This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

This article can be cited before page numbers have been issued, to do this please use: K. Schwinn, N. Ferré and M. Huix-Rotllant, Phys. Chem. Chem. Phys., 2020, DOI: 10.1039/D0CP01714K.
1 Introduction

Cryptochromes are a type of flavoproteins proposed as candidates to explain magnetoreception of animals, plants and bacteria. The main hypothesis is that a biradical is formed upon blue-light absorption by flavin adenine dinucleotide (FAD). In protein milieu, the oxidized form of FAD can be reduced, leading to four derive redox forms: anionic and neutral semi-reduced radicals, and anionic and neutral fully reduced forms. All these forms have a characteristic electronic absorption spectrum, with a strong vibrational resolution. Here, we employ a normal mode analysis at the electrostatic embedding QM/MM level of theory to compute the vibrationally resolved absorption spectra of five redox forms of FAD embedded in plant cryptochrome. We show that explicit account of vibrational broadening contributions to electronic transitions are essential to reproduce the experimental spectra. In the case of the neutral radical form of FAD, the absorption spectrum is reproduced only if the presence of a tryptophan radical is considered.

The UV-visible absorption spectrum is an essential tool for detecting the different redox states of FAD in protein. Indeed, the five forms have a characteristic absorption spectrum and lifetime.\(^a\) Simulations of absorption spectra of FAD and its redox and acid-base derivatives requires explicit account of vibrations in order to reproduce the color of absorption and emission. This was first shown by Saalfrank and coworkers for the fully oxidized form of riboflavin.\(^\text{11}\) Recently, several groups have computed the vibrationally resolved absorption spectra for five redox forms of lumichrome-10H in gas phase,\(^\text{12}\) lumiflavin in several solvents,\(^\text{13}\) and the absorption and emission spectra for isoalloxazine, lumichrome, lumiflavin and riboflavin in solution using implicit continuum solvent models.\(^\text{2}\)

Even though they are detectable in solution, most of the redox derivatives of FAD (especially for the radical forms) are stabilized in protein.\(^\text{14}\) Currently, a systematic simulation of the absorption spectra of FAD and its redox derivatives in protein milieu is missing, which would be essential to interpret the origin of the peaks. Pioneer work by Cannuccia and coworkers,\(^\text{15}\) and more recently Solov'yov and coworkers,\(^\text{16}\) have performed a theoretical study of the vertical transitions at the QM/MM level for the oxidized form of flavin mononucleotide and FAD respectively. However, none of these studies included explicitly the effect of nuclear vibrations in their simulations.

Here, we perform the first simulation of vibrationally resolved electronic absorption spectrum of FAD in five redox forms in plant cryptochrome. For this purpose, we have implemented our recently developed method for computing analytic second derivatives of electrostatic embedding QM/MM energy to unrestricted single-determinant wavefunctions,\(^\text{17}\) necessary to simulate the vi-
brationally resolved spectra of radical species. Our theoretical absorption spectra are in good agreement to the experimental spectra. For FADH\textsuperscript{+}, the spectra of tyrosil and tryptophanyl radicals have to be included to reproduce the experimental spectra.

2 Methodology

Simulations of vibrationally resolved spectra requires the computation of the Hessian for the whole QM/MM system. Currently, there are few methods capable to compute efficiently the QM/MM Hessian for a large macromolecule like a protein, mostly relying on approximations\cite{NoseRH2017,BoeseH2014}. We have recently developed one of the first analytic electrostatic embedding QM/MM Hessians, which is both efficient and without approximations.\cite{DiazG2019} Here, we summarize the main equations following the electrostatic potential fitting (ESPF) method and its derivatives, focusing on the new implementation for unrestricted wavefunctions. The reader is referred to the previous literature for further details on closed shell ESPF equations and analytic QM/MM interaction energy derivatives.\cite{DiazG2019,BoeseH2014}

2.1 Electrostatic potential fitting method

The ESPF method is an electrostatic embedding QM/MM method,\cite{BoeseH2014} which adds a one-electron potential in the usual (gas-phase) Fock operator ($\hat{F}$),

$$\hat{F} = \hat{F}_0 + \hat{h}.$$  \hspace{1cm} (1)

The ESPF hamiltonian is constructed from a (truncated) multipole order expansion. The lowest-order ESPF hamiltonian ($\hat{h}$) is given by

$$\hat{h} = \sum_{\sigma A} (Z_A - \tilde{Q}_A) \phi_{\sigma A},$$ \hspace{1cm} (2)

where $N_{QM}$ is the number of QM atoms, $Z_A$ is the nuclear charge, $\phi_{\sigma A}$ is the external field felt by QM atom $A$ and the atomic charge operator, the matrix elements of which are defined as $Q_A = \sum_{\sigma}(|T^{\sigma}T|^{-1}T^{\sigma})_{\sigma A} V_k$. Here, $V_k$ are electrostatic integrals calculated on a grid surrounding the QM subsystem and $T$ the electrostatic kernel.\cite{BoeseH2014} Usually, the lowest order approximation is a good approximation for the QM/MM interaction.\cite{DiazG2019} Hereafter, we focus on the lowest order, despite being straightforward to generalize it to higher orders in the multipole expansion.\cite{DiazG2019}

The ESPF charge operator is spin independent, so the only spin dependency comes from the density matrix. For a single determinant wavefunction expanded in molecular orbitals, the ESPF QM/MM energy is then obtained as

$$E = Tr[\hat{P}\hat{F}_0] + E_{MM} + Tr[\hat{P}\hat{H}] = E_0 + E_{MM} + E_{ESPF}, \hspace{1cm} (3)$$

in which $\hat{P} = \hat{P}_0 + \hat{P}^{\delta}$ is the atomic orbital density matrix, and $\hat{P}^{\delta} = \sum_{\sigma} C_{\alpha}^{\delta} C_{\beta}^{\delta}$ ($\sigma = \alpha, \beta$). The $C$ are the molecular orbital coefficients obtained from the solution of the secular equation with the Fock operator defined in Eq. (1). The ESPF energy interaction term can be rewritten as

$$E_{ESPF} = Tr[\hat{P}\hat{H}] = \sum_{\sigma A} q_{\sigma A} \phi_{\sigma A}, \hspace{1cm} (4)$$

in which we defined the atomic charge as the sum of the ESPF charge and the nuclear charge,

$$q_A = Z_A - Tr[P\tilde{Q}_A]. \hspace{1cm} (5)$$

In these equations, the classical analogue of the energy expression Eq. (3) becomes apparent.

In case of unrestricted wavefunctions, the $\alpha$ and $\beta$ contributions to the ESPF charges can be defined as $Q_{\alpha N} = Tr[P^{\sigma}\tilde{Q}_A]$. Since classical force fields are spin-independent, the ESPF energy expression Eq. (4) remains unchanged, and the total atomic charge approximating the QM atom $A$ charge distribution is simply the sum of $\alpha$ and $\beta$ components,

$$q_A = Z_A - Q_{\alpha N} - Q_{\beta N}.$ \hspace{1cm} (6)$$

2.2 Analytic second derivatives of the ESPF energy

The analytic first and second derivatives of the ESPF energy have been described in Refs. \cite{DiazG2019} and \cite{BoeseH2014} respectively. The ESPF energy depends on the coordinates of the QM and MM atoms. Therefore, its second derivative with respect to these coordinates contains four blocks,

$$\nabla^2 E = \begin{bmatrix} E_{\sigma \sigma}^{\sigma \sigma} + E_{\sigma \sigma}^{\sigma \delta} + E_{\delta \sigma}^{\sigma \sigma} + E_{\delta \sigma}^{\sigma \delta} & E_{\sigma \delta}^{\sigma \sigma} + E_{\sigma \delta}^{\sigma \delta} & E_{\sigma \delta}^{\delta \sigma} + E_{\sigma \delta}^{\delta \delta} & E_{\delta \sigma}^{\sigma \sigma} + E_{\delta \sigma}^{\sigma \delta} & E_{\delta \sigma}^{\delta \sigma} + E_{\delta \sigma}^{\delta \delta} \\ E_{\sigma \sigma}^{\delta \sigma} + E_{\sigma \delta}^{\delta \sigma} & E_{\sigma \delta}^{\sigma \sigma} + E_{\sigma \delta}^{\sigma \delta} & E_{\sigma \delta}^{\delta \sigma} + E_{\sigma \delta}^{\delta \delta} & E_{\delta \sigma}^{\sigma \sigma} + E_{\delta \sigma}^{\sigma \delta} & E_{\delta \sigma}^{\delta \sigma} + E_{\delta \sigma}^{\delta \delta} \\ E_{\sigma \sigma}^{\delta \sigma} & E_{\sigma \delta}^{\delta \sigma} & E_{\sigma \delta}^{\delta \sigma} & E_{\delta \sigma}^{\delta \sigma} & E_{\delta \sigma}^{\delta \delta} \\ E_{\sigma \sigma}^{\delta \delta} & E_{\sigma \delta}^{\delta \delta} & E_{\sigma \delta}^{\delta \delta} & E_{\delta \sigma}^{\delta \delta} & E_{\delta \sigma}^{\delta \delta} \end{bmatrix}, \hspace{1cm} (7)$$

in which $E_{\sigma \delta}^{\rho \rho} = \frac{\partial^2 E}{\partial \rho \partial \rho}$. The $\rho$ corresponds to the $\chi$ coordinate of a QM atom, while the tilde symbolizes an MM atom. The analytic second derivatives of $E_0$ have been lengthy discussed in the literature,\cite{DiazG2019} while the $E_{MM}$ derivatives have a trivial expression. Here we focus on the derivative of the ESPF energy term, which can be written in orders of external potential derivatives as

$$\nabla^2 E_{ESPF} = \begin{bmatrix} \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\delta A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\delta A} \\ \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\delta A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\delta A} \\ \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\delta A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\delta A} \\ \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\sigma A} \phi_{\delta A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\sigma A} & \sum_A q_{\sigma A} \phi_{\delta A} \phi_{\delta A} \end{bmatrix}, \hspace{1cm} (8)$$

in which the derivatives in parenthesis means derivatives at fixed MO coefficients. The first term in Equation (8) corresponds to variations of the MM external field at fixed QM atomic charges (the unique term in mechanical embedding methods), which is equivalent to the second-derivative of the classical electrostatic interaction. The second term and third terms include the charge responses to external field variations, which is due to the fact that the ESPF atomic charges are fluctuating depending on the system MM geometry. In other words, the ESPF charges are polarizable.

The atomic charge derivatives $q_{\sigma A}^\alpha$ and $q_{\sigma A}^\delta$ include the derivatives of the density matrix and of the ESPF charge matrix,

$$q_{\sigma A}^\alpha = -Tr[P^\sigma \tilde{Q}_A] - Tr[P \tilde{Q}_A],$$

$$q_{\sigma A}^\delta = -Tr[P^\delta \tilde{Q}_A] - Tr[P \tilde{Q}_A]. \hspace{1cm} (9)$$
The second equation requires, in principle, the construction of the density matrix derivatives with respect to MM perturbations. The latter are usually obtained by solving a set of coupled-perturbed equations for all MM perturbations. Since the number of MM atoms is usually much larger than the number QM atoms, we have introduced in Ref. [23] an auxiliary set of coupled-perturbed equations, the Q-vector equations, which allow to compute this term at an almost negligible cost. The Q-vector equations scale with the number of QM atoms, and take the usual form of the coupled perturbed equations

\[
\sum_{j,b} (A - B)_{ia,jb} \hat{Q}_{A,jb} = -Q_{A,ia}.
\]

in which \(i, j\) are indexes of occupied molecular orbitals, and \(a, b\) are indexes of virtual orbitals, and \((A-B)\) is the usual response matrix (for an explicit expression, see Eq. 13 of Ref. [23]). The solution of this equation, \(\hat{Q}_{A,jb}\), is then used to construct the ESPF charge derivative with respect to MM atoms,

\[
q_{A}^{i} = -2 \sum_{B,ia} \hat{Q}_{A,ia} \theta_{B,ia}.
\]

In the case of QM/MM energy second derivatives based on an unrestricted determinant, we can use similar arguments as we did in Equation (6). Since the force field derivatives do not depend on the spin, all charge and charge derivatives of Eq. (8) involves the sum of \(\alpha\) and \(\beta\) contributions. The Q-vector equation (Eq. 10) is split into two expressions

\[
\sum_{j,b,\sigma=\alpha,\beta} (A - B)_{ia,\alpha,jb,\sigma=\alpha,\beta} \hat{Q}_{A,jb,\sigma} = -Q_{A,ia,\alpha},
\]

\[
\sum_{j,b,\sigma=\alpha,\beta} (A - B)_{ia,\alpha,jb,\sigma=\alpha,\beta} \hat{Q}_{A,jb,\sigma} = -Q_{A,ia,\beta},
\]

in which \(\sigma = \alpha, \beta\). Accordingly, the total charge derivative with respect to MM coordinates becomes

\[
q_{A}^{i} = - \sum_{B,ia,\sigma=\alpha,\beta} Q_{A,ia,\sigma} \theta_{B,ia,\sigma}.
\]

3 Computational details

The ESPF QM/MM hessian for unrestricted wavefunctions described above has been implemented in a modified version of Gaussian16 interfaced with a modified version of TINKER 6.3.1. [17,23,24,27] The X-ray structure of Arabidopsis Thaliana plant cryptochrome (RCSB PDBID: 2J4D) has been used as molecular model. Hydrogen atoms have been added to the standard protonation state at physiological pH. For the QM computations, B3LYP/TZVP/TZVPfit level of theory has been employed for DFT and TD-DFT calculations. [29,33] For the MM computations, Amber99 force field has been used with Van der Waals parameters for FAD taken from Ref. [37]. The vibrationally resolved spectrum for FAD has been obtained by Fourier transforming the time-dependent auto-correlation function for the multi-dimensional harmonic oscillator, as implemented in Gaussian16. [27,28] The vertical gradient approximation has been used in all cases. [29]

4 Results

The QM/MM model is defined as FAD (QM subsystem) and the protein scaffold (MM subsystem), from which crystallographic water molecules have been excluded (see Figure 1). The overall system has been optimized with a quadratically convergent algorithm for the QM part and microiterations for the MM part. From the minimum energy structure of FAD, the subsequent structures have been constructed for the reduced forms and each of them reoptimized. All structures have been confirmed via a frequency calculation, leading to a positive semi-definite internal Hessian matrix in all cases.

4.1 Geometries

In the series FAD \(\rightarrow\) FAD\(^{+}\) \(\rightarrow\) FAD\(^{+}\) \(\rightarrow\) FADH\(^{−}\) \(\rightarrow\) FADH\(_{2}\), the lowest occupied molecular orbital (LUMO) becomes a singly occupied molecular orbital (SOMO) after the first electron attachment and the highest occupied molecular orbital (HOMO) after the second electron attachment. The orbital involved in this transformation is essentially anti-bonding between \(N_{3}-C_{4a}\) and bonding between \(C_{4a}-C_{10a}\). This transforms the \(\alpha\)-dimine bond \((N_{3}=C_{4a}-C_{10a}=N_{1})\) of the oxidized form in a ethene-1,2-diamine (\(N_{3}-C_{4a}=C_{10a}-N_{1}\)) for the fully reduced forms. [2]

In Figure 2 the effect of the protein scaffold on the chromophore optimized (local) minimum energy structure is compared with respect to the gas phase minimum energy structure. For the protein, this corresponds to a local minimum of the internal energy, which is taken as reference to do the harmonic expansion. Overall, one observes only minor structural differences when the chromophore is embedded in protein with respect to the gas phase. This is evident for bond distances, in which changes of less than 0.1 Å are observed for all five forms. Conversely, angles and dihedral angles show important variations for all cases. This is especially important in the case of the fully reduced forms, FAD\(^{+}\) and FADH\(_{2}\). These forms contain secondary amines at centres \(N_{3}\) and/or \(N_{1}\) atoms. Secondary amines tend to break
the planarity of the chromophore both in gas phase or in solution (see Fig. 3). Indeed, in the gas phase these two forms find its minimum energy structure in the so-called “butterfly” bending conformation. This structural feature is hindered when the chromophore is embedded in protein, which favours a rather planar structure. Despite the weak electrostatic effects on the π system and thus the chromophore excitations,[11] the conformational constraints have a strong impact on the peak position (see further in the next sections).

4.2 Absorption spectra of FAD and its redox derivatives embedded in protein

In Figure 4, the total theoretical vibrationally resolved absorption spectrum is shown in the range between 300 nm and 800 nm. In addition, the contributions of the most important electronic configurations are also explicitly shown. For all redox forms, the lowest-energy peak is represented by a single electronic transition of the type π → π* centred on the isoalloxazine ring. The second peak of FAD, FAD\textsuperscript{−} and FADH\textsuperscript{+} shows a complex pattern, involving three to four electronic transitions. For FADH\textsuperscript{−} and FADH\textsubscript{2}, the second transition finds its maximum above 300 nm, and it corresponds to a π → π* transition in both cases.

For all redox forms of FAD, the frontier molecular orbital π\textsubscript{1} orbital plays a central role. This orbital exhibits bonding character between C\textsubscript{4}−C\textsubscript{10a} and anti-bonding character between N\textsubscript{10}−C\textsubscript{10a}. For FAD, π\textsubscript{1} is the lowest unoccupied molecular orbital (LUMO). For the radical forms, π\textsubscript{1} is a singly occupied molecular orbital (SOMO). For the reduced forms, π\textsubscript{1} is the highest occupied molecular orbital (HOMO). For the oxidized and anionic radical forms, a π\textsubscript{2} → π\textsubscript{1} transition is observed in the region of 450 nm. This transition is red-shifted in the series FAD → FAD\textsuperscript{−} → FADH\textsuperscript{+} since the energetic gap between π\textsubscript{2} and π\textsubscript{1} becomes smaller when FAD is reduced. Indeed, the absorption maxima are predicted at around 460 nm, 440 nm and 650 nm for FAD, FAD\textsuperscript{−} and FADH\textsuperscript{+} respectively, corresponding to an absorption color ranging from dark blue, to blue and finally red. For the fully reduced forms, the π\textsubscript{1} → π\textsubscript{1} becomes the bright transition, which is blue-shifted with respect to the first transition of the oxidized and semi-reduced forms. Indeed, the absorption maxima are found in the near-UV, around 325 nm for FAD\textsuperscript{−} and 345 nm for FADH\textsubscript{2}. All these transitions have a similar intensity, corresponding to a broad bright peak of 100 nm.

A second intense peak appears between 300 nm and 800 nm for the FAD, FAD\textsuperscript{−} and FADH\textsuperscript{+} species. For the reduced forms, this peak is located outside the considered range, between 200 nm and 300 nm, corresponding to a π\textsubscript{3} → π\textsubscript{3} transition. For the oxidized and radical forms, this peak is found around 300 nm and 500 nm and it includes 3 to 4 different electronic transitions. For the radical forms, the most intense transition corresponds to the electron transfer from the SOMO π\textsubscript{1} to the vacant orbital π\textsubscript{2}. This transition is the same as the first bright transition for the reduced forms, although it is around 50 nm blue-shifted for the radical species. Indeed, the π\textsubscript{3} → π\textsubscript{2} absorption maximum is found at around 340 nm for FAD\textsuperscript{−} and 360 nm for FADH\textsuperscript{+}. This type of transition is not present in FAD, as
in this form, the $\pi_1$ is a vacant orbital. For FAD, the $\pi_1$ is accepting electrons from lower lying $\pi$ orbitals and from oxygen lone-pairs located on the phosphate ($n_{\text{phosphate}}$) and on the ribitol ($n_{\text{ribitol}}$).

These transitions, which would normally be dark, are appearing as peaks of a similar intensity of the $\pi \rightarrow \pi^*$ transitions due to two main reasons: (i) these transitions are strongly mixed with $\pi \rightarrow \pi^*$ transitions, thus becoming slightly bright thanks to an intensity-borrowing mechanism, and (ii) the wavefunction overlap between the ground and excited states is large. A similar trend is observed for the radical forms.

5 Discussion

The UV-visible absorption spectra of flavins and derivatives (riboflavin, lumichrome, lumiflavin, isalloxazine, etc.) have been lengthy explored in the literature, both from an experimental [11,12,46] and theoretical point of view [13,14]. Theoretically, these have been studied mainly in gas phase and within an implicit description of the solvent. Some of these forms, especially for the radical forms, are usually stabilized when embedded in proteins. To our knowledge, only two theoretical studies have attempted to compute the absorption spectrum in protein, the work of Canuccia et al. [15] and the work of Solov'yov et al. [16], albeit only for the oxidized form and without explicitly accounting for the vibrational structure. Here, we attempt to compare the theoretical absorption spectra in protein and in gas phase with the experimental spectra of FAD and its redox derivatives in protein.

This is shown in Fig. 5. Since many flavoproteins share similar spectral properties [47], some of the experimental data have been taken from different forms of cryptochromes or photolyases: data for FAD and FAD$^-$ have been taken from Mosquito Cryptochrome 1 (Anopheles gambiae), and FADH* and FADH$^-$ from a photolyase of E. Coli. The four forms are similar to the absorption spectrum of Arabidopsis Thaliana cryptochrome 1 and have been taken from Ref. 45. The spectrum of FADH$_2$ has been taken from solution measurements, and it is reproduced from data taken from Ref. 10.

In the following, we discuss each form separately.

**FAD:** The oxidized form of FAD is the most stable one, both in solution and in protein. For this reason, it has been extensively studied in the literature both experimentally [11,12,46,48] and theoretically [13,14]. FAD has two main absorption bands, the first band in the blue visible light region and a second band in
The absorption spectrum of FADH\textsuperscript{−} has a complex structure, with a peak at 650 nm and another at 450 nm, compared to the more simple absorption spectrum of FAD\textsuperscript{−} with a peak at 500 nm. In solution, the peak at 650 nm in FADH\textsuperscript{−} is intensified, and a shoulder is observed at 450 nm. The absorption spectrum of FAD\textsuperscript{−} is shifted to 500 nm when both TrpH and TyrO are present, indicating a red shift of about 100 nm compared to the gas phase absorption spectrum. This is consistent with the experimental data, which show that the absorption spectrum of FAD\textsuperscript{−} is shifted to 500 nm when both TrpH and TyrO are present. The absorption spectrum of FAD\textsuperscript{−} is shifted to 500 nm when both TrpH and TyrO are present, indicating a red shift of about 100 nm compared to the gas phase absorption spectrum. This is consistent with the experimental data, which show that the absorption spectrum of FAD\textsuperscript{−} is shifted to 500 nm when both TrpH and TyrO are present.
FADH\textsuperscript{−}: The fully reduced state is not favored in vivo and in vitro requires a strong reducing agent. In protein the reduced anion is formed after a second electron transfer from tryptophan, usually after a second photon absorption from FADH\textsuperscript{•}. Reoxidation happens in the presence of molecular oxygen. In solution the anionic reduced form has a maximum absorption at 256 nm and shoulders at 288 and 350 nm. A study in E. coli photolyase showed an absorption band around 360 nm, with very weak absorbance in the visible region.

Our simulations both in protein and in gas phase match the experimental spectrum, both in term of shape and vibrational progression. In the experimental spectrum, a shoulder is appearing at around 450 nm, which is probably due to some residual species (like a small concentration of Trp\textsuperscript{•}) or an almost dark π → π\textsuperscript{*} transition as found in experimental Stark spectra and theoretical simulations of FADH\textsuperscript{−} in solution. This feature is not captured by our QM/MM simulations, probably due to an underestimation of the intensity of this transition.

FADH\textsubscript{2}: The neutral reduced form is formed after a second proton transfer from tryptophan from the FADH\textsuperscript{−} form. In protein, its spectrum hardly differs from FADH\textsuperscript{−} and its characterization is mostly limited to solution studies, where it has a distinguished shoulder at 400 nm.

The experimental measurements for this form shown in Fig. 5 were done in solution. These match well with the gas-phase calculation, but the protein calculation shows a blue-shift of this peak of around 100 nm, leading to a similar excitation energy than FADH\textsuperscript{−}. This large blue-shift does not have an electrostatic origin, but it can be rather attributed to the enforced planar structure due to steric hindrance of the protein, since the electronic transition of FADH\textsubscript{2} in the gas phase at the protein optimized conformation leads to the same absorption maximum. Unfortunately, there is no clear experimental evidence of this form in cryptochrome protein to determine whether this shift of the QM/MM simulations is realistic.

6 Conclusions

The simulation of UV-visible absorption spectrum in complex media is a difficult task for theoretical chemistry. Here, we have developed and implemented the analytic second-derivatives of electrostatic embedding QM/MM ESPF method for unrestricted wavefunctions, extending our previous developments to open-shell systems. This method allows, among other molecular properties, the vibrational analysis of minimum energy structures and the simulation of vibrationally resolved spectra in protein. We have applied this methodology to simulate the absorption spectrum of FAD and four redox derivatives (two radicals and two reduced forms, one of each anionic and one neutral) embedded in the Arabidopsis Thaliana cryptochrome protein.

The theoretical absorption spectrum in protein shows a perfect agreement with experiments for the non-radical forms, reproducing well both the intensity, peak position and vibrational peak structure of the electronic transitions found between 300 nm and 600 nm. On the contrary, the agreement in the case of the radical forms is only moderate. As discussed recently in the literature, the signature of tryptophan and tyrosine radicals could be found in the absorption spectra of FAD\textsuperscript{−} and FADH\textsuperscript{•}. On the one hand, for FAD\textsuperscript{−}, a tyrosil peak appears at the same position as a π → π\textsuperscript{*} transition. The position and shape of both peaks is similar, and therefore we can not conclude that a tyrosil signature is present. On the other hand, the theoretical FADH\textsuperscript{•} absorption spectrum in protein can only be matched to the experimental spectrum considering the lowest intense peak of Trp\textsuperscript{•} radical at around 500 nm. We estimate the proportion between the two species as 1:1.

In conclusion, we demonstrate that vibrational analysis at the electrostatic embedding QM/MM level is a valuable tool to interpret the complex absorption spectra arising in proteins. In the future, we plan to extend our method to the simulation of vibrational spectroscopy (infra-red, non-resonant Raman, etc.) as well as extending the vibrational analysis to other kind of wavefunctions.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge financial support by the “Agence Nationale pour la Recherche” through the project BIOMAGNET (ANR-16CE29-0008-01). “Centre de Calcul Intensif d’Aix-Marseille” is acknowledged for granting access to its high performance computing resources.
Notes and references

1. X. Yu, H. Liu, J. Kleinit and C. Lin, *The Arabidopsis Book*, 2010, 2010, year.
2. S. Braatsch and G. Klug, *Photosynthesis Research*, 2004, 79, 45–57.
3. I. Chaves, R. Pokorny, M. Byrerd, N. Hoang, T. Ritz, K. Brettel, L.-O. Essen, G. T. J. van der Horst, A. Batschauer and M. Ahmadd, *Ann. Rev. Plant. Biol.*, 2011, 62, 335–364.
4. C. Lin, D. Top, C. C. Manahan, M. W. Young and B. R. Crane, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, 115, 3822–3827.
5. M. Liedvogel and H. Mouitzen, *J. R. Soc. Interface*, 2010, 7, S147–S162.
6. A. Pinzon-Rodriguez, S. Bensch and R. Muheim, *J. R. Soc. Interface*, 2018, 15, 20180058.
7. P. Mondal, K. Schwinn and M. Huix-Rotllant, *J. Photochem. Photobiol. A: Chem.*, 2020, 387, 112164.
8. I. A. Solov’yov, T. Domatricea, A. R. Moughal Shahi and K. Schulten, *J. Am. Chem. Soc.*, 2012, 134, 18046–18052.
9. P. Mondal and M. Huix-Rotllant, *Phys. Chem. Chem. Phys.*, 2019, 21, 8874–8882.
10. Y.-T. Kao, C. Saxena, T.-F. He, L. Guo, L. Wang, A. Sancar and D. Zhong, *J. Am. Chem. Soc.*, 2008, 130, 13132–13139.
11. B. Klauamünzer, D. Kröner and P. Saalfrank, *J. Phys. Chem. B*, 2010, 114, 10826–10834.
12. Y. Ai, C. Zhao, J. Xing, Y. Liu, Z. Wang, J. Jin, S. Xia, G. Cui and X. Wang, *J. Phys. Chem. A*, 2018, 122, 7954–7961.
13. M. P. Kabir, Y. Orozco-Gonzalez and S. Gozem, *Phys. Chem. Chem. Phys.*, 2019, 21, 16526–16537.
14. F. Müller, *Free Radial. Biol. Med.*, 1987, 3, 215 – 230.
15. E. Cannuccia, O. Pulci, R. D. Sole and M. Cascella, *Chem. Phys.*, 2011, 389, 35–38.
16. C. Nielsen, M. S. Norby, J. Jongted and I. A. Solov’yov, *J. Phys. Chem. Lett.*, 2018, 9, 3618–3623.
17. K. Schwinn, N. Ferré and M. Huix-Rotllant, *J. Chem. Theor. Comput.*, 2020, 0, null.
18. S. Dapprich, I. Komáromi, K. Byun, K. Morokuma and F. J. Molec. Struc.: THEOCHEM, 1999, 461-462, 1–21.
19. Q. Cui and M. Karplus, *J. Chem. Phys.*, 2000, 112, 1133–1149.
20. Q. Cui and M. Karplus, *J. Phys. Chem. B*, 2000, 104, 3721–3743.
21. A. Ghysels, H. L. Woodcock, J. D. Larkin, B. T. Miller, Y. Shao, J. Kong, D. V. Neck, V. V. Speybroeck, M. Waroquier and B. R. Brooks, *J. Chem. Theor. Comput.*, 2011, 7, 496–514.
22. T. Giovannini, L. Grazioili, M. Ambrosetti and C. Cappelli, *J. Chem. Theory Comput.*, 2019, 15, 5495–5507.
23. K. Schwinn, N. Ferré and M. Huix-Rotllant, *J. Chem. Phys.*, 2019, 151, 041102.
24. N. Ferré and J. G. Ángyán, *Chem. Phys. Lett.*, 2002, 356, 331–339.
25. F. Melaccio, M. Olivucci, R. Lindh and N. Ferré, *Int. J. Quant. Chem.*, 2011, 111, 3339–3346.
26. J. A. Pope, R. Krishnan, H. B. Schlegel and J. S. Binkley, *Int. J. Quant. Chem.*, 1979, 16, 225–241.
27. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Peterson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparrini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16 Revision A.01, 2016, Gaussian Inc. Wallingford CT.
28. J. W. Ponder, TINKER: Software tools for molecular design, version 6.3, 2010, [http://dasher.wustl.edu/tinker/](http://dasher.wustl.edu/tinker/)
29. A. D. Becke, J. *Chem. Phys.*, 1993, 98, 5648–5652.
30. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, 37, 785–789.
31. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, 58, 1200–1211.
32. A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, 97, 2571–2577.
33. A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, 100, 5829–5835.
34. K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, 240, 283 – 290.
35. K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, 97, 119–124.
36. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Gleihr, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.
37. J. Antony, D. Medvedev and A. Stuchebrukhov, *J. Am. Chem. Soc.*, 2000, 122, 1057.
38. A. Biaiardi, J. Blöino and V. Barone, *J. Chem. Theor. Comput.*, 2013, 9, 4097–4115.
39. P. Macak, Y. Luo and H. Ågren, *Chem. Phys. Lett.*, 2000, 330, 447 – 456.
40. Y. S. Choong and V. Massey, *J. Biol. Chem.*, 1980, 255, 8672–7.
41. C. Lin, D. Robertson, M. Ahmad, A. Raibekas, M. Jorns, P. Dutton and A. Cashmore, *Science*, 1995, 269, 968–970.
42. B. Giovani, M. Byrdin, M. Ahmad and K. Brettel, *Nat. Struct. Mol. Biol.*, 2003, **10**, 489–490.
43. S.-H. Song, B. Dick, A. Penzkofer, R. Pokorny, A. Batschauer and L.-O. Essen, *J. Photochem. Photobiol. B*, 2006, **85**, 1–16.
44. T. Kottke, A. Batschauer, M. Ahmad and J. Heberle, *Biochemistry*, 2006, **45**, 2472–2479.
45. B. Liu, H. Liu, D. Zhong and C. Lin, *Curr. Opin. Plant Biol.*, 2010, **13**, 578–586.
46. P. Müller, J.-P. Bouly, K. Hitomi, V. Balland, E. D. Getzoff, T. Ritz and K. Brettel, *Sci. Rep.*, 2014, **4**, 5175.
47. P. Macheroux, *UV-Visible Spectroscopy as a Tool to Study Flavoproteins*, Humana Press, 1999, pp. 1–7.
48. J. Koziol, *Experientia*, 1965, **21**, 189–190.
49. J. Koziol, *Photochem. Photobiol.*, 1966, **5**, 41–54.
50. J. Koziol, *Photochem. Photobiol.*, 1969, **9**, 45–53.
51. A. Weigel, A. L. Dobryakov, M. Veiga and J. L. P. Lustres, *J. Phys. Chem. A*, 2008, **112**, 12054–12065.
52. P. Zirak, A. Penzkofer, T. Mathes and P. Hegemann, *Chem. Phys.*, 2009, **358**, 111–122.
53. T. Kottke, J. Heberle, D. Hehn, B. Dick and P. Hegemann, *Biophys. J.*, 2003, **84**, 1192–1201.
54. E. Sikorska, I. V. Khmelinskii, J. Koput, J. L. Bourdelange and M. Sikorski, *J. Molec. Struct.*, 2004, **697**, 137–141.
55. A. Penzkofer, *J. Photochem. Photobiol. A: Chem.*, 2016, **314**, 114–124.
56. C. B. Martin, X. Shi, M.-L. Tsao, D. Karweik, J. Brooke, C. M. Hadad and M. S. Platz, *J. Phys. Chem. B*, 2002, **106**, 10263–10271.
57. M. Takahashi, Y. Ishikawa, J.-i. Nishizawa and H. Ito, *Chem. Phys. Lett.*, 2005, **401**, 475–482.
58. S. Salzmann, J. Tatichen and C. M. Marian, *J. Photochem. Photobiol. A: Chem.*, 2008, **198**, 221–231.
59. S. Salzmann and C. M. Marian, *Chem. Phys. Lett.*, 2008, **463**, 400–404.
60. J.-y. Hasegawa, S. Bureekaew and H. Nakatsuji, *J. Photochem. Photobiol. A: Chem.*, 2007, **189**, 205–210.
61. M. M. N. Wolf, C. Schumann, R. Gross, T. Domratcheva and R. Diller, *J. Phys. Chem. B*, 2008, **112**, 13424–13432.
62. C. Neiss, P. Saalfrank, M. Parac and S. Grimme, *J. Phys. Chem. A*, 2003, **107**, 140–147.
63. M. Dittric, P. L. Reddistino and K. Schulten, *J. Phys. Chem. B*, 2005, **109**, 13006–13013.
64. Y.-K. Choe, S. Nagase and K. Nishimoto, *J. Comput. Chem.*, 2007, **28**, 727–739.
65. K. Maeda, A. J. Robinson, K. B. Henbest, H. J. Hogben, T. Biskup, M. Ahmad, E. Schleicher, S. Weber, C. R. Timmel and P. J. Hore, *Proc. Natl. Acad. Sci.*, 2012, **109**, 4774–4779.
66. F. Lacombat, A. Espagne, N. Dozova, P. Plaza, P. Müller, K. Brettel, S. Franz-Badur and L.-O. Essen, *J. Am. Chem. Soc.*, 2019, **141**, 13394–13409.
67. E. J. Baerends, O. V. Gritsenko and R. van Meer, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16408–16425.
68. P. Müller and M. Ahmad, *J. Biol. Chem.*, 2011, **286**, 21033–21040.
69. L. J. van Wilderen, G. Silkstone, M. Mason, J. J. van Thor and M. T. Wilson, *FEBS Open Bio*, 2015, **5**, 885–892.
70. J. Wang, X. Du, W. Pan, X. Wang and W. Wu, *J. Photochem. Photobiol. C*, 2015, **22**, 84–102.
71. P. Hemmerich, C. Veeger and H. C. S. Wood, *Angew. Chem.*, 1965, **4**, 671–687.
72. R. F. Pauszek, G. Kodali, M. S. U. Siddiqui and R. J. Stanley, *J. Am. Chem. Soc.*, 2016, **138**, 14880–14889.
QM/MM Hessian
\[ \frac{\partial^2 E_{QM/MM}}{\partial R^2} \]

Physical Chemistry Chemical Physics

FAD

\[ +e^- \]

\[ +H^+ \]

FADH\(_2\)

\[ +H^+ \]

FADH\(^-\)

\[ +e^- \]

FADH

\[ +H^+ \]

FADH\(^-\)

\[ +H^+ \]

FAD

Cryptochrome

Exp.

QM/MM

Exp.

Exp.

Exp.

Exp.

Exp.