}}$, we performed the same QM calculations without including the effects of the MM subsystem.

The site energy $\mathscr{E}^{\mathrm{site}}$ of each chlorophyll was then obtained by means of Eq. (19), which is essential to avoid double counting of the fast internal motions of the pigments when computing the spectrum, while still retaining the pigment identity arising from its particular dynamic environment. The site energies of Chls *a*603 and *a*609 are shown in Fig. 5(e). The difference in mean of chlorophylls *a*603 and *a*609 between the WT and the mutant is $\langle \mathscr{E}_{a603} \rangle_{\mathrm{WT}} - \langle \mathscr{E}_{a603} \rangle_{\mathrm{H111N}} = 30 \pm 41\ \mathrm{cm}^{-1}$ and $\langle \mathscr{E}_{a609} \rangle_{\mathrm{WT}} - \langle \mathscr{E}_{a609} \rangle_{\mathrm{H111N}} = -18 \pm 43\ \mathrm{cm}^{-1}$ (the mean difference value and the associated standard error pooled among the replicas are reported). These values indicate that we can consider the site energy of both chlorophylls to be practically the same in the mutant and the WT. By contrast, the electronic coupling difference for the *a*603-*a*609 pair is estimated as $\langle V_{a603-a609} \rangle_{\mathrm{WT}} - \langle V_{a603-a609} \rangle_{\mathrm{H111N}} = -55 \pm 13\ \mathrm{cm}^{-1}$, which is around a 40% change in the absolute coupling of the WT [see also Fig. 5(f)]. Therefore, corroborating the geometrical insights obtained from the orientational factor [Fig. 5(c)], we obtain a significantly different coupling between the WT and the mutant, with the mutant presenting higher excitonic interactions than the WT. This confirms the intuition of Guardini *et al.*,[22] who proposed a strengthening of excitonic interactions for the *a*603-*a*609 pair in CP29-H111N.

Considering the local character of the geometrical changes associated with the mutation (i.e., only the relative arrangement of *a*603 and *a*609 is substantially altered) and the increased electronic coupling between the two chlorophylls, we can compute the absorption spectrum of CP29-H111N by employing the very same exciton Hamiltonians employed for CP29-WT, with the only difference that the *a*603-*a*609 coupling is increased by 55 cm^{-1} in the mutant. The absorption spectra of CP29-WT and CP29-H111N are then computed with the cumulant expansion theory truncated at second order (see Sec. IV). The spectral density of chlorophylls was modeled as a linear combination of 48 high-frequency modes and an overdamped Brownian oscillator term, taking the parameters from FLN experimental data.[111] This choice was made to remain consistent with our earlier optical spectra simulations of light-harvesting complex II (LHCII) and light-harvesting complex stress-related protein 1 (LHCSR1).[26,72]

The calculated absorption spectra for CP29-WT and CP29-H111N are reported in Fig. 5(g). Notably, the red shift observed by Guardini *et al.* (from 674.8 to 685.6 nm) is well reproduced by our calculations (from 670.9 to 688.2 nm), demonstrating the effectiveness of our approach. Therefore, our computational model of CP29-H111N validates the interpretation of Guardini *et al.*,[22] namely, that a strengthening of excitonic interactions of the a603-a609 pair is responsible for the red shift observed in the absorption spectrum. We also note that this computational approach does not just reproduce the experimental results, but it complements them by explicitly tracking down the molecular sources, giving rise to the different observables that have been probed experimentally.

VI. CONCLUDING REMARKS AND FUTURE OUTLOOKS

In this Perspective, we have presented a general overview of the challenges encountered when light-harvesting complexes are modeled through atomistic approaches and highlighted the peculiarities that characterize these pigment–protein complexes. We have presented an effective computational strategy to simulate their optical spectroscopy, which combines molecular dynamics simulation, quantum chemical methodologies, and line shape theories. The predictive power achieved by such an integrated strategy has been demonstrated on the comparison of absorption spectra of wild-type CP29 and a particular CP29 mutant that has been studied recently by experimental means.

Despite the remarkable power offered by this strategy, there is still much room for improvement in many of the computational aspects discussed in this Perspective.

One possibility is to exploit machine learning methods to substantially reduce the computational cost associated with excitonic properties while retaining an accuracy almost equal to that of expensive quantum chemical methods.[125–132] Machine learning frameworks have also proven to be able to incorporate the description of the environment for complex systems, including pigment–protein complexes.[133,134] This opens a route to accurate calculations of static and dynamics disorders, which require a large number of calculations.

Another important aspect is the extent of MD sampling. In most applications, crystal structures or short MD simulations of the complexes are used, which can lead to significant biases toward the selected initial structure. Instead, to properly explore the whole conformational space of the complex, including all possible configurations of the pigments, enhanced sampling simulations should be used.[135–138] In addition, in this case, machine learning techniques can represent a speedup by providing useful reaction coordinates upon which the sampling is improved.[139–142]

ACKNOWLEDGMENTS

The authors acknowledge funding from the European Research Council, under Grant No. ERC-AdG-786714 (LIFETimeS). The authors thank Ulrich Kleinekathöfer for providing the spectral densities calculated through a DFTB/MM MD simulation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

© Author(s) 2022

REFERENCES

[1] R. E. Blankenship, *Molecular Mechanisms of Photosynthesis* (John Wiley & Sons, 2014).
[2] G. D. Scholes, G. R. Fleming, A. Olaya-Castro, and R. van Grondelle, "Lessons from nature about solar light harvesting," Nat. Chem. **3**, 763–774 (2011).
[3] R. Croce and H. van Amerongen, "Natural strategies for photosynthetic light harvesting," Nat. Chem. Biol. **10**, 492–501 (2014).
[4] L. Dall'Osto, M. Bressan, and R. Bassi, "Biogenesis of light harvesting proteins," Biochim. Biophys. Acta, Bioenerg. **1847**, 861–871 (2015).
[5] T. Renger, M. E.-A. Madjet, M. Schmidt am Busch, J. Adolphs, and F. Müh, "Structure-based modeling of energy transfer in photosynthesis," Photosynth. Res. **116**, 367–388 (2013).
[6] T. Renger and F. Müh, "Understanding photosynthetic light-harvesting: A bottom up theoretical approach," Phys. Chem. Chem. Phys. **15**, 3348 (2013).
[7] C. Curutchet and B. Mennucci, "Quantum chemical studies of light harvesting," Chem. Rev. **117**, 294–343 (2017).
[8] S. J. Jang and B. Mennucci, "Delocalized excitons in natural light-harvesting complexes," Rev. Mod. Phys. **90**, 035003 (2018).
[9] L. Cupellini, M. Bondanza, M. Nottoli, and B. Mennucci, "Successes & challenges in the atomistic modeling of light-harvesting and its photoregulation," Biochim. Biophys. Acta, Bioenerg. **1861**, 148049 (2020).
[10] T. L. C. Jansen, "Computational spectroscopy of complex systems," J. Chem. Phys. **155**, 170901 (2021).
[11] G. R. Fleming and R. van Grondelle, "Femtosecond spectroscopy of photosynthetic light-harvesting systems," Curr. Opin. Struct. Biol. **7**, 738–748 (1997).
[12] Y.-C. Cheng and G. R. Fleming, "Dynamics of light harvesting in photosynthesis," Annu. Rev. Phys. Chem. **60**, 241–262 (2009).
[13] X. Pan, Z. Liu, M. Li, and W. Chang, "Architecture and function of plant light-harvesting complexes II," Curr. Opin. Struct. Biol. **23**, 515–525 (2013).
[14] G. D. Scholes and C. Smyth, "Perspective: Detecting and measuring exciton delocalization in photosynthetic light harvesting," J. Chem. Phys. **140**, 110901 (2014).
[15] T. P. J. Krüger and R. van Grondelle, "Design principles of natural light-harvesting as revealed by single molecule spectroscopy," Physica B **480**, 7–13 (2016).
[16] T. Mirkovic, E. E. Ostroumov, J. M. Anna, R. van Grondelle, Govindjee, and G. D. Scholes, "Light absorption and energy transfer in the antenna complexes of photosynthetic organisms," Chem. Rev. **117**, 249–293 (2017).
[17] T. Kondo, W. J. Chen, and G. S. Schlau-Cohen, "Single-molecule fluorescence spectroscopy of photosynthetic systems," Chem. Rev. **117**, 860–898 (2017).
[18] C. König and J. Neugebauer, "Quantum chemical description of absorption properties and excited-state processes in photosynthetic systems," ChemPhysChem **13**, 386–425 (2011).
[19] S. Jurinovich, L. Viani, C. Curutchet, and B. Mennucci, "Limits and potentials of quantum chemical methods in modelling photosynthetic antennae," Phys. Chem. Chem. Phys. **17**, 30783–30792 (2015).
[20] F. Segatta, L. Cupellini, M. Garavelli, and B. Mennucci, "Quantum chemical modeling of the photoinduced activity of multichromophoric biosystems," Chem. Rev. **119**, 9361–9380 (2019).
[21] N. Liguori, R. Croce, S. J. Marrink, and S. Thallmair, "Molecular dynamics simulations in photosynthesis," Photosynth. Res. **144**, 273–295 (2020).
[22] Z. Guardini, M. Bressan, R. Caferri, R. Bassi, and L. Dall'Osto, "Identification of a pigment cluster catalysing fast photoprotective quenching response in CP29," Nat. Plants **6**, 303–313 (2020).
[23] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," Nucleic Acids Res. **28**, 235–242 (2000).
[24] N. Liguori, V. Novoderezhkin, L. M. Roy, R. van Grondelle, and R. Croce, "Excitation dynamics and structural implication of the stress-related complex LHCSR3 from the green alga *Chlamydomonas reinhardtii*," Biochim. Biophys. Acta, Bioenerg. **1857**, 1514–1523 (2016).
[25] F. Cardoso Ramos, M. Nottoli, L. Cupellini, and B. Mennucci, "The molecular mechanisms of light adaption in light-harvesting complexes of purple bacteria revealed by a multiscale modeling," Chem. Sci. **10**, 9650–9662 (2019).
[26] I. G. Prandi, V. Sláma, C. Pecorilla, L. Cupellini, and B. Mennucci, "Structure of the stress-related LHCSR1 complex determined by an integrated computational strategy," Commun. Biol. **5**, 145 (2022).
[27] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with AlphaFold," Nature **596**, 583–589 (2021).
[28] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer *et al.*, "Accurate prediction of protein structures and interactions using a three-track neural network," Science **373**, 871–876 (2021).
[29] M. Ceccarelli, P. Procacci, and M. Marchi, "An *ab initio* force field for the cofactors of bacterial photosynthesis," J. Comput. Chem. **24**, 129–142 (2003).
[30] I. G. Prandi, L. Viani, O. Andreussi, and B. Mennucci, "Combining classical molecular dynamics and quantum mechanical methods for the description of electronic excitations: The case of carotenoids," J. Comput. Chem. **37**, 981–991 (2016).
[31] O. Andreussi, I. G. Prandi, M. Campetella, G. Prampolini, and B. Mennucci, "Classical force fields tailored for QM applications: Is it really a feasible strategy?," J. Chem. Theory Comput. **13**, 4636–4648 (2017).
[32] M. Lapillo, E. Cignoni, L. Cupellini, and B. Mennucci, "The energy transfer model of nonphotochemical quenching: Lessons from the minor CP29 antenna complex of plants," Biochim. Biophys. Acta, Bioenerg. **1861**, 148282 (2020).
[33] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, "Unsupervised learning methods for molecular simulation data," Chem. Rev. **121**, 9722–9758 (2021).
[34] K. Pearson, "On lines and planes of closest fit to systems of points in space," London, Edinburgh Dublin Philos. Mag. J. Sci. **2**, 559–572 (1901).
[35] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science **290**, 2319–2323 (2000).
[36] E. Cignoni, M. Lapillo, L. Cupellini, S. Acosta-Gutiérrez, F. L. Gervasio, and B. Mennucci, "A different perspective for nonphotochemical quenching in plant antenna complexes," Nat. Commun. **12**, 7152 (2021).
[37] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," Phys. Rev. Lett. **72**, 3634–3637 (1994).
[38] V. Daskalakis, S. Papadatos, and T. Stergiannakos, "The conformational phase space of the photoprotective switch in the major light harvesting complex II," Chem. Commun. **56**, 11215–11218 (2020).
[39] X. Li, R. M. Parrish, F. Liu, S. I. L. Kokkila Schumacher, and T. J. Martínez, "An ab initio exciton model including charge-transfer excited states," J. Chem. Theory Comput. **13**, 3493–3504 (2017).
[40] M. Nottoli, S. Jurinovich, L. Cupellini, A. T. Gardiner, R. Cogdell, and B. Mennucci, "The role of charge-transfer states in the spectral tuning of antenna complexes of purple bacteria," Photosynth. Res. **137**, 215–226 (2018).
[41] O. Andreussi, S. Knecht, C. M. Marian, J. Kongsted, and B. Mennucci, "Carotenoids and light-harvesting: From DFT/MRCI to the Tamm–Dancoff approximation," J. Chem. Theory Comput. **11**, 655–666 (2015).
[42] F. Segatta, L. Cupellini, S. Jurinovich, S. Mukamel, M. Dapor, S. Taioli, M. Garavelli, and B. Mennucci, "A quantum chemical interpretation of two-dimensional electronic spectroscopy of light-harvesting complexes," J. Am. Chem. Soc. **139**, 7558–7567 (2017).
[43] A. Anda, T. Hansen, and L. De Vico, "Q_y and Q_x absorption bands for bacteriochlorophyll a molecules from LH2 and LH3," J. Phys. Chem. A **123**, 5283–5292 (2019).
[44] V. V. Poddubnyy, M. I. Kozlov, and I. O. Glebov, "The origin of the red shift of Q_y band of chlorophylls *d* and *f*," Chem. Phys. Lett. **778**, 138792 (2021).
[45] C.-M. Suomivuori, H. Fliegl, E. B. Starikov, T. S. Balaban, V. R. I. Kaila, and D. Sundholm, "Absorption shifts of diastereotopically ligated chlorophyll dimers of photosystem I," Phys. Chem. Chem. Phys. **21**, 6851–6858 (2019).
[46] A. Sirohiwal, R. Berraud-Pache, F. Neese, R. Izsák, and D. A. Pantazis, "Accurate computation of the absorption spectrum of chlorophyll *a* with pair natural orbital coupled cluster methods," J. Phys. Chem. B **124**, 8761–8771 (2020).
[47] A. Sirohiwal, F. Neese, and D. A. Pantazis, "How can we predict accurate electrochromic shifts for biochromophores? A case study on the photosynthetic reaction center," J. Chem. Theory Comput. **17**, 1858–1873 (2021).

© Author(s) 2022

[48]Z. Hashemi and L. Leppert, "Assessment of the *ab initio* Bethe–Salpeter equation approach for the low-lying excitation energies of bacteriochlorophylls and chlorophylls," J. Phys. Chem. A **125**, 2163–2172 (2021).
[49]B. M. Bold, M. Sokolov, S. Maity, M. Wanko, P. M. Dohmen, J. J. Kranz, U. Kleinekathöfer, S. Höfener, and M. Elstner, "Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes," Phys. Chem. Chem. Phys. **22**, 10500–10518 (2020).
[50]S. Maity, B. M. Bold, J. D. Prajapati, M. Sokolov, T. Kubař, M. Elstner, and U. Kleinekathöfer, "DFTB/MM molecular dynamics simulations of the FMO light-harvesting complex," J. Phys. Chem. Lett. **11**, 8660–8667 (2020).
[51]M. K. Lee, K. B. Bravaya, and D. F. Coker, "First-principles models for biological light-harvesting: Phycobiliprotein complexes from cryptophyte algae," J. Am. Chem. Soc. **139**, 7803–7814 (2017).
[52]F. Lipparini and B. Mennucci, "Hybrid QM/classical models: Methodological advances and new applications," Chem. Phys. Rev. **2**, 041303 (2021).
[53]E. Boulanger and W. Thiel, "Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: A fully polarizable model," J. Chem. Theory Comput. **8**, 4527–4538 (2012).
[54]E. Boulanger and W. Thiel, "Toward QM/MM simulation of enzymatic reactions with the drude oscillator polarizable force field," J. Chem. Theory Comput. **10**, 1795–1809 (2014).
[55]M. A. Hagras and W. J. Glover, "Polarizable embedding for excited-state reactions: Dynamically weighted polarizable QM/MM," J. Chem. Theory Comput. **14**, 2137–2144 (2018).
[56]R. A. Bryce, R. Buesnel, I. H. Hillier, and N. A. Burton, "A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges," Chem. Phys. Lett. **279**, 367–371 (1997).
[57]F. Lipparini and V. Barone, "Polarizable force fields and polarizable continuum model: A fluctuating charges/PCM approach. 1. Theory and implementation," J. Chem. Theory Comput. **7**, 3711–3724 (2011).
[58]T. Giovannini, R. R. Riso, M. Ambrosetti, A. Puglisi, and C. Cappelli, "Electronic transitions for a fully polarizable QM/MM approach based on fluctuating charges and fluctuating dipoles: Linear and corrected linear response regimes," J. Chem. Phys. **151**, 174104 (2019).
[59]A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol. **103**, 227–249 (1976).
[60]M. A. Thompson and G. K. Schenter, "Excited states of the bacteriochlorophyll *b* dimer of *Rhodopseudomonas viridis*: A QM/MM study of the photosynthetic reaction center that includes MM polarization," J. Phys. Chem. **99**, 6374–6386 (1995).
[61]J. Gao, "Energy components of aqueous solution: Insight from hybrid QM/MM simulations using a polarizable solvent model," J. Comput. Chem. **18**, 1061–1071 (1997).
[62]C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, and B. Mennucci, "Electronic energy transfer in condensed phase studied by a polarizable QM/MM model," J. Chem. Theory Comput. **5**, 1838–1848 (2009).
[63]D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal, and B. Mennucci, "A QM/MM approach using the AMOEBA polarizable embedding: From ground state energies to electronic excitations," J. Chem. Theory Comput. **12**, 3654–3661 (2016).
[64]J. M. H. Olsen and J. Kongsted, "Chapter 3: Molecular properties through polarizable embedding," Adv. Quantum Chem. **61**, 107–143 (2011).
[65]J. Dziedzic, Y. Mao, Y. Shao, J. Ponder, T. Head-Gordon, M. Head-Gordon, and C.-K. Skylaris, "TINKTEP: A fully self-consistent, mutually polarizable QM/MM approach based on the AMOEBA force field," J. Chem. Phys. **145**, 124106 (2016).
[66]M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini, and B. Mennucci, "Polarizable embedding QM/MM: The future gold standard for complex (bio)systems?," Phys. Chem. Chem. Phys. **22**, 14433–14448 (2020).
[67]M. Nottoli, L. Cupellini, F. Lipparini, G. Granucci, and B. Mennucci, "Multiscale models for light-driven processes," Annu. Rev. Phys. Chem. **72**, 489–513 (2021).
[68]B. T. Thole, "Molecular polarizabilities calculated with a modified dipole interaction," Chem. Phys. **59**, 341 (1981).
[69]F. Lipparini, "General linear scaling implementation of polarizable embedding schemes," J. Chem. Theory Comput. **15**, 4312–4317 (2019).
[70]S. Jurinovich, C. Curutchet, and B. Mennucci, "The Fenna–Matthews–Olson protein revisited: A fully polarizable (TD)DFT/MM description," ChemPhysChem **15**, 3194–3204 (2014).
[71]L. Cupellini, S. Jurinovich, M. Campetella, S. Caprasecca, C. A. Guido, S. M. Kelly, A. T. Gardiner, R. Cogdell, and B. Mennucci, "An *ab initio* description of the excitonic properties of LH2 and their temperature dependence," J. Phys. Chem. B **120**, 11348–11359 (2016).
[72]V. Sláma, L. Cupellini, and B. Mennucci, "Exciton properties and optical spectra of light harvesting complex II from a fully atomistic description," Phys. Chem. Chem. Phys. **22**, 16783–16795 (2020).
[73]P.-F. Loos, A. Scemama, and D. Jacquemin, "The quest for highly accurate excitation energies: A computational perspective," J. Phys. Chem. Lett. **11**, 2374–2383 (2020).
[74]S. Maity, P. Sarngadharan, V. Daskalakis, and U. Kleinekathöfer, "Time-dependent atomistic simulations of the CP29 light-harvesting complex," J. Chem. Phys. **155**, 055103 (2021).
[75]J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," Chem. Rev. **105**, 2999–3093 (2005).
[76]B. Mennucci, "Polarizable continuum model," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 386–404 (2012).
[77]S. Jurinovich, L. Viani, I. G. Prandi, T. Renger, and B. Mennucci, "Towards an *ab initio* description of the optical spectra of light-harvesting antennae: Application to the CP29 complex of photosystem II," Phys. Chem. Chem. Phys. **17**, 14405–14416 (2015).
[78]F. Lipparini, G. Scalmani, L. Lagardère, B. Stamm, E. Cancès, Y. Maday, J.-P. Piquemal, M. J. Frisch, and B. Mennucci, "Quantum, classical, and hybrid QM/MM calculations in solution: General implementation of the ddCOSMO linear scaling strategy," J. Chem. Phys. **141**, 184108 (2014).
[79]M. Nottoli, R. Nifosì, B. Mennucci, and F. Lipparini, "Energy, structures, and response properties with a fully coupled QM/AMOEBA/ddCOSMO implementation," J. Chem. Theory Comput. **17**, 5661–5672 (2021).
[80]M. Corbella, L. Cupellini, F. Lipparini, G. D. Scholes, and C. Curutchet, "Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α-polypeptide chains," ChemPhotoChem **3**, 945–956 (2019).
[81]C. R. Jacob and J. Neugebauer, "Subsystem density-functional theory," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 325–362 (2014).
[82]C. König and J. Neugebauer, "Protein effects on the optical spectrum of the Fenna–Matthews–Olson complex from fully quantum chemical calculations," J. Chem. Theory Comput. **9**, 1808–1820 (2013).
[83]Z. Q. You and C. P. Hsu, "Theory and calculation for the electronic coupling in excitation energy transfer," Int. J. Quantum Chem. **114**, 102–115 (2014).
[84]L. Cupellini, M. Corbella, B. Mennucci, and C. Curutchet, "Electronic energy transfer in biomacromolecules," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **9**, e1392 (2019).
[85]P. López-Tarifa, N. Liguori, N. van den Heuvel, R. Croce, and L. Visscher, "Coulomb couplings in solubilised light harvesting complex II (LHCII): Challenging the ideal dipole approximation from TDDFT calculations," Phys. Chem. Chem. Phys. **19**, 18311–18320 (2017).
[86]M. E. Madjet, A. Abdurahman, and T. Renger, "Intermolecular Coulomb couplings from ab initio electrostatic potentials: Application to optical transitions of strongly coupled pigments in photosynthetic antennae and reaction centers," J. Phys. Chem. B **110**, 17268–17281 (2006).
[87]T. Renger and F. Müh, "Theory of excitonic couplings in dielectric media," Photosynth. Res. **111**, 47–52 (2012).
[88]C. Curutchet, G. D. Scholes, B. Mennucci, and R. Cammi, "How solvent controls electronic energy transfer and light harvesting: Toward a quantum-mechanical description of reaction field and screening effects," J. Phys. Chem. B **111**, 13253–13265 (2007).
[89]C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hossein-Nejad, G. D. Scholes, and B. Mennucci, "Photosynthetic light-harvesting is tuned by the heterogeneous polarizable environment of the protein," J. Am. Chem. Soc. **133**, 3078–3084 (2011).
[90]C.-P. Hsu, "The electronic couplings in electron transfer and excitation energy transfer," Acc. Chem. Res. **42**, 509–518 (2009).

© Author(s) 2022

[91]C.-H. Yang and C.-P. Hsu, "A multi-state fragment charge difference approach for diabatic states in electron transfer: Extension and automation," J. Chem. Phys. **139**, 154104 (2013).
[92]L. Cupellini, S. Caprasecca, C. A. Guido, F. Müh, T. Renger, and B. Mennucci, "Coupling to charge transfer states is the key to modulate the optical bands for efficient light harvesting in purple bacteria," J. Phys. Chem. Lett. **9**, 6892–6899 (2018).
[93]D. Abramavicius, B. Palmieri, D. V. Voronine, F. Šanda, and S. Mukamel, "Coherent multidimensional optical spectroscopy of excitons in molecular aggregates; quasiparticle versus supermolecule perspectives," Chem. Rev. **109**, 2350–2408 (2009).
[94]V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems* (John Wiley & Sons, 2008).
[95]T.-C. Dinh and T. Renger, "Towards an exact theory of linear absorbance and circular dichroism of pigment-protein complexes: Importance of non-secular contributions," J. Chem. Phys. **142**, 034104 (2015).
[96]A. Gelzinis, D. Abramavicius, and L. Valkunas, "Absorption lineshapes of molecular aggregates revisited," J. Chem. Phys. **142**, 154107 (2015).
[97]J. Ma and J. Cao, "Förster resonance energy transfer, absorption and emission spectra in multichromophoric systems. I. Full cumulant expansions and system-bath entanglement," J. Chem. Phys. **142**, 094106 (2015).
[98]L. Cupellini, F. Lipparini, and J. Cao, "Absorption and circular dichroism spectra of molecular aggregates with the full cumulant expansion," J. Phys. Chem. B **124**, 8610–8617 (2020).
[99]N. De Mitri, S. Monti, G. Prampolini, and V. Barone, "Absorption and emission spectra of a flexible dye in solution: A computational time-dependent approach," J. Chem. Theory Comput. **9**, 4507–4516 (2013).
[100]T. J. Zuehlsdorff and C. M. Isborn, "Combining the ensemble and Franck-Condon approaches for calculating spectral shapes of molecules in solution," J. Chem. Phys. **148**, 024110 (2018).
[101]D. Loco and L. Cupellini, "Modeling the absorption lineshape of embedded systems from molecular dynamics: A tutorial review," Int. J. Quantum Chem. **119**, e25726 (2019).
[102]T. Renger, A. Klinger, F. Steinecker, M. Schmidt am Busch, J. Numata, and F. Müh, "Normal mode analysis of the spectral density of the Fenna–Matthews–Olson light-harvesting protein: How the protein dissipates the excess energy of excitons," J. Phys. Chem. B **116**, 14565–14580 (2012).
[103]T. Renger and R. A. Marcus, "On the relation of protein dynamics and exciton relaxation in pigment–protein complexes: An estimation of the spectral density and a theory for the calculation of optical spectra," J. Chem. Phys. **116**, 9997–10019 (2002).
[104]M. Rätsep and A. Freiberg, "Electron–phonon and vibronic couplings in the FMO bacteriochlorophyll *a* antenna complex studied by difference fluorescence line narrowing," J. Lumin. **127**, 251–259 (2007).
[105]M. Rätsep, J. Pieper, K.-D. Irrgang, and A. Freiberg, "Excitation wavelength-dependent electron–phonon and electron–vibrational coupling in the CP29 antenna complex of green plants," J. Phys. Chem. B **112**, 110–118 (2008).
[106]P. Macak, Y. Luo, and H. Ågren, "Simulations of vibronic profiles in two-photon absorption," Chem. Phys. Lett. **330**, 447–456 (2000).
[107]F. J. Avila Ferrer and F. Santoro, "Comparison of vertical and adiabatic harmonic approaches for the calculation of the vibrational structure of electronic spectra," Phys. Chem. Chem. Phys. **14**, 13549–13563 (2012).
[108]D. Padula, M. H. Lee, K. Claridge, and A. Troisi, "Chromophore-dependent intramolecular exciton–vibrational coupling in the FMO complex: Quantification and importance for exciton dynamics," J. Phys. Chem. B **121**, 10026–10035 (2017).
[109]V. Macaluso, L. Cupellini, G. Salvadori, F. Lipparini, and B. Mennucci, "Elucidating the role of structural fluctuations, and intermolecular and vibronic interactions in the spectroscopic response of a bacteriophytochrome," Phys. Chem. Chem. Phys. **22**, 8585–8594 (2020).
[110]S. Valleau, A. Eisfeld, and A. Aspuru-Guzik, "On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations," J. Chem. Phys. **137**, 224103 (2012).
[111]V. I. Novoderezhkin, M. A. Palacios, H. Van Amerongen, and R. Van Grondelle, "Energy-transfer dynamics in the LHCII complex of higher plants: Modified redfield approach," J. Phys. Chem. B **108**, 10363–10375 (2004).
[112]C. Olbrich, J. Strümpfer, K. Schulten, and U. Kleinekathöfer, "Theory and simulation of the environmental effects on FMO electronic transitions," J. Phys. Chem. Lett. **2**, 1771–1776 (2011).
[113]S. Shim, P. Rebentrost, S. Valleau, and A. Aspuru-Guzik, "Atomistic study of the long-lived quantum coherences in the Fenna-Matthews-Olson complex," Biophys. J. **102**, 649–660 (2012).
[114]S. Chandrasekaran, M. Aghtar, S. Valleau, A. Aspuru-Guzik, and U. Kleinekathöfer, "Influence of force fields and quantum chemistry approach on spectral densities of BChl *a* in solution and in FMO proteins," J. Phys. Chem. B **119**, 9995–10004 (2015).
[115]T. J. Zuehlsdorff, H. Hong, L. Shi, and C. M. Isborn, "Influence of electronic polarization on the spectral density," J. Phys. Chem. B **124**, 531–543 (2019).
[116]S.-Y. Lu, T. J. Zuehlsdorff, H. Hong, V. P. Aguirre, C. M. Isborn, and L. Shi, "The influence of electronic polarization on nonlinear optical spectroscopy," J. Phys. Chem. B **125**, 12214–12227 (2021).
[117]S. M. Blau, D. I. G. Bennett, C. Kreisbeck, G. D. Scholes, and A. Aspuru-Guzik, "Local protein solvation drives direct down-conversion in phycobiliprotein PC645 via incoherent vibronic transport," Proc. Natl. Acad. Sci. U. S. A. **115**, E3342–E3350 (2018).
[118]M. Bondanza, L. Cupellini, F. Lipparini, and B. Mennucci, "The multiple roles of the protein in the photoactivation of orange carotenoid protein," Chem **6**, 187–203 (2020).
[119]S. Maity, V. Daskalakis, M. Elstner, and U. Kleinekathöfer, "Multiscale QM/MM molecular dynamics simulations of the trimeric major light-harvesting complex II," Phys. Chem. Chem. Phys. **23**, 7407–7417 (2021).
[120]T. J. Zuehlsdorff, S. V. Shedge, S.-Y. Lu, H. Hong, V. P. Aguirre, L. Shi, and C. M. Isborn, "Vibronic and environmental effects in simulations of optical spectroscopy," Annu. Rev. Phys. Chem. **72**, 165–188 (2021).
[121]S. V. Shedge, T. J. Zuehlsdorff, A. Khanna, S. Conley, and C. M. Isborn, "Explicit environmental and vibronic effects in simulations of linear and nonlinear optical spectroscopy," J. Chem. Phys. **154**, 084116 (2021).
[122]C. W. Kim, B. Choi, and Y. M. Rhee, "Excited state energy fluctuations in the Fenna–Matthews–Olson complex from molecular dynamics simulations with interpolated chromophore potentials," Phys. Chem. Chem. Phys. **20**, 3310–3319 (2018).
[123]A. Singharoy, C. Maffeo, K. H. Delgado-Magnero, D. J. K. Swainsbury, M. Sener, U. Kleinekathöfer, J. W. Vant, J. Nguyen, A. Hitchcock, B. Isralewitz *et al.*, "Atoms to phenotypes: Molecular design principles of cellular energy metabolism," Cell **179**, 1098–1111 (2019).
[124]V. Mascoli, V. Novoderezhkin, N. Liguori, P. Xu, and R. Croce, "Design principles of solar light harvesting in plants: Functional architecture of the monomeric antenna CP29," Biochim. Biophys. Acta, Bioenerg. **1861**, 148156 (2020).
[125]F. Häse, L. M. Roch, P. Friederich, and A. Aspuru-Guzik, "Designing and understanding light-harvesting devices with machine learning," Nat. Commun. **11**, 4587 (2020).
[126]J. Westermayr and P. Marquetand, "Machine learning for electronically excited states of molecules," Chem. Rev. **121**, 9873–9926 (2020).
[127]P. O. Dral and M. Barbatti, "Molecular excited states through a machine learning lens," Nat. Rev. Chem. **5**, 388–405 (2021).
[128]F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, "Physics-inspired structural representations for molecules and materials," Chem. Rev. **121**, 9759–9815 (2021).
[129]V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian process regression for materials and molecules," Chem. Rev. **121**, 10073–10141 (2021).
[130]C.-I. Wang, I. Joanito, C.-F. Lan, and C.-P. Hsu, "Artificial neural networks for predicting charge transfer coupling," J. Chem. Phys. **153**, 214113 (2020).
[131]M. Krämer, P. M. Dohmen, W. Xie, D. Holub, A. S. Christensen, and M. Elstner, "Charge and exciton transfer simulations using machine-learned Hamiltonians," J. Chem. Theory Comput. **16**, 4061–4070 (2020).
[132]A. Farahvash, C.-K. Lee, Q. Sun, L. Shi, and A. P. Willard, "Machine learning Frenkel Hamiltonian parameters to accelerate simulations of exciton dynamics," J. Chem. Phys. **153**, 074111 (2020).

© Author(s) 2022

[133]F. Häse, S. Valleau, E. Pyzer-Knapp, and A. Aspuru-Guzik, "Machine learning exciton dynamics," Chem. Sci. **7**, 5139–5147 (2016).
[134]M. S. Chen, T. J. Zuehlsdorff, T. Morawietz, C. M. Isborn, and T. E. Markland, "Exploiting machine learning to efficiently predict multidimensional optical spectra in complex environments," J. Phys. Chem. Lett. **11**, 7559–7568 (2020).
[135]C. Abrams and G. Bussi, "Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration," Entropy **16**, 163–199 (2013).
[136]O. Valsson, P. Tiwary, and M. Parrinello, "Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint," Annu. Rev. Phys. Chem. **67**, 159–184 (2016).
[137]Y. Miao and J. A. McCammon, "Unconstrained enhanced sampling for free energy calculations of biomolecules: A review," Mol. Simul. **42**, 1046–1055 (2016).
[138]A. S. Kamenik, S. M. Linker, and S. Riniker, "Enhanced sampling without borders: On global biasing functions and how to reweight them," Phys. Chem. Chem. Phys. **24**, 1225 (2022).
[139]C. Wehmeyer and F. Noé, "Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics," J. Chem. Phys. **148**, 241703 (2018).
[140]C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande, "Variational encoding of complex dynamics," Phys. Rev. E **97**, 062412 (2018).
[141]F. Giberti, G. A. Tribello, and M. Ceriotti, "Global free-energy landscapes as a smoothly joined collection of local maps," J. Chem. Theory Comput. **17**, 3292–3308 (2021).
[142]S.-T. Tsai, Z. Smith, and P. Tiwary, "SGOOP-d: Estimating kinetic distances and reaction coordinate dimensionality for rare event systems from biased/unbiased simulations," J. Chem. Theory Comput. **17**, 6757–6765 (2021).

© Author(s) 2022