



# Infrared spectroscopy from electrostatic embedding QM/MM: local normal mode analysis of blue-light-induced infrared spectra of arabidopsis thaliana plant cryptochrome

Miquel Huix-Rotllant, Karno Schwinn, Nicolas Ferré

## ► To cite this version:

Miquel Huix-Rotllant, Karno Schwinn, Nicolas Ferré. Infrared spectroscopy from electrostatic embedding QM/MM: local normal mode analysis of blue-light-induced infrared spectra of arabidopsis thaliana plant cryptochrome. *Physical Chemistry Chemical Physics*, 2021, 23, pp.1666. 10.1039/D0CP06070D . hal-03066310

**HAL Id: hal-03066310**

**<https://hal.science/hal-03066310>**

Submitted on 15 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cite this: DOI: 00.0000/xxxxxxxxxx

# Infrared spectroscopy from electrostatic embedding QM/MM: local normal mode analysis of blue-light-induced infrared spectra of arabidopsis thaliana plant cryptochrome<sup>†</sup>

Miquel Huix-Rotllant<sup>\*a</sup>, Karno Schwinn<sup>a</sup> and Nicolas Ferré<sup>a</sup>

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

Infrared (IR) spectroscopy of biological macromolecules is an undoubtedly valuable tool for analyzing chemical reactions. Currently, there is a lack of theoretical methods able to model successfully and efficiently simulate and interpret the origin of the spectral signatures. Here, we develop a new method for IR vibrational spectroscopy based on analytic second derivatives of electrostatic embedding QM/MM energy, the computation of electric dipole moments with respect to nuclear perturbations and the localization of normal modes. In addition to the spectrum, the method can provide the origin of each peak from clearly identified molecular motions. As proof of concept, we analyze the IR spectra of flavin adenine dinucleotide in water and in *arabidopsis thaliana* cryptochrome protein for four redox forms, and the difference IR spectrum before and after illumination with blue light. We show that the main peaks in the difference spectrum are due to N–H hydrogen out-of-plane motions and hydrogen bendings.

## 1 Introduction

Infrared (IR) spectroscopy is the simplest vibrational spectroscopy technique, which provides undoubtedly relevant information of biological macromolecules (proteins, lipids, nucleic acids, etc.)<sup>1,2</sup> Several levels of sophistication have been developed for extracting more information related to the nuclear motions. Fourier transform IR (FTIR) is currently routinely applied to extract this vibrational information of macromolecules of biological interest.<sup>3</sup> Two-dimensional and higher dimensional IR is providing information relevant to establish the structure and its biological function.<sup>4</sup> Time-resolved IR spectra can provide a full picture of conformational changes during phase transitions or chemical reactions.<sup>5</sup>

Frequently, the IR spectra macromolecules in general and proteins in particular is difficult to interpret, due to the number of spectral signatures appearing in the IR spectra. Proteins are macromolecular chains of aminoacids,

and give rise to nine types of bands, labeled as amides A, B and I–VII.<sup>1,6</sup> When two distinct forms of the protein exist (either due to a chemical reaction or a conformational change induced by an external perturbation), the IR difference spectra becomes useful for ruling out the IR vibrations and concentrating on the different nuclear motions between the two protein forms.<sup>7,8</sup> This has been successfully applied to photoactive proteins, to detect intermediate species of rhodopsins,<sup>9</sup> photoactive yellow proteins,<sup>10</sup> myoglobins,<sup>11</sup> cryptochromes,<sup>12</sup> green fluorescent proteins,<sup>13</sup> photosynthetic proteins,<sup>14</sup> etc.

Despite the simpler spectra that difference IR offers, it is still challenging to interpret the microscopic origin of the vibrational transitions. In this respect, theoretical simulations are particularly adapted to such a purpose.<sup>15,16</sup> Two distinct strategies exist for computing IR spectra, namely, the normal mode analysis (NMA),<sup>17,18</sup> and the Fourier transform of time-dependent (TD) dipole autocorrelation functions extracted from molecular dynamics.<sup>19,20</sup> Both methods have their strengths and weaknesses. While the TD approaches are more easily applied to large macromolecules and extended systems and deals better with anharmonic effects in the IR spectrum, the NMA analysis pro-

<sup>a</sup>Aix-Marseille Univ, CNRS, ICR, Marseille, France. E-mail: miquel.huixrotllant@univ-amu.fr

<sup>†</sup> Electronic Supplementary Information (ESI) available: Theoretical IR spectra in the range between 0 and 4000 cm<sup>−1</sup>. See DOI: 10.1039/cXCP00000x/

vides a more detailed microscopic picture of the vibrational motions involved in each vibrational transition. Currently, NMA analysis is routinely applied to large molecular systems,<sup>21</sup> using all-electron tight-binding approaches,<sup>22,23</sup> purely MM force field second derivatives,<sup>24</sup> and quantum mechanics/molecular mechanics (QM/MM) methods.<sup>25–31</sup>

Here, we develop a new approach for computing the IR spectra based on NMA analysis from a linear scaling formulation of electrostatic embedding QM/MM.<sup>32–35</sup> This method is efficient enough to be routinely applied to extract IR spectra for full proteins. For the sake of interpretation, we localize the normal modes in fragments, following the procedure described in Ref. 36. An alternative method of localization of normal modes by Kraka and coworkers has been recently shown extremely powerful to dissect and analyze the contributions of complex IR spectra.<sup>37–39</sup> As a test case, we simulate and interpret the origin of the IR spectra of flavin adenine dinucleotide (FAD) in water and *arabidopsis thaliana* cryptochrome protein in four different forms. We conclude by analyzing the difference spectrum before and after blue-light illumination of cryptochrome.

## 2 Methods

### 2.1 Electrostatic embedding QM/MM energy based on electrostatic potential fitting charges

The definition of the electrostatic embedding QM/MM energy based on electrostatic potential fitting (ESPF) charges has been presented elsewhere.<sup>35,40</sup> Here, we summarize the main equations for completeness. In QM/MM, the total energy is expressed by

$$E = E_{QM} + E_{MM}. \quad (1)$$

The  $E_{QM}$  here is considered as the solution of a single determinant equation, usually expressed in the form of an eigenvalue problem using an atomic orbital basis,

$$\mathbf{F}_0 \mathbf{C} = \epsilon \mathbf{S} \mathbf{C}, \quad (2)$$

in which  $\mathbf{F}_0$  is the QM Fock operator,  $\mathbf{C}$  is the molecular orbital (MO),  $\epsilon$  is the orbital energy and  $\mathbf{S}$  the atomic orbital overlap. The MM energy in classical force fields is usually written in terms of bonded terms, electrostatic and van der Waals,<sup>41</sup>

$$E_{MM} = E_{bnd} + E_{elec} + E_{vdW}. \quad (3)$$

The electrostatic interaction leads to the main interaction term with the QM system, and is expressed as a classical Coulomb interaction between point charges,

$$E_{elec} = \sum_{P>Q}^N \frac{q_P q_Q}{|\mathbf{R}_P - \mathbf{R}_Q|}, \quad (4)$$

in which  $P, Q$  run over all the atoms of the system  $N$  (both QM and MM). Hereafter, capital indexes  $A, B, \dots$  refer to QM centers,  $I, J, \dots$  to MM centers and  $P, Q, \dots$  unspecified. The point charge  $q_P$  is the partial atomic charges of center  $P$  and  $\mathbf{R}_P$  its Cartesian position. In most classical force fields, the point charges are fixed values. In electrostatic embedding QM/MM, the QM atomic charges are polarizable. In this case, we can split Eq. 4 in two sums

$$E_{elec} = E_{elec,MM} + \Delta E, \quad (5)$$

where  $E_{elec,MM}$  is the purely MM electrostatic term

$$E_{elec,MM} = \sum_{J>I}^{N_{MM}} \frac{q_J q_I}{|\mathbf{R}_J - \mathbf{R}_I|}, \quad (6)$$

and  $\Delta E$  is the QM-MM electrostatic interaction

$$\begin{aligned} \Delta E &= \sum_A^{N_{QM}} q_A(\mathbf{R}) \sum_I^{N_{MM}} \frac{q_I}{|\mathbf{R}_A - \mathbf{R}_I|} \\ &= \sum_A q_A(\mathbf{R}) \phi_A(\mathbf{R}). \end{aligned} \quad (7)$$

Here, we used the short-hand notation for the “external” electrostatic potential on a QM center  $A$  as  $\phi_A$ . The QM partial charges  $q_A(\mathbf{R})$  are functions of the state of the system. They are obtained by tracing the QM density matrix with the quantum atomic charge operator,

$$q_A(\mathbf{R}) = Z_A - Q_A(\mathbf{R}) = Z_A - \sum_{\mu\nu} P_{\mu\nu}(\mathbf{R}) Q_{A;\mu\nu}(\mathbf{R}_{QM}). \quad (8)$$

In this equation,  $Z_A$  is the nuclear charge. It should be noted that the density matrix depend on all the atoms of the system (MM or QM), while the atomic charge operator only depends on the QM positions. This density matrix is obtained by solving Eq. 2 with a Fock operator containing the electrostatic QM-MM interaction,

$$\mathbf{F} = \mathbf{F}_0 + \mathbf{h} = \mathbf{F}_0 - \sum_A \phi_A(\mathbf{R}) Q_A(\mathbf{R}_{QM}). \quad (9)$$

Thus, the dependence of the density matrix on all atoms of the system comes from the presence of the external electrostatic potential in the Fock operator. The interaction energy can thus be obtained by

$$\Delta E = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu} + \sum_A Z_A \phi_A. \quad (10)$$

The QM/MM interaction is defined once the form of the atomic charge operator is given. Here, we employ the charge-conserving electrostatic potential fitting charges.<sup>35</sup> In essence, interaction hamiltonian is modified to conserve

the total charge of the QM fragment,

$$h_{\mu\nu} = \sum_A (\Phi_{av} - \phi_A) Q_{A;\mu\nu} - \Phi_{av} S_{\mu\nu}. \quad (11)$$

Here,  $\Phi_{av} = N_{QM}^{-1} \sum_A \phi_A$  is the average external potential. The charge operator is obtained by fitting the electrostatic integrals on a grid  $\mathbf{V}$ ,

$$\mathbf{Q} = (\mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{T}^\dagger \mathbf{V}, \quad (12)$$

in which  $T_{A;k} = |\mathbf{R}_A - \mathbf{r}_k|^{-1}$  is the electrostatic kernel between the QM center A and the grid point k, and  $\mathbf{V}$  is given by

$$V_{k;\mu\nu} = \int d\mathbf{r} \chi_\mu^*(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \chi_\nu(\mathbf{r}), \quad (13)$$

where  $\chi_\mu$  is the atomic orbital basis. The grid points are usually taken as atom-centered Lebedev spheres.<sup>35,42–44</sup>

## 2.2 Harmonic frequencies from electrostatic embedding ESPF QM/MM

At a minimum energy structure  $\mathbf{R}_0$ , the second-order Taylor expansion of the energy is given by

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \frac{1}{2!} \Delta \mathbf{R}^T \left. \frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{R} \partial \mathbf{R}'} \right|_{\mathbf{R}=\mathbf{R}_0} \Delta \mathbf{R} + \mathcal{O}(\mathbf{R}^3), \quad (14)$$

in which  $\Delta \mathbf{R} = \mathbf{R} - \mathbf{R}_0$  are the Cartesian displacements from the reference geometry. Transforming these in mass-weighted normal coordinates,  $\mathbf{Q} = \mathbf{L} \mathbf{M}^{1/2} \Delta \mathbf{R}$ , in which  $M_{PQ}^{1/2} = \delta_{PQ} m_P^{1/2}$  ( $m_P$  is the atomic mass of center P), and  $\mathbf{L}$  is obtained from the diagonalization of the Hessian matrix

$$\mathbf{L}^\dagger \mathbf{H} \mathbf{L} = \boldsymbol{\omega}^2, \quad (15)$$

where  $\mathbf{H} = \mathbf{M}^{-1/2} (\partial^2 E(\mathbf{R}) / \partial \mathbf{R} \partial \mathbf{R}') \mathbf{M}^{-1/2}$ . The frequencies  $\boldsymbol{\omega}$  can be interpreted as the harmonic frequencies of the normal mode displacement  $\mathbf{Q}$ . Thus, the second derivative at a minimum energy structure contains the essential information to build the IR spectrum. We have recently derived an analytic form for the second derivative of the QM/MM energy (Eq. 1) based on ESPF charges.<sup>32–35</sup> The second derivatives of the QM energy are well established,<sup>45</sup> while the second derivatives of the MM energy are trivially computed. The main caveat in building the QM/MM hessian is building the complete second derivative of the QM/MM interaction energy, Eq. 10. Indeed, this energy depends on the coordinates of QM and MM type of atoms, and thus, the complete Hessian of the interaction energy is given by

$$\frac{\partial^2 \Delta E}{\partial \mathbf{R} \partial \mathbf{R}'} = \begin{bmatrix} \frac{\partial^2 \Delta E}{\partial \mathbf{R} \partial \mathbf{R}'} & \frac{\partial^2 \Delta E}{\partial \mathbf{R} \partial \tilde{\mathbf{R}}} \\ \frac{\partial^2 \Delta E}{\partial \tilde{\mathbf{R}} \partial \mathbf{R}} & \frac{\partial^2 \Delta E}{\partial \tilde{\mathbf{R}} \partial \tilde{\mathbf{R}}} \end{bmatrix}, \quad (16)$$

where  $\mathbf{R}$  and  $\tilde{\mathbf{R}}$  refers to QM and MM atoms respectively. The expression of each block is given by

$$\begin{aligned} \frac{\partial^2 \Delta E}{\partial X_C \partial Y_D} &= \sum_A Z_A \phi_A^{X_C Y_D} + \sum_A Q_A (\Phi_{av} - \phi_A)^{X_C Y_D} \\ &+ \sum_A Q_A^{X_C Y_D} (\Phi_{av} - \phi_A) \\ &+ \sum_A Q_A^{Y_D} (\Phi_{av} - \phi_A)^{X_C} \\ &+ \sum_A Q_A^{X_C} (\Phi_{av} - \phi_A)^{Y_D} \end{aligned} \quad (17)$$

$$\begin{aligned} \frac{\partial^2 \Delta E}{\partial X_C \partial Y_I} = \frac{\partial^2 \Delta E}{\partial Y_I \partial X_C} &= \sum_A Q_A (\Phi_{av} - \phi_A)^{X_C Y_I} \\ &+ \sum_A Q_A^{X_C} (\Phi_{av} - \phi_A)^{Y_I} \\ &+ \sum_A Z_A \phi_A^{X_C Y_I}, \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{\partial^2 \Delta E}{\partial X_I \partial Y_J} &= \sum_A Z_A \phi_A^{X_I Y_J} + \sum_A Q_A (\Phi_{av} - \phi_A)^{X_I Y_J} \\ &+ \sum_A Q_A^{X_I} (\Phi_{av} - \phi_A)^{Y_J}. \end{aligned} \quad (19)$$

Thus, to compute the full QM/MM Hessian, one requires the external potential and its derivatives up to second order with respect to MM and QM atoms. These are trivially computed since they have an analytic classical form. Besides, one requires the derivatives of the atomic charges up to first order with respect to QM and MM atoms and up to second order with respect to QM atoms only. It can be shown that only geometric second derivatives of the atomic charges at fixed MO coefficients are necessary to construct the Hessian.<sup>33–35</sup> In the following, we concentrate on the full derivative of atomic partial charges.

## 2.3 Derivatives of atomic partial charges

The atomic partial charge derivative require the derivative of the density matrix with respect to all nuclear perturbations of the total system. This introduces serious bottlenecks in the computation of the Hessian, that we can easily overcome with the Q-vector method discussed in Ref. 32, to which the reader is referred for the details. For completeness, we summarize the main findings relevant to the computation of the IR method. In section 2.4, we extend the Q-vector method to electric field perturbations.

The derivatives of atomic partial charges defined in Eq.

8 with respect to QM and MM atoms are respectively given by

$$\frac{\partial q_A}{\partial \mathbf{R}} = -\sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial \mathbf{R}} Q_{A;\mu\nu} - \sum_{\mu\nu} P_{\mu\nu} \frac{\partial Q_{A;\mu\nu}}{\partial \mathbf{R}}, \quad (20)$$

and

$$\frac{\partial q_A}{\partial \tilde{\mathbf{R}}} = -\sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial \tilde{\mathbf{R}}} Q_{A;\mu\nu}. \quad (21)$$

Thus, to construct the derivative of atomic partial charges, one requires the derivative of the ESPF operator with respect to QM atoms and the derivative of the density matrix with respect to all atoms of the system. While the derivative of the ESPF charge operator has a trivial expression (see Refs. 32,33,35,40), the derivative of the density matrix with respect to MM atoms can easily become a bottleneck of the computation.<sup>32</sup> The derivative of the density matrix  $\partial \mathbf{P} / \partial \mathbf{R} = (\partial \mathbf{C} / \partial \mathbf{R}) \mathbf{C}^\dagger + \mathbf{C} (\partial \mathbf{C}^\dagger / \partial \mathbf{R})$  require the derivatives of the MO coefficients, which are obtained in a perturbative way from the unperturbed MO coefficients,

$$\frac{\partial C_{\mu i}}{\partial X_C} = \sum_p C_{\mu p} U_{pi}^{X_C}. \quad (22)$$

The rotation matrices  $\mathbf{U}$  are obtained by solving the so-called coupled-perturbed self-consistent field equations.<sup>46</sup> In principle, there are  $3(N_{\text{QM}} + N_{\text{MM}})$  of such type to be solved. Fortunately, one can workaround this scaling by solving an auxiliary set of coupled perturbed equations, the so-called Q-vector equations, and thus only  $4N_{\text{QM}}$  equations have to be solved. In essence, the rotation matrix of MO coefficients with respect to MM atoms is build from  $\mathbf{U}_{pi}^{X_I} = \sum_A \tilde{Q}_{A;pi} (\Phi_{av} - \phi_A)^{X_I}$ , where  $\tilde{Q}$  is the solution of the auxiliary coupled-perturbed equation in which the bare atomic charge operator is used as perturbation.<sup>32</sup>

## 2.4 Induced electric dipole moment derivatives

The infrared spectrum is computed using the well-known expression for the integrated absorption coefficient at 0 K,<sup>47,48</sup>

$$A = \frac{N_A}{6c^2 \epsilon_0 \hbar} \sum_k \omega_k \left| \langle v_i | \frac{\partial \boldsymbol{\mu}}{\partial \Delta X_k} | v_f \rangle \right|^2, \quad (23)$$

in which  $i$  and  $f$  refer to the initial and final state,  $\omega$  is the normal mode harmonic frequency,  $c$  is the light velocity,  $\epsilon_0$  is the permittivity in vacuum. The induced dipole with respect to a normal mode displacement  $\Delta \mathbf{X}$  is given by,

$$\frac{\partial \boldsymbol{\mu}}{\partial \Delta \mathbf{X}_k} = \mathbf{L}_k^\dagger \mathbf{M}^{-1/2} \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{R}}. \quad (24)$$

Here,  $\boldsymbol{\mu}$  is the induced dipole moment from an external electric field. Indeed, in the presence of an external static

electric field  $\vec{\mathcal{E}}$ , the total energy can be expanded in the Taylor series

$$E(\vec{\mathcal{E}}; \mathbf{R}) = E(\mathbf{0}; \mathbf{R}) - \sum_f \mu_f(\mathbf{R}) \vec{\mathcal{E}}_f + \mathcal{O}(\vec{\mathcal{E}}^2) \quad (25)$$

in which  $f = x, y, z$  indicates the direction of the electric field. Many derivations of electric field derivative using analytic derivative method exist in the literature. Here, we follow the derivation of Amos.<sup>49</sup> The induced electric dipole moment can be defined as the derivative with respect to the induced electric field operator  $\mathbf{e}^f$

$$\mu_f(\mathbf{R}) = -\sum_{\mu\nu} P_{\mu\nu} \langle \mu | \mathbf{r}_f | \nu \rangle = \sum_{\mu\nu} P_{\mu\nu}(\mathbf{R}) e_{\mu\nu}^f(\mathbf{R}_{\text{QM}}), \quad (26)$$

in which the atomic orbital density matrix  $\mathbf{P}$  is defined as  $P_{\mu\nu} = \sum_i C_{\mu,i} C_{i,\nu}^*$ , where  $C_{\mu,i}$  is the molecular orbital coefficient. Hereafter, indexes  $i, j, \dots$  are used for occupied orbitals,  $a, b, \dots$  are used for virtual orbitals and  $p, q, \dots$  are unspecified. The dipole moment derivatives with respect to a QM atom are given by

$$\mu_f^{X_C} = 4 \sum_{pi} U_{pi}^{X_C} e_{pi}^f + \sum_{\mu\nu} P_{\mu\nu} e_{\mu\nu}^{fX_C} \quad (27)$$

The derivatives of the induced dipole with respect to an MM atom are simply given by

$$\mu_f^{X_I} = 4 \sum_{pi} U_{pi}^{X_I} e_{pi}^f. \quad (28)$$

This equation would require the solution of a set of coupled-perturbed equations for each MM perturbation, which would make the simulation intractable. Similar to the atomic charge derivatives, we can workaround this by solving an auxiliary set of equations, the  $\tilde{Q}$ -vector technique,<sup>32</sup> to obtain

$$\mu_f^{X_I} = 4 \sum_A \sum_{ai} \tilde{Q}_{A;ai} (\Phi_{av} - \phi_A)^{X_I} e_{ai}^f \quad (29)$$

It is to be noted that this corresponds to the dipole response contribution due to MM motions, but does not lead to the full IR spectrum of the MM subsystem. The main contribution is usually obtained from the Fourier transform of the real-time dipole-dipole autocorrelation function.<sup>50</sup> The presented method here is rather useful to represent the QM frequencies with the steric and electrostatic effects of the MM environment taken into account in addition to the mode-mixing between the QM and MM systems.

## 2.5 Localization of normal modes

In Ref. 36, we described a strategy to decompose normal modes in fragments. The method is based on the fact that submatrices of a positive semi-definite matrix are also positive semi-definite. The total Hessian at a true minimum (Eq. 14) is positive semi-definite, the frequencies of which are interpreted as the characteristic frequencies of vibration. Therefore, the submatrices grouping the atoms of a pre-defined fragment can be interpreted as the local modes of the fragment. The local normal modes can be obtained from the fragment Hessians, defined from the total Hessian as  $[H_F]_{A,B} = \delta_{A \in F, B \in F} \partial^2 E / \partial \mathbf{R}_A \partial \mathbf{R}_B$ . Here, the total Hessian is organized in such a way that the atoms of a fragment are consecutively appearing in the Hessian blocks. From the fragment Hessians, we obtained the local normal modes and frequencies,

$$\mathbf{L}_F^\dagger \mathbf{H}_F \mathbf{L}_F = \boldsymbol{\omega}_F^2. \quad (30)$$

Note that the local Hessian and the local normal modes have the dimension of the total Hessian. The sum  $\mathbf{L}_{loc} = \sum_F^{N_{\text{frag}}} \mathbf{L}_F$  gives a block diagonal transformation for the total Hessian,

$$\mathbf{H}_{loc} = \mathbf{L}_{loc}^\dagger \mathbf{H} \mathbf{L}_{loc}. \quad (31)$$

The transformed local Hessian  $\mathbf{H}_{loc}$  consists of block diagonal matrices for the fragments that are simply the fragment frequencies, and the off-diagonal blocks lead to the fragment vibrational couplings. In addition, the total normal modes can be decomposed as a linear combination of local modes

$$\mathbf{L}_i = \sum_j \alpha_j^i [\mathbf{L}_{loc}]_j. \quad (32)$$

Finally, the local normal modes can be used in Eq. 24 to decompose the IR spectrum in terms of fragments.

## 3 Computational details

The IR spectrum has been obtained by the static normal mode analysis of the geometry corresponding to a minimum of the internal energy, from which a harmonic expansion is done. All calculations have been performed at the B3LYP/6-31G\*/Amber99 level for the QM/MM calculations.<sup>51–54</sup> FAD force field parameters have been taken from Ref. 55. Water solvent effects in the infrared spectra have been included with implicit solvent using a polarizable continuum model (PCM).

For frequency calculations, a scaling factor of 0.96 has been applied to account for anharmonic effects.<sup>56</sup> This scaling factor is obtained for B3LYP/6-31G\* calculations in gas phase. For the current QM/MM computation this is insufficient, and so an extra shift of the QM/MM IR spectra is required to compare to experiments. This results in a reasonable agreement between experiments and

theory: while intensities are in general well reproduced, frequencies need between 20-70  $\text{cm}^{-1}$  shifts. Ideally, one would have to generalize the concept of scaling factors for QM/MM in the spirit proposed by Morokuma and coworkers and reparameterize these for each couple of functional, basis set and force field.<sup>25</sup>

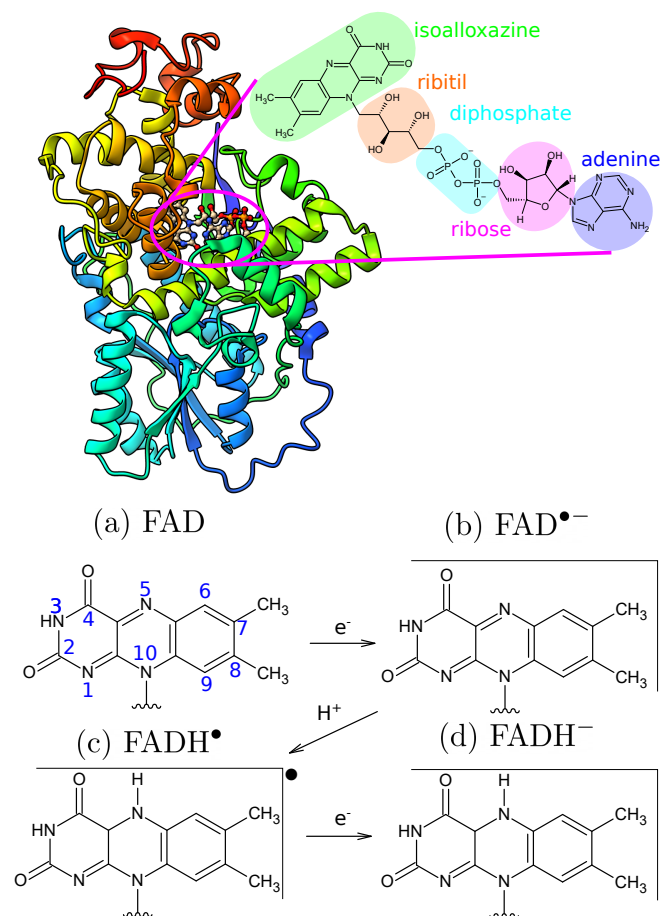
All methods and calculations described here have been performed with a local development version of Gaussian16 (for QM calculations) and Tinker 8.7.1 (for MM calculations), which we have interfaced.<sup>57,58</sup>

## 4 Results and Discussion

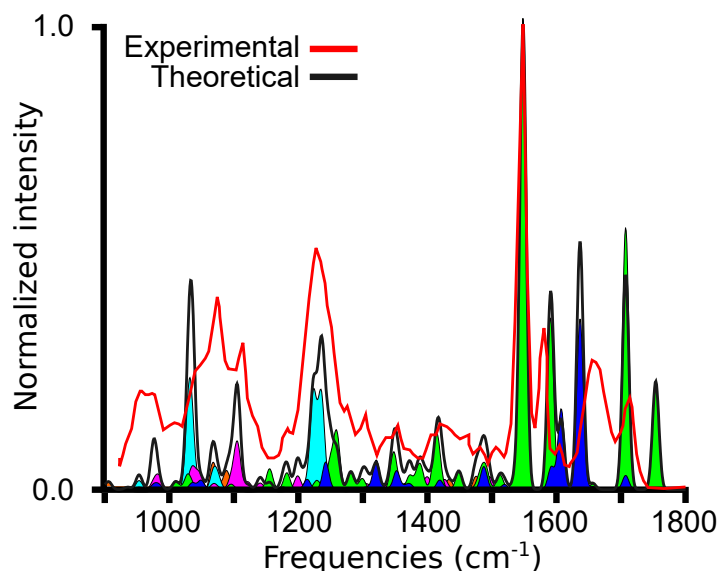
The structures considered in this study are shown in Fig. 1. The QM fragment consists of the flavin adenine dinucleotide, while the rest of the protein is treated at the MM level. For analysis, the FAD vibrations have been analyzed in terms of local normal modes of 5 fragments: isoalloxazine ring (green), ribitol (orange), diphosphate (cyan), ribose (magenta) and adenine (blue). Four redox forms have been considered in this study. Usually, FAD is in its fully oxidized form in the rest state of the protein. Upon blue light absorption, an electron is transferred forming the radical anion  $\text{FAD}^{\bullet-}$ . This is a basic form, that rapidly gets a proton leading to the neutral radical  $\text{FADH}^\bullet$ .<sup>12</sup> A second photon can be absorbed to produce the fully reduced anionic form  $\text{FADH}^-$ . The four redox forms have distinct absorption and emission spectra, which are usually used to characterize the different forms.<sup>34,59–62</sup>

### 4.1 Analysis of FAD spectrum in water

In Fig. 2, we compare the theoretical and experimental IR spectrum of oxidized FAD in water. The theoretical simulations have been performed using an implicit solvent continuum model. Overall, a reasonable agreement is observed between both spectra in the region between 900 and 1800  $\text{cm}^{-1}$ , both in terms of frequencies and intensities. In this frequency window, all vibrations are of in-plane type. From the fragment analysis, two distinct regions can be observed: above 1300  $\text{cm}^{-1}$  the frequencies correspond mainly to isoalloxazine and adenine, while below 1300  $\text{cm}^{-1}$  the vibrations are dominated by the diphosphate and, to a lesser extent, from ribose and ribitol fragments. The band around 1050  $\text{cm}^{-1}$  correspond essentially to two stretching vibrations of  $\text{P}=\text{O}$ , mixed with in-plane ribose, adenine and isoalloxazine, while the band around 1100  $\text{cm}^{-1}$  is a sugar  $\text{C}-\text{C}$  stretching. The band around 1250  $\text{cm}^{-1}$  corresponds to two  $\text{O}=\text{P}-\text{O}$  rocking motions, close in energy to some in-plane motions of adenine and isoalloxazine. The region between 1300-1500  $\text{cm}^{-1}$  correspond essentially to low intensity  $\text{C}-\text{C}$  and  $\text{C}-\text{N}$  in-plane stretchings of adenine and isoalloxazine, mixed respec-



**Fig. 1** Secondary structure of cryptochrome 3 of *arabidopsis thaliana* (PDBID: 2J4D). The close-up image corresponds to the flavin adenine dinucleotide (FAD) with the representation of the different fragments for which the infrared spectra is decomposed. Besides the fully oxidized form (FAD), three reduced forms can exist in protein, a semi-quinonic radical anion ( $\text{FAD}^{\bullet-}$ ), a neutral semi-quinonic radical ( $\text{FADH}^{\bullet}$ ), and the fully reduced hydroquinone anion ( $\text{FADH}^-$ ).



**Fig. 2** Comparison of the experimental (red) and theoretical (black) infrared spectra of FAD in water in the range 900-1800  $\text{cm}^{-1}$  (for the full spectrum see the Supporting Information). The theoretical IR spectrum has been decomposed in fragments, shown as colored areas as described in Fig. 1. The intensities of both experimental and theoretical spectra have been normalized to the one for the most intense peak. For the sake of comparison, the theoretical spectra has been downscaled 23.99  $\text{cm}^{-1}$ . Experimental data has been extracted from Ref. 63.

tively with ribose and ribitol. The main vibrational peak at 1572  $\text{cm}^{-1}$  is localized exclusively on isoalloxazine. It correspond to  $\text{C}_7\text{--C}_8$  and  $\text{C}_{5a}\text{--C}_{9a}$  stretchings combined with  $\text{N}_{10}\text{--C}_{10a}$  stretching. A minor contribution of  $\text{N}_5\text{--C}_{4a}$  is observed. A similar less-intense vibration is found at 1614  $\text{cm}^{-1}$ , in which the  $\text{N}_5\text{--C}_{4a}$  stretching is major and  $\text{N}_{10}\text{--C}_{10a}$  stretching is minor. These vibrations are almost isoenergetic and resonant to several  $\text{C--N}$  stretchings of the adenine ring. The most intense peak of adenine is found at 1660  $\text{cm}^{-1}$ , and corresponds to a  $\text{NH}_2$  bending combined with  $\text{C--NH}_2$  stretching. Experimentally, this is observed as a broader band, that simulations were not able to reproduce. Finally, carbonyl stretchings  $\text{C}_2=\text{O}$  and  $\text{C}_4=\text{O}$  localized on isoalloxazine are found between 1700-1800  $\text{cm}^{-1}$ .

## 4.2 IR spectra of FAD and its reduced forms in protein

The protein has a strong impact in the IR spectrum of FAD. First, the steric effects of the protein constraint FAD to adopt a closed conformation, in which FAD and adenine are face to face. Second, the electrostatic effects change the characteristic frequencies of vibrations as well as the intensities, especially due to the resonance between the prosthetic group and the protein aminoacids. In this section, we analyze the IR spectrum of FAD and its three reduced forms

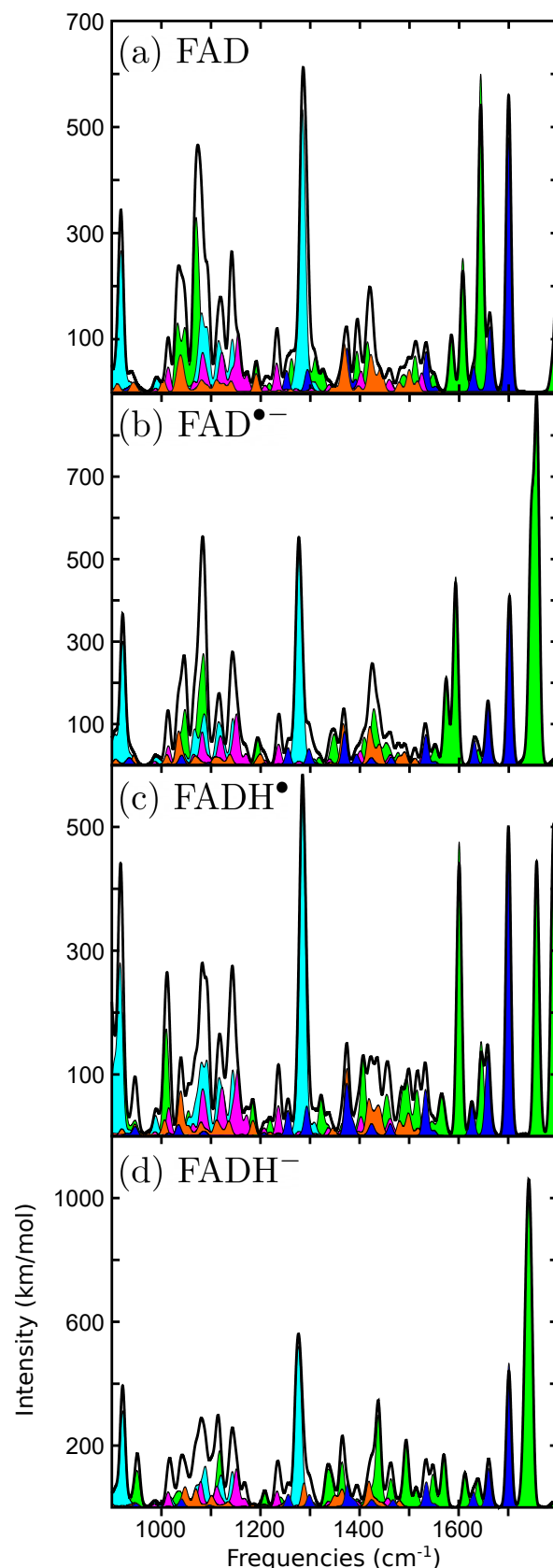
in cryptochrome protein. These results are shown in Fig. 3. In the following, we discuss each spectrum separately.

#### 4.2.1 FAD

The oxidized FAD spectrum in protein is distinct from the spectrum in water. The intensity of isoalloxazine and adenine peaks are diminished, and now the main peak of absorption corresponds to the phosphate. In addition, a stronger mixture of the frequencies in the region from 900  $\text{cm}^{-1}$  to 1500  $\text{cm}^{-1}$  is observed. The peak at 918  $\text{cm}^{-1}$  corresponds to a P–O–P stretching of the bridging oxygen between the two phosphates, resonant with a close by crystallographic water. In the region between 1000  $\text{cm}^{-1}$  and 1100  $\text{cm}^{-1}$  a strong mixture of vibrations is observed. The intense peaks at 1031, 1047, 1067, 1068 and 1074  $\text{cm}^{-1}$  correspond to five vibrations  $\text{N}_3\text{--H}$  hydrogen out-of-plane motion mixed with  $\text{N}_1\text{--C}_2$  stretching are observed, strongly mixed with ribitol stretchings. Around 1100  $\text{cm}^{-1}$ , the P=O stretchings are found, 50  $\text{cm}^{-1}$  upscaled with respect to the IR spectrum in water. These are now strongly mixed with ribose and ribitol vibrations. The main IR peak corresponds to the two O=P–O rocking motions found at 1282 and 1287  $\text{cm}^{-1}$ . These are again upscaled with respect to spectra in water by 30  $\text{cm}^{-1}$ . The two vibrations at 1584 and 1607  $\text{cm}^{-1}$  correspond to the  $\text{C}_7\text{--C}_8$  and  $\text{C}_{5a}\text{--C}_{9a}$  stretchings combined in this case, the  $\text{N}_5\text{--C}_{4a}$  is combined with  $\text{N}_1\text{--C}_{10a}$ . These C–N stretchings are found in two degenerate vibrations at 1643  $\text{cm}^{-1}$ . The higher part of the spectrum correspond to the  $\text{NH}_2$  bending vibrations found at 1661 and 1700  $\text{cm}^{-1}$ .

#### 4.2.2 $\text{FAD}^{\bullet-}$

The anionic radical is obtained by an electron transfer to FAD. In protein, the structure is planar and close to the oxidized FAD and the radical is delocalized all over the isoalloxazine ring.<sup>34,59</sup> The lower peak is found at 921  $\text{cm}^{-1}$  and corresponds to the P–O–P stretching, slightly upscaled with respect to oxidized FAD. In the region 1000–1200  $\text{cm}^{-1}$  five distinct peaks are observed. At 1013  $\text{cm}^{-1}$  a sugar stretching is observed. Between 1032–1050  $\text{cm}^{-1}$  an isoalloxazine  $\text{--CH}_3$  bending motions mixed with ribitol stretchings. The other three peaks in this region correspond to  $\text{N}_3\text{--H}$  hydrogen out-of-plane motion mixed with  $\text{N}_1\text{--C}_2$  stretching found between 1050 and 1085  $\text{cm}^{-1}$  as for FAD. These vibrations mix also with the P=O stretchings. The intense peak at 1277 and 1283  $\text{cm}^{-1}$  correspond to two O=P–O rocking, and is found at the same position as oxidized FAD. The region between 1300–1500  $\text{cm}^{-1}$  corresponds to several stretchings between isoalloxazine mixed with ribitol, and adenine mixed with sugar. The isoalloxazine frequencies found at 1573  $\text{cm}^{-1}$  and 1593



**Fig. 3** Infrared spectrum of FAD and its reduced forms in protein in the range 900–1800  $\text{cm}^{-1}$  (for the full spectrum see the Supporting Information). The black line is the total IR spectrum, while the shaded areas correspond to the different FAD fragments as depicted in Fig. 1. For this spectra, no shift of frequencies was applied.



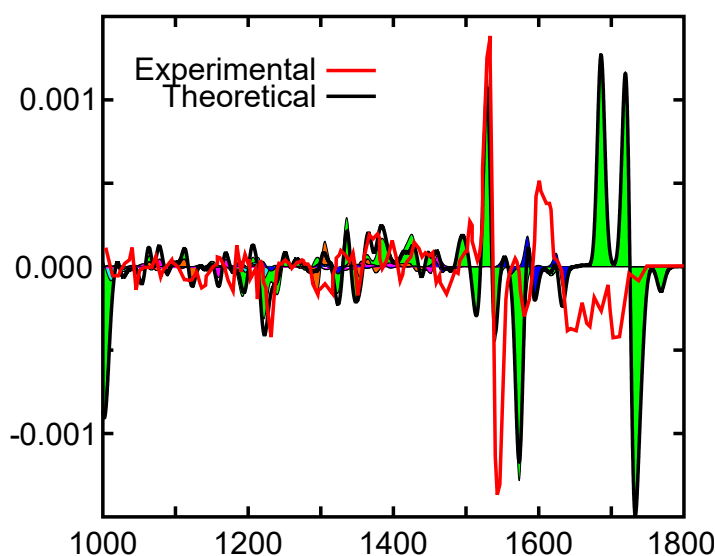
$\text{cm}^{-1}$  correspond to the  $\text{C}_7\text{--C}_8$  and  $\text{C}_{5a}\text{--C}_{9a}$  stretchings combined with  $\text{N}_1\text{--C}_{10a}$  and to lesser extend the  $\text{N}_5\text{--C}_{4a}$  stretching. The adenine  $\text{NH}_2$  bending vibrations are found at 1657 and 1701  $\text{cm}^{-1}$ . The most intense peak at 1745  $\text{cm}^{-1}$  correspond to the two  $\text{C=O}$  vibrations of isoalloxazine, which are about 50  $\text{cm}^{-1}$  downscaled with respect to the oxidized form.

#### 4.2.3 FADH<sup>•</sup>

The neutral radical is obtained by a proton transfer to  $\text{FAD}^{\bullet-}$  in position  $\text{N}_5$ . The lowest IR peak corresponds to the  $\text{P--O--P}$  stretching, and is found at 916  $\text{cm}^{-1}$ , closer to the oxidized FAD peak. Nearby, a small peak corresponding  $\text{N}_3\text{--H}$  hydrogen out-of-plane motion appears at 920  $\text{cm}^{-1}$ , about 80  $\text{cm}^{-1}$  downscaled with respect to FAD or  $\text{FAD}^{\bullet-}$ . In the region of 1010  $\text{cm}^{-1}$ , the  $\text{N}_5\text{--H}$  hydrogen out-of-plane motion. This vibration was not present in the previous forms, since the  $\text{N}_5$  was deprotonated. The next four peaks in the IR spectrum correspond to a ribitol  $\text{C--O}$  stretchings found at 1038 and 1040  $\text{cm}^{-1}$ , a ribose  $\text{C--O}$  stretching at 1082  $\text{cm}^{-1}$  overlapping with a mixed  $\text{P=O}$  stretching with ribitol and isoalloxazine  $\text{CH}_3$  motions found at 1090  $\text{cm}^{-1}$ , with a similar motion found at 1144  $\text{cm}^{-1}$ . These are slightly more spread than for FAD and  $\text{FAD}^{\bullet-}$ . The main peak of absorption is the two  $\text{O=P--O}$  rocking motions found at 1283 and 1286  $\text{cm}^{-1}$ . This peak was equivalently found in the same place and intensity for FAD and  $\text{FAD}^{\bullet-}$ . In the region between 1300 and 1600  $\text{cm}^{-1}$  several peaks of equivalent intensity appear. Around 1600  $\text{cm}^{-1}$ , two peaks were appearing for FAD and  $\text{FAD}^{\bullet-}$  corresponding to several resonant vibrations involving  $\text{N}_5\text{--C}_{4a}$  and  $\text{N}_1\text{--C}_{10a}$ . Due to the presence of a proton on  $\text{N}_5$ , this resonance is broken. An intense peak at 1600  $\text{cm}^{-1}$  corresponds to a main contribution of the  $\text{N}_1\text{--C}_{10a}$  mixed with an  $\text{N}_5\text{--H}$  bending, whereas the lower intensity peak at 1569  $\text{cm}^{-1}$  corresponds to the main  $\text{N}_5\text{--H}$  bending motion. The  $\text{N}_5\text{--C}_{4a}$  stretching appears at 1644  $\text{cm}^{-1}$ , almost degenerate to the  $\text{NH}_2$  bending motions of adenine which are found at 1657 and 1700  $\text{cm}^{-1}$ . The two  $\text{C=O}$  stretchings of isoalloxazine are appearing now at 1755 and 1790  $\text{cm}^{-1}$ , with a small contribution from the  $\text{N}_5\text{--H}$  bending.

#### 4.2.4 FADH<sup>-</sup>

The fully reduced anion is obtained by an electron transfer to  $\text{FADH}^{\bullet}$ . The main spectral features of the previous form are still found in the IR spectrum of  $\text{FADH}^-$ . Indeed, the lowest absorbing peak is the  $\text{P--O--P}$  stretching of the bridging oxygen found at 921  $\text{cm}^{-1}$ . The  $\text{N}_5\text{--H}$  hydrogen out-of-plane motion is found at 952  $\text{cm}^{-1}$ , that is, 50  $\text{cm}^{-1}$  red shifted with respect to the  $\text{FADH}^{\bullet}$ . The  $\text{N}_3\text{--H}$  hy-



**Fig. 4** Comparison of the theoretical and experimental difference IR spectra between  $\text{FADH}^{\bullet}$  and FAD in the range 1000-1800  $\text{cm}^{-1}$  (for the full spectrum see the Supporting Information). For the sake of comparison, theoretical difference spectrum has been computed first with the unscaled spectra of Fig. 3 and subsequently downscaled by 70  $\text{cm}^{-1}$  and scaled to match the intensity of the largest positive experimental peak.

drogen out-of-plane motion is in this case found between 1038 and 1115  $\text{cm}^{-1}$ , similar to oxidized FAD. In this region between 1000 and 1100  $\text{cm}^{-1}$ , sugar  $\text{C--O}$  stretchings as well as  $\text{O=P--O}$  symmetric stretchings are found, with less intensity than in the previous forms. The two  $\text{O=P--O}$  rocking motions are found at 1271, 1273, 1277 and 1281  $\text{cm}^{-1}$ , similar to the previous redox forms. The first  $\text{N}_5\text{--C}_{4a}$  stretching mixed with  $\text{N}_{10}\text{--C}_{10a}$  stretching and  $\text{N}_5\text{--H}$  bending is found at 1365  $\text{cm}^{-1}$ , about 150  $\text{cm}^{-1}$  downscaled than the previous forms. In this region, they are strongly mixed with ribitol stretchings. The main  $\text{N}_5\text{--H}$  bending motion is found at 1438  $\text{cm}^{-1}$ , about a 100  $\text{cm}^{-1}$  downscaled with respect to  $\text{FADH}^{\bullet}$ . Similar motions are found at 1463 and 1547  $\text{cm}^{-1}$ , in this case mixed with  $\text{CH}_3$  hydrogen bending motions. A peak corresponding to a synchronous hydrogen bending motions in plane of  $\text{C}_6\text{--H}$  and  $\text{C}_9\text{--H}$  are observed at 1570  $\text{cm}^{-1}$ , and a  $\text{C}_{4a}\text{--C}_{10a}$ ,  $\text{C}_{5a}\text{--C}_{9a}$  and  $\text{C}_7\text{--C}_8$  synchronous stretching is found at 1612  $\text{cm}^{-1}$ . These two peaks were not appearing in the previous forms. The adenine  $\text{NH}_2$  bending is found at 1701  $\text{cm}^{-1}$ , while the  $\text{C=O}$  stretchings are found at 1738 and 1745  $\text{cm}^{-1}$ .

### 4.3 Difference spectra between bright and dark cryptochrome

The IR difference spectra are frequently used in literature to analyze two states of a protein, usually after a reac-

tion has occurred. In Ref. 12, Ahmad and coworkers analyzed the IR difference spectrum between the bright and dark forms of cryptochrome 1. The dark state corresponds essentially to the oxidized form of FAD, while the bright state corresponds to FADH<sup>•</sup>. From the spectra presented in Fig. 3, we can compute the theoretical difference IR spectra, as shown in Fig. 4. Overall, a reasonable agreement is observed between experiments and theory. The difference spectrum is dominated by isoalloxazine vibrations, although minor contributions are also observed from adenine in the region of 1600 cm<sup>-1</sup>, ribitol and ribose. The diphosphate vibrations are almost completely hindered out in this energy window. The difference spectrum peak at 1220 cm<sup>-1</sup> can be attributed to the N<sub>5</sub>–H hydrogen out-of-plane motion, while the main peak difference located at 1550 cm<sup>-1</sup> can thus be interpreted as the N<sub>5</sub>–H bending motion, which changes the intensity and position of the C<sub>5a</sub>–N<sub>5</sub> and C<sub>9a</sub>–N<sub>10</sub> stretchings.

## 5 Conclusion

We have presented a strategy to efficiently compute the IR spectrum at the ESPF-QM/MM level, by extending the Q-vector method coupled-perturbed equations for nuclear perturbations to electric dipoles. This allows an efficient computation of the induced dipole moments on the QM atoms in addition to the response dipole due to MM perturbations.

To illustrate the method, we have analyzed the IR spectrum of oxidized FAD in water, as well as oxidized and reduced forms in protein. In addition, we have computed the difference IR spectrum between the bright and dark structures of cryptochrome. Overall, a reasonable agreement is observed between theory and experiments. We have analyzed the origin of the peaks in terms of fragment local normal modes, which allowed us to effectively determine the origin of each peak in the spectrum. We show that the difference spectrum only contains signals of the isoalloxazine ring.

In conclusion, the ESPF QM/MM method is a powerful method to obtain vibrational spectroscopies in protein. In the future, the method described here will be extended to other spectroscopies like static non-resonant Raman or vibrational circular dichroism.

## 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.

## Author contributions

MHR designed the research. MHR, KS and NF developed and implemented the method. MHR performed the calculations, analyzed the data and wrote the article. All authors have participated in the discussions.

## Notes and references

- 1 A. Barth and C. Zscherp, *Quarterly Reviews of Biophysics*, 2002, **35**, 369–430.
- 2 B. H. Stuart, in *Infrared Spectroscopy of Biological Applications: An Overview*, American Cancer Society, 2012, ch. , pp. 1–17.
- 3 H. H. Mantsch and R. N. McElhaney, *J. Mol. Struct.*, 1990, **217**, 347–362.
- 4 N. T. Hunt, *Chem. Soc. Rev.*, 2009, **38**, 1837–1848.
- 5 M. L. Groot and R. Van Grondelle, in *Femtosecond Time-Resolved Infrared Spectroscopy*, Springer Netherlands, Dordrecht, 2008, pp. 191–200.
- 6 A. Barth, *Biochim. Biophys. Acta - Bioenerg.*, 2007, **1767**, 1073 – 1101.
- 7 J. Grdadolnik, *Vibrational Spectroscopy*, 2003, **31**, 279 – 288.
- 8 J. Grdadolnik and Y. Maréchal, *Vibrational Spectroscopy*, 2003, **31**, 289 – 294.
- 9 K. A. Bagley, L. Eisenstein, T. G. Ebrey and M. Tsuda, *Biochem.*, 1989, **28**, 3366–3373.
- 10 Y. Imamoto, Y. Shirahige, F. Tokunaga, T. Kinoshita, K. Yoshihara and M. Kataoka, *Biochem.*, 2001, **40**, 8997–9004.
- 11 D. Braunstein, K. Chu, K. Egeberg, H. Frauenfelder, J. Mourant, G. Nienhaus, P. Ormos, S. Sligar, B. Springer and R. Young, *Biophys. J.*, 1993, **65**, 2447 – 2454.
- 12 P. Müller and M. Ahmad, *J. Biol. Chem.*, 2011, **286**, 21033–21040.
- 13 J. J. van Thor, G. Y. Georgiev, M. Towrie and J. T. Sage, *J. Biol. Chem.*, 2005, **280**, 33652–33659.
- 14 M. Di Donato and M. L. Groot, *Biochim. Biophys. Acta (BBA) - Bioenergetics*, 2015, **1847**, 2 – 11.
- 15 N. A. Besley, *Philos. Trans. Royal Soc. A*, 2007, **365**, 2799–2812.
- 16 K. B. Beć, J. Grabska, C. W. Huck and Y. Ozaki, *Molecular Spectroscopy*, Wiley-VCH Verlag GmbH & Co. KGaA, 2019, pp. 353–388.
- 17 A. P. Scott and L. Radom, *J. Phys. Chem.*, 1996, **100**, 16502–16513.
- 18 *Normal Mode Analysis*, ed. Q. Cui and I. Bahar, Chap-

- man and Hall/CRC, 2006.
- 19 M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.
  - 20 P.-A. Cazade, J. Huang, J. Yosa, J. J. Szymczak and M. Meuwly, *Int. Rev. Phys. Chem.*, 2012, **31**, 235–264.
  - 21 J. Neugebauer, M. Reiher, C. Kind and B. A. Hess, *J. Comput. Chem.*, 2002, **23**, 895–910.
  - 22 H. A. Witek and K. Morokuma, *J. Comput. Chem.*, 2004, **25**, 1858–1864.
  - 23 H. Yu and Q. Cui, *J. Chem. Phys.*, 2007, **127**, 234504.
  - 24 M. Levitt, C. Sander and P. S. Stern, *Int. J. Quant. Chem.*, 1983, **24**, 181–199.
  - 25 S. Dapprich, I. Komáromi, K. Byun, K. Morokuma and F. J. J. *Molec. Struct.: THEOCHEM*, 1999, **461–462**, 1–21.
  - 26 Q. Cui and M. Karplus, *J. Chem. Phys.*, 2000, **112**, 1133–1149.
  - 27 Q. Cui and M. Karplus, *J. Phys. Chem. B*, 2000, **104**, 3721–3743.
  - 28 L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding and K. Morokuma, *Chem. Rev.*, 2015, **115**, 5678–5796.
  - 29 T. Giovannini, L. Grazioli, M. Ambrosetti and C. Cappelli, *J. Chem. Theor. Comput.*, 2019, **15**, 5495–5507.
  - 30 M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2020, **22**, 14433–14448.
  - 31 U. N. Morzan, D. J. Alonso de Armiño, N. O. Foglia, F. Ramírez, M. C. González Lebrero, D. A. Scherlis and D. A. Estrin, *Chem. Rev.*, 2018, **118**, 4071–4113.
  - 32 K. Schwinn, N. Ferré and M. Huix-Rotllant, *J. Chem. Phys.*, 2019, **151**, 041102.
  - 33 K. Schwinn, N. Ferré and M. Huix-Rotllant, *J. Chem. Theor. Comput.*, 2020, **16**, 3816–3824.
  - 34 K. Schwinn, N. Ferré and M. Huix-Rotllant, *Phys. Chem. Chem. Phys.*, 2020, **22**, 12447–12455.
  - 35 M. Huix-Rotllant and N. Ferré, *J. Chem. Theor. Comput.*, 2020, (accepted, doi: 10.1021/acs.jctc.0c01075).
  - 36 M. Huix-Rotllant and N. Ferré, *J. Chem. Theor. Comput.*, 2016, **12**, 4768–4777.
  - 37 Y. Tao, C. Tian, N. Verma, W. Zou, C. Wang, D. Cremer and E. Kraka, *J. Chem. Theor. Comput.*, 2018, **14**, 2558–2569.
  - 38 E. Kraka, W. Zou and Y. Tao, *WIREs Computat. Molec. Sci.*, 2020, **10**, e1480.
  - 39 *J. Mol. Mod.*, 2020, **26**, 281.
  - 40 N. Ferré and J. G. Ángyán, *Chem. Phys. Lett.*, 2002, **356**, 331–339.
  - 41 A. D. Mackerell, *J. Comput. Chem.*, 2004, **25**, 1584–1604.
  - 42 Z. C. Holden, R. M. Richard and J. M. Herbert, *J. Chem. Phys.*, 2013, **139**, 244108.
  - 43 Z. C. Holden, B. Rana and J. M. Herbert, *J. Chem. Phys.*, 2019, **150**, 144115.
  - 44 Z. C. Holden, *Long-range effects in QM/MM calculations: Ewald summation in non-minimal basis sets*, 2015, Ph.D. thesis, Ohio State University, Columbus, OH.
  - 45 N. C. Handy, D. J. Tozer, G. J. Laming, C. W. Murray and R. D. Amos, *Isr. J. Chem.*, 1993, **33**, 331–344.
  - 46 J. A. Pople, R. Krishnan, H. B. Schlegel and J. S. Binkley, *Int. J. Quant. Chem.*, 1979, **16**, 225–241.
  - 47 E. Wilson, *Molecular vibrations : the theory of infrared and Raman vibrational spectra*, Dover Publications, New York, 1980.
  - 48 P. Norman, K. Ruud and T. Saue, *Principles and practices of molecular properties : theory, modeling and simulations*, John Wiley & Sons, Hoboken, NJ, 2018.
  - 49 R. Amos, *Chem. Phys. Lett.*, 1986, **124**, 376–381.
  - 50 J. P. Boon and M. Köhler, *J. Chem. Phys.*, 1969, **51**, 3681–3688.
  - 51 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
  - 52 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
  - 53 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
  - 54 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. Ghoreishi, 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.
  - 55 J. Antony, D. Medvedev and A. Stuchebrukhov, *J. Am. Chem. Soc.*, 2000, **122**, 1057.
  - 56 *NIST Standard Reference Database*, August 2020, Number 101 Release 21, Ed. Russell D. Johnson III, <http://cccbdb.nist.gov/>.
  - 57 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, 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. Lipparini, 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, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision A.01*, 2016, Gaussian Inc. Wallingford CT.
- 58 J. W. Ponder, *TINKER: Software tools for molecular design, version 8.7.1*, 2019, <http://dasher.wustl.edu/tinker/>.
- 59 P. Mondal, K. Schwinn and M. Huix-Rotllant, *J. Photochem. Photobiol. A: Chem.*, 2020, **387**, 112164.
- 60 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.
- 61 Y.-T. Kao, C. Tan, S.-H. Song, N. Öztürk, J. Li, L. Wang, A. Sancar and D. Zhong, *J. Am. Chem. Soc.*, 2008, **130**, 7695–7701.
- 62 M. P. Kabir, Y. Orozco-Gonzalez and S. Gozem, *Phys. Chem. Chem. Phys.*, 2019, **21**, 16526–16537.
- 63 M. Spexard, D. Immeln, C. Thöing and T. Kottke, *Vib. Spectrosc.*, 2011, **57**, 282 – 287.