



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

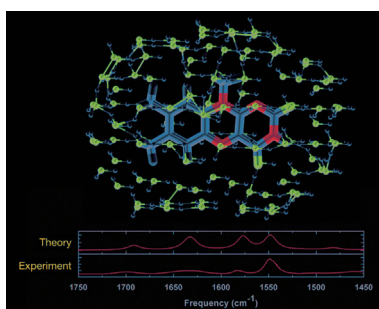
The effect of hydrogen-bonding on flavin's infrared absorption spectrum

Mohammad Pabel Kabir^a, Yoelvis Orozco-Gonzalez^a, Gary Hastings^{b,c,*}, Samer Gozem^{a,*}^a Department of Chemistry, Georgia State University, Atlanta, GA 30302, United States^b Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30302, United States^c Center for Nano-Optics, Georgia State University, Atlanta, GA 30302, United States

HIGHLIGHTS

- Hydrogen bonding induced modifications to flavin's IR spectrum are calculated.
- Using a new computational tool, we assess potential flavin-water interactions.
- Calculations using cluster and continuum models are compared with experimental data.
- Calculations provide insight into flavin intermolecular interactions in proteins.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 17 February 2021

Received in revised form 17 June 2021

Accepted 21 June 2021

Available online 26 June 2021

Keywords:

Flavin

FMN

FAD

Hydrogen bonding

Infrared spectroscopy

FTIR

ABSTRACT

Cluster and continuum solvation computational models are employed to model the effect of hydrogen bonding interactions on the vibrational modes of lumiflavin. Calculated spectra were compared to experimental Fourier-transform infrared (FTIR) spectra in the diagnostic 1450–1800 cm⁻¹ range, where intense $\nu_{C=C}$, $\nu_{C=N}$, $\nu_{C_2=O}$, and $\nu_{C_4=O}$ stretching modes of flavin's isoalloxazine ring are found. Local mode analysis is used to describe the strength of hydrogen-bonding in cluster models. The computations indicate that $\nu_{C=C}$ and $\nu_{C=N}$ mode frequencies are relatively insensitive to intermolecular interactions while the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ modes are sensitive to direct (and also indirect for $\nu_{C_2=O}$) hydrogen-bonding interactions. Although flavin is neutral, basis sets without the diffuse functions provide incorrect relative frequencies and intensities. The 6-31+G* basis set is found to be adequate for this system, and there is limited benefit to considering larger basis sets. Calculated vibrational mode frequencies agree with experimentally determined frequencies in solution when cluster models with multiple water molecules are used. Accurate simulation of relative FTIR band intensities, on the other hand, requires a continuum (or possibly quantum mechanical/molecular mechanical) model that accounts for long-range electrostatic effects. Finally, an experimental peak at ca. 1624 cm⁻¹ that is typically assigned to the $\nu_{C_2=O}$ vibrational stretching mode has a complicated shape that suggests multiple underlying contributions. Our calculations show that this band has contributions from both the C₆–C₇ and C₂ = O stretching vibrations.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Flavin-dependent proteins are found in all kingdoms of life, where they are responsible for a wide range of biological processes [1–4]. Flavoproteins typically bind either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), both of which are derivatives of riboflavin (vitamin B₂). Flavin acts as a redox agent

* Corresponding authors at: Department of Chemistry, Georgia State University, Atlanta, GA 30302, United States (S. Gozem). Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30302, United States (G. Hastings).

E-mail addresses: ghastings@gsu.edu (G. Hastings), sgozem@gsu.edu (S. Gozem).

in many enzymatic reactions [5,6], although there are also multiple instances where flavin catalyzes reactions with no redox change [7,8]. Flavin has also been implicated as the active chromophore in a number of photoreceptors and light-responsive enzymes [9–13].

The physical and chemical changes accompanying catalytic and photochemical processes in flavoproteins and the resulting changes to the microenvironment such as H-bonding interactions could be probed using Fourier-transform infrared (FTIR) spectroscopy. However, since proteins are macromolecules with multiple functional groups, their congested FTIR spectra make it difficult to isolate specific interactions. Flavins are multi-redox, photoactive compounds that have multiple protonation states, making it possible to probe changes in their FTIR spectra as a result of some chemical or physical change using FTIR difference spectroscopy (DS) [14,15]. FTIR DS techniques [16,17] are sensitive to small changes in intra- and inter-molecular interactions that may occur during any process and are therefore excellent tools to probe flavin redox chemistry or photochemistry [18].

The interpretation of experimental FTIR DS in proteins is not always straightforward, in part because vibrational frequencies and intensities are affected by electrostatic and H-bonding interactions within the protein. On the other hand, it is precisely these changes that make redox molecules like flavin useful probes of their protein microenvironment. Interpretation of experimental FTIR DS requires knowledge of how electrostatics and H-bonding modify flavin's FTIR spectrum. This knowledge can be obtained from computational modeling.

One approach to interpreting how specific interactions modulate the FTIR spectra of a protein-bound molecule (e.g., flavin) is with hybrid quantum mechanical / molecular mechanical (QM/MM) methods [19–22]. Several QM/MM methods and tools have been developed to simulate FTIR DS of protein-bound molecules more efficiently and to help assign peaks in experimental spectra to specific molecular modes [23–26]. There are also some recent instances where such QM/MM methods were employed to simulate the FTIR spectra of flavins in different proteins [26–30]. However, while such calculations are specific to the protein being modeled, it is also desirable to look at a straightforward model system to try to understand how nearby interactions at different positions influence flavin's vibrational spectra.

The goal of this study is to answer the following two questions:

1. Can explicit solvent or continuum solvation models reproduce the relative frequencies and intensities of prominent peaks in flavin's IR spectrum in an aqueous solution?
2. How do specific non-bonding interactions with nearby polar molecules affect the vibrational frequencies and intensities of flavin's isoalloxazine ring?

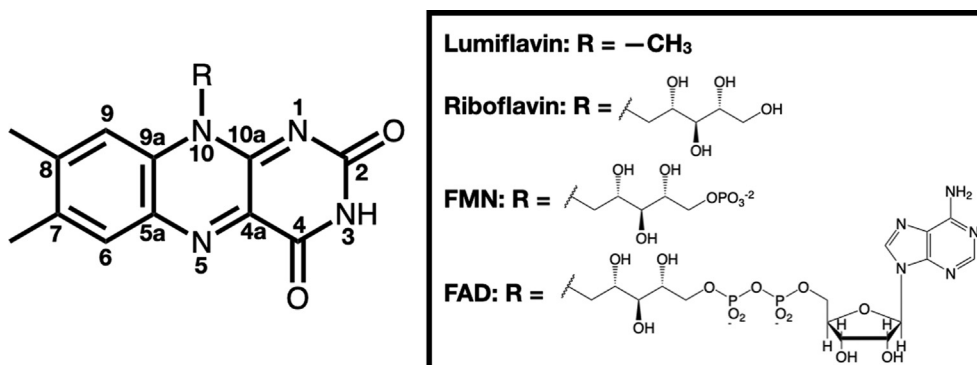
We will focus on the four most prominent vibrational modes of flavin's isoalloxazine ring, which all appear in the 1450–1800 cm^{-1} region for flavin in solution. Several experimental FTIR spectra for FMN or FAD in solution are shown in Fig. S1 in the supplementary information (SI) [31–37]. The four bands are observed near 1548 cm^{-1} ($\nu_{\text{C}=\text{C}}$), 1580 cm^{-1} ($\nu_{\text{C}=\text{N}}$), 1640 cm^{-1} ($\nu_{\text{C}_2=\text{O}}$), and 1700 cm^{-1} ($\nu_{\text{C}_4=\text{O}}$) for samples in D_2O . All experimental spectra in Fig. S1 display bands within 5 cm^{-1} of these indicated values [31–34], with the exact frequency depending on whether experiments employ FMN or FAD or whether the samples are in D_2O or H_2O (the $\nu_{\text{C}_2=\text{O}}$ and $\nu_{\text{C}_4=\text{O}}$ bands are upshifted in H_2O). The corresponding normal mode atomic motions are shown in Fig. S2 of the SI (the suggested assignments refer to the dominant vibrational mode). Since we focus on vibrational modes of the isoalloxazine ring of flavin, we employ lumiflavin as a model system (see Scheme 1 for structures).

2. Methodology

The main goal of this study is to determine how intermolecular interactions between flavin and other polar molecules influence flavin's vibrational mode frequencies. As a starting point to disentangle the effect of such interactions, we modeled the FTIR spectrum of lumiflavin interacting with one water molecule placed at different positions around the isoalloxazine ring. To find such potential interactions, we used the approach outlined in Fig. 1.

The approach in Fig. 1 is related to previously reported Electrostatic Spectral Tuning Maps (ESTM) [38,39], and employs the open-source python library pyvdwsurface [40]. In this work, the approach is used to search for local geometry minima that represent different interactions between lumiflavin and water molecules. While 92 structures were optimized, many of these structures were similar, indicating there are only a few local minima for isoalloxazine-water interactions. Specifically, eight positions were identified where water had a tendency to go. Waters placed in the hydrophilic side of lumiflavin ended up in one of the five positions shown in orange in Fig. 2, while waters placed near the hydrophobic side of lumiflavin ended up in one of the three positions shown in blue in Fig. 2.

In addition to those eight structures that each included a single water molecule, we optimized several cluster models with multiple water molecules. Specifically, we chose a structure with two water molecules both H-bonded to one carbonyl (green structures in Fig. 2) and three cluster models with more than two water molecules (magenta structures in Fig. 2). The cluster models include a model with all waters near the hydrophilic side (6W; a sixth water was added near N_3 to stabilize the two nearby water molecules) as well as two models with waters both near the hydrophilic and hydrophobic sides of lumiflavin (8W and 9W).



Scheme 1. The tricyclic structure common to FMN, FAD, riboflavin, and lumiflavin.

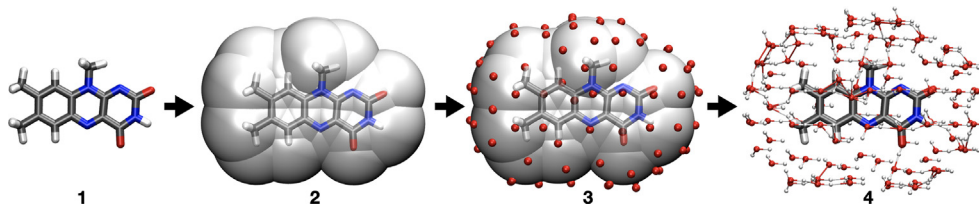


Fig. 1. Approach used to find potential lumiflavin-water interactions. We first optimized lumiflavin in the gas phase (panel 1). We then used a surface that is 3.4 Å from all atoms of the molecule (panel 2) and located points on the surface at a density of 1 point per 5 Å² (panel 3). Note that 3.4 Å is slightly longer than the length of a typical H-bond distance between heavy atoms. We placed an oxygen atom at each point and added two hydrogen atoms to each oxygen to get a total of 106 water positions (panel 4). For lumiflavin in the presence of each one of these waters, we ran an unconstrained geometry optimization and frequency calculation (a total of 106 geometry optimization and frequency calculations were performed, each with lumiflavin and one single water molecule). Of the 106 calculations, 92 converged to a structure where the water is H-bonded or non-covalently bound to the lumiflavin.

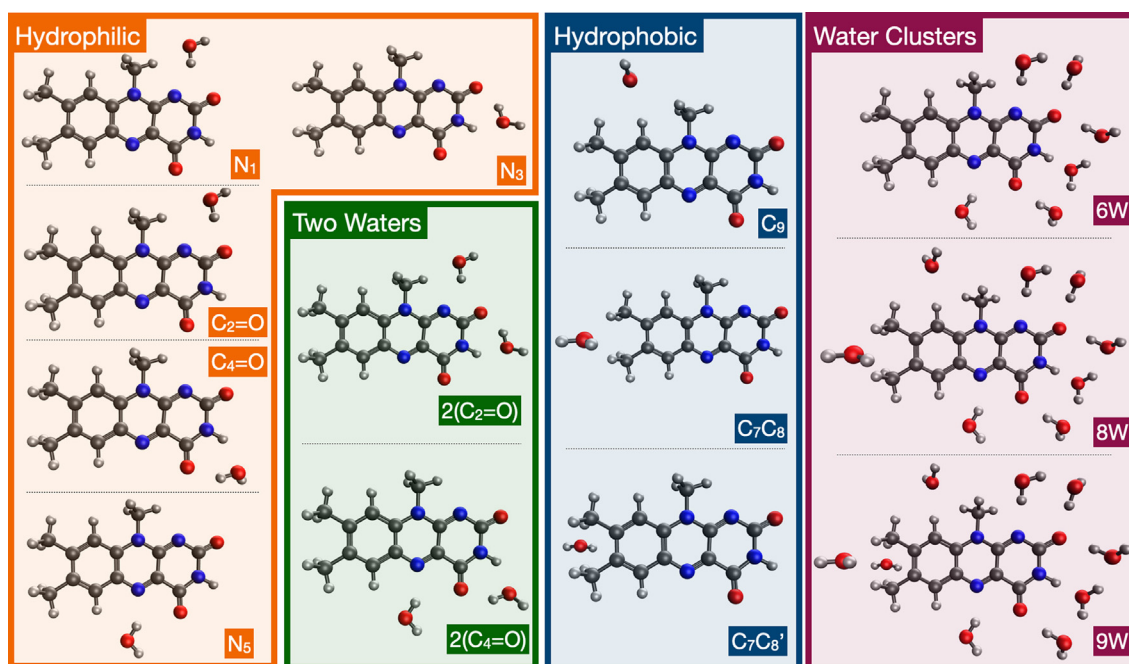


Fig. 2. Lumiflavin + water structures employed in this study. Structures in orange and blue regions were found using the approach outlined in Fig. 1. Structures in the green region include two water molecules, while structures in the magenta region include six to nine water molecules. The labels, which will be used throughout this work, are indicated next to each structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

All geometry optimizations and frequency calculations employed the B3LYP density functional and 6-31+G* basis set, unless indicated otherwise. B3LYP remains one of the most widely used functionals for the calculation of ground-state frequencies and has been found to be consistently reliable in many molecules when used with a constant scaling factor [41–45]. Water molecules were treated at the QM level in all cases. In addition to the models shown in Fig. 2, lumiflavin's geometry and frequencies were computed using a polarizable continuum model (PCM) using the integral equation formalism (IEF) accounting for a water solvent [46]. Gas-phase models serve as a reference for the effect of solvation or specific H-bonding interactions. For all models, only positive frequencies were calculated.

FTIR experiments are normally conducted for samples in D₂O (Fig. S1) to avoid bands from intense water bending vibrations that can overwhelm flavin bands of interest. Therefore, our frequency calculations included an isotope effect by replacing all hydrogen atoms likely to undergo isotope exchange with deuterium atoms. Specifically, the flavin N₃ hydrogen and all water hydrogens were replaced with deuterium for the frequency calculations. All calculations were undertaken using Gaussian 16 software [47]. A comparison between frequency calculations with and without

deuteration are reported in Fig. S3 of the SI, and reproduce the experimentally observed differences between flavin spectra in H₂O and D₂O.

To quantify the strength of H-bonding interactions in the models shown in Fig. 2, local mode force constants [48,49] (k_a) were computed using LMODEA [50,51]. k_a is a powerful metric for representing bond strength [52,53] between two atoms described by an internal bond length coordinate (q). Local mode theory has also been widely used to determine the strength of non-covalent interactions such as H-bonding [51,54–57].

Using LMODEA, k_a was computed for hydrogen---acceptor (H---A) non-covalent interactions. k_a can be related to a more chemically intuitive parameter, the bond order (n), through a power relationship [58,59], as long as the bond order is defined with respect to some reference. We employ equation (1) derived by Freindorf et al. [56] to compute the H---A “bond orders” for flavin-water interactions for each of the models in Fig. 2.

$$n = (0.532)k_a^{0.278} \quad (1)$$

Freindorf et al. [56] determined this power relationship between n and k_a through a comprehensive computational study of many H-bonding pairs. Specifically, for H---A non-covalent

interactions, the bond order was calibrated using 0 for no H-bonding interaction and 0.5 for the strongest ($[F\cdots H\cdots F]^-$) H-bond, where the hydrogen is equally shared between two fluorine centers. For bonded interactions, n is 1 for the $F-H$ monomer (full single covalent bond). Using these reference values, the weakest H-bond reported by Freindorf et al. was between ammonia and difluorine ($n = 0.134$) [56]. These values will serve as calibration points for the discussion of H-bonding interactions in this manuscript.

3. Results and discussion

Table 1 displays k_a and n values for all of the H-bonding interactions in the 9W model. n is computed from k_a using Eq. (1). As expected, the strongest interactions (largest k_a and n) are between water molecules and carbonyl oxygens. The values reported in Table 1 are consistent with typical H-bond interaction strengths reported in Ref. [56]. The $N_3-H\cdots OH_2$ interaction is also relatively strong, likely due to the strong polarity of the N_3-H bond caused by aromaticity of N_3 . As similar strength has been observed in H-bonding interactions involving imidazole [56].

The weakest non-covalent interactions in Table 1, as expected, are $C-H\cdots$ water interactions, which do not constitute H-bonds. The bond orders of these interactions, however, are of similar magnitude (or even stronger) than interactions between ammonia and non-polar difluorine reported by Freindorf et al. ($n = 0.134$) [56]. This is likely due to the small dipole in $C-H$ bonds, which creates a sufficiently strong dipole interaction with the water to keep it nearby during geometry optimizations.

To determine whether H-bonding interaction strengths are consistent in the other models shown in Fig. 2, n was computed for each H-bonding interaction in all the models. The results are included in Table S3 in the SI. The calculations show that n is consistent for each type of interaction in the different models, with the

Table 1

Computed bond distance (q), local mode force constant (k_a), and bond order (n) for various H-bonding interactions in 9W.

Interaction	Bond	q (Å)	k_a (mdyn/Å)	n
N_1-HOH	N_1-H	2.093	0.122	0.30
N_3-H-OH_2	$H-O$	1.853	0.231	0.35
$C_2=O-HOH$	$O-H$	1.874	0.199	0.34
$C_2=O-HOH'$	$O-H$	1.713	0.334	0.39
$C_4=O-HOH$	$O-H$	1.868	0.209	0.34
N_5-HOH	N_5-H	2.069	0.134	0.30
C_9-H-OH_2	$H-O$	2.388	0.048	0.23
$C_7-Methyl-H-OH_2$	$H-O$	2.649	0.023	0.19
$C_8-Methyl-H-OH_2$	$H-O$	2.705	0.033	0.21
$C_7-Methyl-H-OH_2'$	$H-O$	2.584	0.018	0.17
$C_8-Methyl-H-OH_2'$	$H-O$	2.594	0.006	0.13

exception of H-bonding interactions near N_1 ; Since 6W, 8W, and 9W models include two water molecules in the vicinity of N_1 and $C_2=O$, the waters are well oriented in those systems to form a strong H-bond with each of N_1 and $C_2=O$. However, in the case of the N_1 and $C_2=O$ models with one water molecule (shown in orange in Fig. 2), there is no water network to support two individual H-bonds and the single water interacts less strongly with both the N_1 and $C_2=O$ of flavin, yielding smaller bond orders for those interactions (Table S3).

Fig. 3 shows experimental (bottom panel, for FMN) and calculated (all other panels, for lumiflavin) IR spectra in the 1750–1450 cm^{-1} region. The PCM model in Fig. 3 includes only the lumiflavin and PCM solvent, while all other models include explicit QM water molecules without continuum solvation. All computed frequencies were broadened using gaussians with an 8 cm^{-1} full-width at half-maximum (FWHM) to simulate the broadening observed experimentally for the intense $\nu_{C=C}$ band near 1548 cm^{-1} . Initially, a constant scaling factor of 0.964 was applied to all computed frequencies (Fig. 3A) [45]. While those calculations reproduce the main features observed experimentally, the calcu-

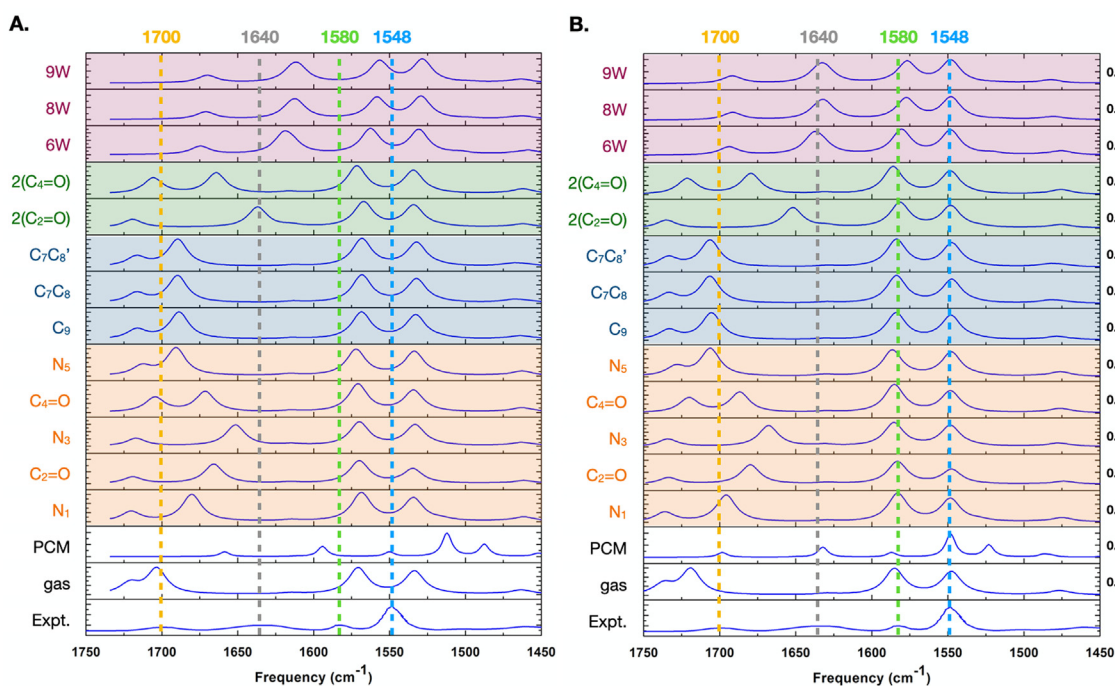


Fig. 3. Experimental [32] (bottom panel) and computed (all other panels, each panel representing a different model) IR spectra in the 1750–1450 cm^{-1} region. Four bands dominate the spectra and are assigned to $\nu_{C=C}$ (blue), $\nu_{C=N}$ (green), $\nu_{C=O}$ (grey), and ν_{C-H} (yellow). Colored dashed lines are used to help compare calculated and experimental spectra. The background is colored in a way that corresponds to the color scheme in Fig. 2 to indicate the model used. In part A, calculated frequencies were scaled by 0.964, while in part B the frequencies were scaled so that the $\nu_{C=C}$ calculated frequency is equal to 1548 cm^{-1} . The scale factors are indicated on the right side of panel B. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lated bands in most models are consistently downshifted relative to the experimental bands. Therefore, the calculations were rescaled such that the calculated $\nu_{C=C}$ peak frequency is found at 1548 cm^{-1} , matching the experimentally determined frequency (Fig. 3B).

In all models using explicit QM water molecules, a scaling factor between 0.972 and 0.977 is needed to match the $\nu_{C=C}$ peak frequency. However, the PCM model required a scaling factor of 0.987, considerably larger than the recommended 0.964 scaling factor [45]. Of greater concern is that the PCM spectrum displays an extra, relatively intense peak near 1520 cm^{-1} (Fig. 3B) that is not present in the experimental spectrum (or any of the other calculated spectra). This additional PCM calculated peak is predominantly due to a coupled $C_6 = C_7/C_{10a} = N_1$ stretching vibration (see Scheme 1 for atom number scheme, and see Fig. S4 left for the normal mode atomic motions).

In the 6W–9W cluster models, there is a second vibration with a relatively strong intensity that is only slightly offset from the $C_2 = O$ peak at around 1624 cm^{-1} . This mode is due to a combination of C_6-C_7 and $C_2 = O$ stretching modes and is not easily resolved as an identifiable peak in our broadened spectra. This may explain why, in the experimental spectra, the $\nu_{C_2=O}$ peak near 1640 cm^{-1} appears broader than the $\nu_{C_4=O}$ peak near 1700 cm^{-1} . A shoulder is resolved at around 1640 cm^{-1} in some experiments, especially for FAD spectra but also in the FMN spectra of Iuliano et al. [32] and El-Khoury et al. [33] (in D_2O , see Fig. S1). The contribution of such a mode to this peak has also been discussed in another computational study [60].

In principle, the main internal coordinates that contribute to the mode that displays a peak at 1624 cm^{-1} can be determined using a decomposition of normal modes into local modes [61,62]. However, in larger polycyclic molecules such as lumiflavin, the choice of a set non-redundant set of internal coordinates required to perform such a decomposition is not obvious.

To more easily compare calculated and experimental frequencies for the different molecular models, Fig. 4A shows a plot of the calculated frequencies of each of the four modes ($\nu_{C=C}$, $\nu_{C=N}$, $\nu_{C_2=O}$, and $\nu_{C_4=O}$) for the different models, using a constant scaling factor of 0.964. With this scaling factor, none of the models give calculated frequencies that agree well with experiments. However, all models show a similar calculated frequency for the $\nu_{C=C}$ mode (which is underestimated relative to the experiment based on the chosen scale factor). This indicates that the $\nu_{C=C}$ modes are not sensitive to any of the flavin-water intermolecular interactions. Fig. 4B plots the data using a tailored scaling factor (the same scaling factors shown in Fig. 3B). Using this frequency scaling, the PCM calculated spectrum is more in line with the experimental data (except the additional peak discussed above). An important obser-

vation, which may be useful in calculations for flavin in proteins, is that the calculated frequency differences between the $\nu_{C=C}$ and $\nu_{C=N}$ modes are similar for all the molecular models (Fig. 4B). That is, the computed $\nu_{C=C}$ and $\nu_{C=N}$ mode frequencies are insensitive to the molecular model, and hence are independent of H-bonding interactions.

On the other hand, the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ mode frequencies are very sensitive to the interactions with water molecules (Fig. 4). In the calculations, the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ mode frequencies downshifts 69 cm^{-1} and 57 cm^{-1} , respectively, relative to the gas phase when H-bonded to two water molecules (see data with green background in Fig. 4B compared to the gas phase). Additional waters (i.e., in the 6W, 8W, and 9W cluster models; magenta background in Fig. 4B) downshift the $\nu_{C_2=O}$ frequency further but upshift the $\nu_{C_4=O}$ frequency relative to the model where it is the only H-bonded group. Overall, it appears that the $\nu_{C_2=O}$ frequency is more sensitive to H-bonding (even if that H-bonding is to other nearby atoms) than the $\nu_{C_4=O}$ frequency, which is only downshifted by direct H-bonding. When only the $C_4 = O$ is H-bonded but not $C_2 = O$, the $\nu_{C_4=O}$ frequency is lower than that of $\nu_{C_2=O}$ (see $C_4 = O$ and $2(C_4 = O)$ in Fig. 4B).

In summary, the calculations shown in Fig. 4B indicate that the $\nu_{C=C}$ and $\nu_{C=N}$ frequencies are not sensitive to H-bonding interactions, while the $\nu_{C_2=O}$ frequency displays the highest sensitivity to any H-bonding near flavin's hydrophilic ring. The $\nu_{C_4=O}$ frequency is only shifted by direct H-bonding. These observations are consistent with the flavin spectra in H_2O and D_2O (see Fig. S4); using a deuterated solvent has the largest effect on the $\nu_{C_2=O}$ peak, the next largest effect on the $\nu_{C_4=O}$, and no detectable effect on the $\nu_{C=C}$ and $\nu_{C=N}$ frequencies.

In addition to calculating the vibrational frequencies, we also consider whether the calculations can reproduce the experimentally observed relative intensities of the IR bands. Experimental relative intensities were established by considering the relative areas under the FTIR absorption bands (using the relative heights is misleading since the peaks are not broadened to the same extent). This relative area is established by fitting each absorption band to a gaussian function and then integrating to obtain the area under each fit. The calculated relative intensities are obtained by dividing the computed mode intensities for two peaks. The computed and experimental relative intensities are compared in Fig. 5.

Experimentally, the $\nu_{C=C}$ band is more than 4.15 times more intense than the $\nu_{C=N}$ band. Only the PCM model reproduces this experimental relative intensity ratio, while the gas-phase and atomistic models all indicate $\nu_{C=C}$ and $\nu_{C=N}$ bands of nearly equal intensity (blue data points in Fig. 5). Since the $\nu_{C=C}$ and $\nu_{C=N}$ relative intensities can only be reproduced using a dielectric continuum, this suggests that long-range electrostatic interactions

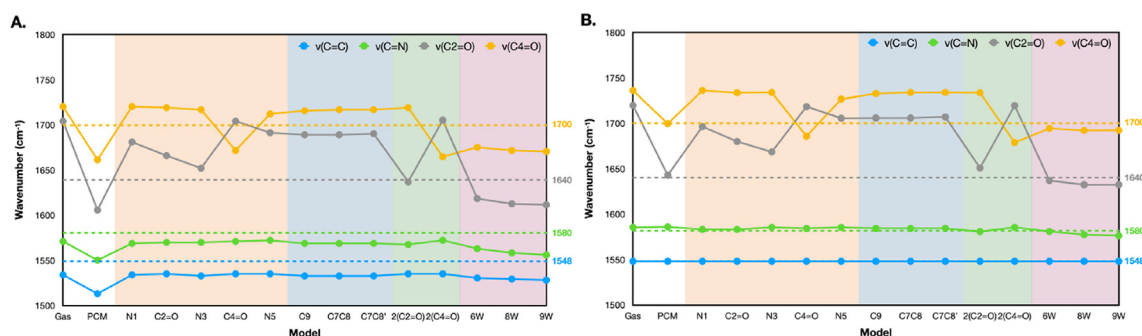


Fig. 4. A. The frequencies of the $\nu_{C=C}$ (blue), $\nu_{C=N}$ (green), $\nu_{C_2=O}$ (grey), and $\nu_{C_4=O}$ (yellow) modes computed with different models. The experimental reference³² is indicated with a dashed line, and the corresponding frequency on the right vertical axis. A single scaling factor of 0.964 is used. The background is colored coded as in Figs. 2 and 3 to indicate the model used. B. The same data but with the frequencies scaled so that the calculated $\nu_{C=C}$ frequency matches the experiment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

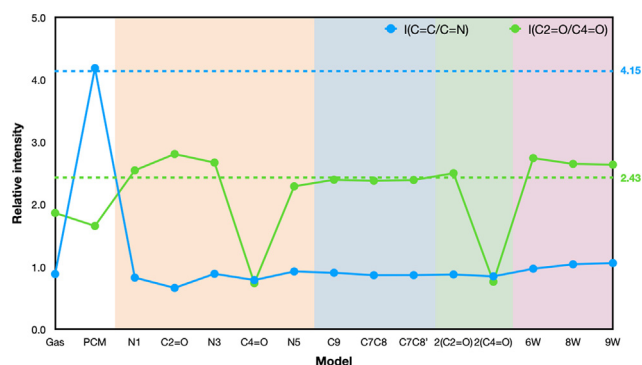


Fig. 5. The relative intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ peaks (blue) and $\nu_{C2=O}$ and $\nu_{C4=O}$ peaks (green) computed with different models. The experimental references are indicated with a dashed line with corresponding relative intensity shown on the right (obtained by finding the relative areas under the peaks from the experimental FTIR spectrum in Refs. [34,32]). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contribute to the relative intensities. This hypothesis is supported by QM/MM calculations [60] that reproduced the large difference in intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ modes only when using a solvent QM/MM model but not in the gas phase. Therefore, it is likely that the different intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ modes of flavin are not due to a single specific flavin-water interaction, but rather are due to a dielectric effect of the collective solvent molecules.

While the $\nu_{C=C}/\nu_{C=N}$ relative intensity calculated using the PCM approach is in line with the experiment, the PCM calculated $\nu_{C2=O}/\nu_{C4=O}$ relative intensity is not (Fig. 5). Many of the other

molecular models, however, do give calculated $\nu_{C2=O}/\nu_{C4=O}$ relative intensities that align well with the experiment (Fig. 5).

Finally, we studied the effect of changing the basis set on the results of the calculations. We pick three representative models that reproduce aspects of the experimental data well; PCM, 6W, and 9W. A comparison of how the calculated mode frequencies depend on the basis set is outlined in Fig. 6A. The frequencies were scaled such that the computed $\nu_{C=C}$ frequency matches the experiment (as in Fig. 3B) for each basis set. How the relative mode intensity ratios vary as a function of basis set is outlined in Fig. 6B.

In the absence of diffuse functions (i.e., with the 6-31G* basis set), the $\nu_{C2=O}$ and $\nu_{C4=O}$ frequencies in the PCM and 6W models agree poorly with the experiment (Fig. 6A). The calculations converge with the 6-31+G* basis set, and there appears to be limited benefit of considering calculations (of the type performed here) with a larger basis set, especially for the 6W and 9W models. Diffuse functions are likely necessary to correctly describe the noncovalent interactions between the flavin and water. We note that several previous studies have used either the 6-31G or 6-31G* basis set to model the FTIR spectrum of flavin, which may explain why these models could not reproduce well the relative frequencies of those prominent vibrational modes.

To better understand the origin of the basis-set dependence observed in Fig. 6, we computed bond orders from k_s s computed using different basis sets (Table 2). We find that computed bond orders are consistent across all the basis sets (within 0.01–0.05), with a notable exception; The N_1 ---HOH interaction is weaker in the case of the 6-31G and 6-31G* basis sets relative to larger basis set calculations. Due to missing diffuse functions, this interaction is weaker in the small basis calculations, which may be the origin of the error observed for those basis sets in Fig. 6. For larger basis sets, on the other hand, it appears that having diffuse and polarization

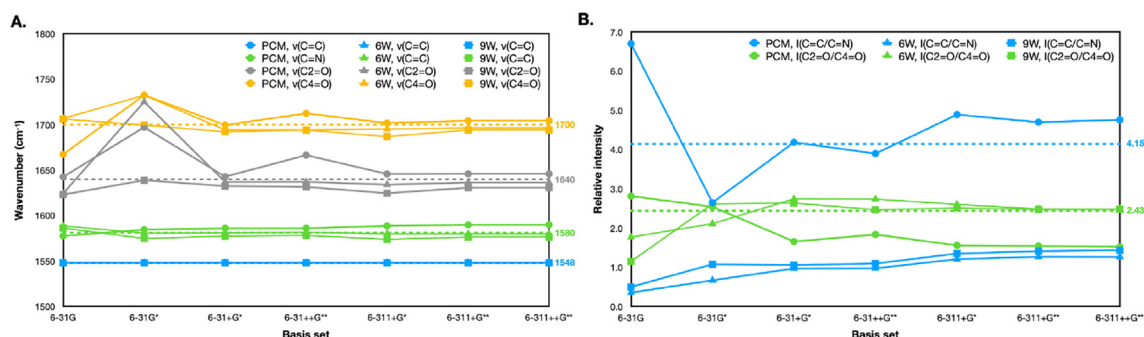


Fig. 6. A. The frequencies of the $\nu_{C=C}$ (blue), $\nu_{C=N}$ (green), $\nu_{C2=O}$ (grey), and $\nu_{C4=O}$ (yellow) modes computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares) models. The experimental reference is indicated with a dashed line and the corresponding frequency on the right. Frequencies were scaled such that the C = C computed frequency matches the experiment. B. The relative intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ peaks (blue) and $\nu_{C2=O}$ and $\nu_{C4=O}$ peaks (green) computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares) models. The experimental reference, indicated with a dashed line and the corresponding relative intensity on the right, is obtained by integrating to find the relative areas under the peaks from the experimental FTIR spectrum from Ref. [32]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Computed bond orders (n) for all H-bonding interactions in 9W computed with different basis sets.

Interaction	6-31G	6-31G*	6-31+G*	6-311+G*	6-311+G**	6-311++G**
N_1 —HOH	0.16	0.18	0.30	0.30	0.29	0.29
N_3 —H—OH ₂	0.36	0.34	0.35	0.35	0.34	0.34
C_2 = O—HOH	0.35	0.34	0.34	0.34	0.33	0.33
C_2 = O—HOH'	0.41	0.40	0.39	0.40	0.38	0.38
C_4 = O—HOH	0.36	0.35	0.34	0.34	0.33	0.33
N_5 —HOH	0.31	0.30	0.30	0.30	0.30	0.30
C_9 —H—OH ₂	0.28	0.27	0.23	0.26	0.23	0.23
C_7 —Methyl—H—OH ₂	0.21	0.21	0.19	0.21	0.19	0.19
C_8 —Methyl—H—OH ₂	0.19	0.19	0.21	0.21	0.19	0.19
C_7 —Methyl—H—OH ₂	0.21	0.22	0.17	0.21	0.19	0.19
C_8 —Methyl—H—OH ₂	0.21	0.20	0.13	0.21	0.20	0.19

functions on hydrogen atoms does not considerably change the description of H-bonding interactions, since 6-31+G* bond orders in Table 2 are largely consistent with results obtained with larger basis sets.

The relative peak intensities for the different models (and PCM in particular) show a much larger variation with the change in basis set (Fig. 6B). Specifically, the excellent agreement of the relative intensities in the case of the PCM model with 6-31+G* appears fortuitous, since a further increase in basis set size leads to an increase of the relative intensities relative to the experiment. Increasing the basis set size gradually increases the $v_{C=C}/v_{C=N}$ relative intensities in the explicit water QM models, but not enough to get close to the experimental relative intensities. The $v_{C_2=O}/v_{C_4=O}$ relative intensities are captured very well by the explicit water QM models, on the other hand. This suggests that while the long-range electrostatic effect of the solvent has an effect on the $v_{C=C}/v_{C=N}$ relative intensities, the $v_{C_2=O}/v_{C_4=O}$ relative intensities are less sensitive to these long-range electrostatics.

4. Conclusions

The reaction mechanisms of many flavoproteins are still poorly understood, and FTIR DS is an important tool for obtaining detailed mechanistic information such as changes in the flavin's intermolecular interactions with nearby amino acids. However, interpreting FTIR DS experiments would require some understanding of how H-bonding, electrostatics, and other intermolecular interactions affect normal mode vibrational frequencies and intensities of molecular groups of the flavin isoalloxazine ring. To start to address this issue, we have modeled the effect of H-bonding interactions on four prominent vibrational modes of a flavin model system. We find that the $v_{C=C}$ and $v_{C=N}$ mode frequencies are not sensitive to any H-bonding interactions, but their relative intensities are affected by the dielectric environment of the solvent. On the other hand, while the $v_{C_2=O}$ and $v_{C_4=O}$ mode frequencies are strongly downshifted by direct (both $v_{C_2=O}$ and $v_{C_4=O}$) and indirect (only $v_{C_2=O}$) H-bonding interactions, their relative intensities are less sensitive to the dielectric environment.

To reproduce the relative frequencies of the four modes in flavin, a continuum solvation model or a cluster model with multiple water molecules is needed (requiring a basis set with diffuse functions), although the continuum solvation model predicts an additional relatively intense peak that does not appear in the experiments. To reproduce the relative intensities of the FTIR bands, on the other hand, it is necessary to use either a continuum model (which is not quantitative) or QM/MM model that includes many water molecules [60].

Finally, we note the utility of the simple approach in Fig. 1, which is related to the ESTM approach [38,39], used in this context for the first time to explore possible flavin-water interaction geometries in an unbiased way. This automated approach may prove useful instead of the manual addition of water molecules at a few selected positions.

CRediT authorship contribution statement

Mohammad Pabel Kabir: Methodology, Formal analysis, Data curation, Writing - original draft. **Yoelvis Orozco-Gonzalez:** Methodology, Software, Writing - review & editing. **Gary Hastings:** Conceptualization, Formal analysis, Resources, Writing - review & editing, Supervision, Funding acquisition, Project administration. **Samer Gozem:** Conceptualization, Validation, Formal analysis, Resources, Writing - original draft, Supervision, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Drs. Elfi Kraka and Marek Freindorf for providing us with a copy of LMODEA. MPK acknowledges a fellowship from the Molecular Basis of Disease Program at Georgia State University. SG acknowledges the National Science Foundation (NSF) XSEDE for computational resources through Research Allocation CHE180027. We also acknowledge the Advanced Research Computing Technology and Innovation Core (ARCTIC) resources, which is supported by the NSF Major Research Instrumentation (MRI) grant number CNS-1920024. This material is based on work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC-0017937 to GH, and by the National Science Foundation under Grant No. CHE-2047667 to SG.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2021.120110>.

References

- [1] G. Gadda, Flavins, in: *Encyclopedia of Biophysics*, Springer, 2013, pp. 771–775.
- [2] E. Romero, J.R. Gomez Castellanos, G. Gadda, M.W. Fraaije, A. Mattevi, Same substrate, many reactions: oxygen activation in flavoenzymes, *Chem. Rev.* 118 (4) (2018) 1742–1769.
- [3] P. Macheroux, B. Kappes, S.E. Ealick, Flavogenomics—a genomic and structural view of flavin-dependent proteins, *FEBS J.* 278 (15) (2011) 2625–2634.
- [4] V. Massey, The chemical and biological versatility of riboflavin, *Biochem. Soc. Trans.* 28 (4) (2000) 283–296.
- [5] C. Walsh, Flavins: at the crossroads of biological redox chemistry, *Acc. Chem. Res.* 13 (5) (1980) 148–155.
- [6] S. Ghisla, V. Massey, Mechanisms of flavoprotein-catalyzed reactions, *Eur. J. Biochem.* 181 (1) (1989) 1–17.
- [7] S. Bornemann, Flavoenzymes that catalyze reactions with no net redox change, *Nat. Prod. Rep.* 19 (6) (2002) 761–772.
- [8] S.O. Mansoorabadi, C.J. Thibodeaux, H.-W. Liu, The diverse roles of flavin coenzymes nature's most versatile thespians, *J. Organ. Chem.* 72 (17) (2007) 6329–6342.
- [9] M. Ahmad, A.R. Cashmore, HY4 gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor, *Nature* 366 (6451) (1993) 162–166.
- [10] M. Gomelsky, G. Klug, BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms, *Trends Biochem. Sci.* 27 (10) (2002) 497–500.
- [11] D. Seth, Photoenzymic repair of UV-damaged DNA: a chemist's perspective, *Chem. Soc. Rev.* 24 (4) (1995) 289–297.
- [12] A. Sancar, Structure and function of DNA photolyase, *Biochemistry* 33 (1) (1994) 2–9.
- [13] J.M. Christie, M. Salomon, K. Nozue, M. Wada, W.R. Briggs, LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide, *Proc. Natl. Acad. Sci. U.S.A.* 96 (15) (1999) 8779–8783.
- [14] T. Kottke, A. Batschauer, M. Ahmad, J. Heberle, Blue-light-induced changes in Arabidopsis cryptochrome 1 probed by FTIR difference spectroscopy, *Biochemistry* 45 (8) (2006) 2472–2479.
- [15] D. Yamada, H. Kandori, FTIR spectroscopy of flavin-binding photoreceptors, *Methods Mol. Biol.* 1146 (2014) 361–376.
- [16] W. Mantle, Reaction-induced infrared difference spectroscopy for the study of protein function and reaction mechanisms, *Trends Biochem. Sci.* 18 (6) (1993) 197–202.

- [17] V.A. Lorenz-Fonfria, Infrared difference spectroscopy of proteins: from bands to bonds, *Chem. Rev.* 120 (7) (2020) 3466–3576.
- [18] G. Wille, M. Ritter, R. Friedemann, W. Mantele, G. Hubner, Redox-triggered FTIR difference spectra of FAD in aqueous solution and bound to flavoproteins, *Biochemistry* 42 (50) (2003) 14814–14821.
- [19] Q. Cui, M. Karplus, Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations, *J. Chem. Phys.* 112 (3) (2000) 1133–1149.
- [20] U.N. Morzan, D.J. Alonso de Armino, N.O. Foglia, F. Ramirez, M.C. Gonzalez Lebrero, D.A. Scherlis, D.A. Estrin, Spectroscopy in complex environments from QM-MM simulations, *Chem. Rev.* 118 (7) (2018) 4071–4113.
- [21] M. Nonella, G. Mathias, P. Tavan, Infrared spectrum of p-benzoquinone in water obtained from a QM/MM hybrid molecular dynamics simulation, *J. Phys. Chem. A* 107 (41) (2003) 8638–8647.
- [22] H.P. Lamichhane, G. Hastings, Calculated vibrational properties of pigments in protein binding sites, *Proc. Natl. Acad. Sci. U. S. A.* 108 (26) (2011) 10526–10531.
- [23] A. Ghysels, H.L. Woodcock 3rd, J.D. Larkin, B.T. Miller, Y. Shao, J. Kong, D.V. Neck, V.V. Speybroeck, M. Waroquier, B.R. Brooks, Efficient calculation of QM/MM frequencies with the mobile block hessian, *J. Chem. Theory Comput.* 7 (2) (2011) 496–514.
- [24] L.M. Thompson, A. Lasoroski, P.M. Champion, J.T. Sage, M.J. Frisch, J.J. van Thor, M.J. Bearpark, Analytical harmonic vibrational frequencies for the green fluorescent protein computed with ONIOM: chromophore mode character and its response to environment, *J. Chem. Theory Comput.* 10 (2) (2014) 751–766.
- [25] K. Schwinn, N. Ferre, M. Huix-Rotllant, Efficient analytic second derivative of electrostatic embedding QM/MM energy: normal mode analysis of plant cryptochrome, *J. Chem. Theory Comput.* 16 (6) (2020) 3816–3824.
- [26] M. Huix-Rotllant, K. Schwinn, N. Ferré, Infrared spectroscopy from electrostatic embedding QM/MM: local normal mode analysis of blue-light-induced infrared spectra of arabidopsis thaliana plant cryptochrome, *PCCP* (2020).
- [27] T. Iwata, T. Nagai, S. Ito, S. Osoegawa, M. Iseki, M. Watanabe, M. Unno, S. Kitagawa, H. Kandori, Hydrogen bonding environments in the photocycle process around the flavin chromophore of the AppA-BLUF domain, *J. Am. Chem. Soc.* 140 (38) (2018) 11982–11991.
- [28] T. Domratcheva, B.L. Grigorenko, I. Schlichting, A.V. Nemukhin, Molecular models predict light-induced glutamine tautomerization in BLUF photoreceptors, *Biophys. J.* 94 (10) (2008) 3872–3879.
- [29] J.J. Goings, P.F. Li, Q.W. Zhu, S. Hammes-Schiffer, Formation of an unusual glutamine tautomer in a blue light using flavin photocycle characterizes the light-adapted state, *P Natl. Acad. Sci. U. S. A.* 117 (43) (2020) 26626–26632.
- [30] B. Rieff, S. Bauer, G. Mathias, P. Tavan, DFT/MM description of flavin IR spectra in BLUF domains, *J. Phys. Chem. B* 115 (38) (2011) 11239–11253.
- [31] V.I. Birss, A.S. Hinman, C.E. Mcgarvey, J. Segal, In-situ ftir thin-layer reflectance spectroscopy of flavin adenine-dinucleotide at a mercury gold electrode, *Electrochim. Acta* 39 (16) (1994) 2449–2454.
- [32] J.N. Iuliano, J.B. French, P.J. Tonge, Vibrational spectroscopy of flavoproteins, *Methods Enzymol.* 620 (2019) 189–214.
- [33] Y. El Khoury, L.J. Van Wilderen, J. Bredenbeck, Ultrafast 2D-IR spectroelectrochemistry of flavin mononucleotide, *J. Chem. Phys.* 142 (21) (2015) 212416.
- [34] Zhao, R., Photochemistry, photophysics and spectroscopy of redox states of flavins relevant to photoactive flavoproteins. 2012.
- [35] M. Spexard, D. Immeln, C. Thioing, T. Kottke, Infrared spectrum and absorption coefficient of the cofactor flavin in water, *Vib. Spectrosc.* 57 (2) (2011) 282–287.
- [36] R.K. Kar, V.A. Borin, Y. Ding, J. Matysik, I. Schapiro, Spectroscopic properties of lumiflavin: a quantum chemical study, *Photochem. Photobiol.* 95 (2) (2019) 662–674.
- [37] C. Thioing, A. Pfeifer, S. Kakorin, T. Kottke, Protonated triplet-excited flavin resolved by step-scan FTIR spectroscopy: implications for photosensory LOV domains, *PCCP* 15 (16) (2013) 5916–5926.
- [38] Y. Orozco-Gonzalez, M. Kabir, S. Gozem, Electrostatic spectral tuning maps for biological chromophores, *J. Phys. Chem. B* 123 (23) (2019) 4813–4824.
- [39] M.P. Kabir, Y. Orozco-Gonzalez, S. Gozem, Electronic spectra of flavin in different redox and protonation states: a computational perspective on the effect of the electrostatic environment, *PCCP* 21 (30) (2019) 16526–16537.
- [40] RT, M. pyvdwsurface, 2014.
- [41] J. Neugebauer, M. Reiher, B.A. Hess, Coupled-cluster Raman intensities: assessment and comparison with multiconfiguration and density functional methods, *J. Chem. Phys.* 117 (19) (2002) 8623–8633.
- [42] A.A. El-Azhary, H.U. Suter, Comparison between optimized geometries and vibrational frequencies calculated by the DFT methods, *J. Phys. Chem.* 100 (37) (1996) 15056–15063.
- [43] R. Demicheli, B. Civalleri, M. Ferrabone, R. Dovesi, On the performance of eleven DFT functionals in the description of the vibrational properties of aluminosilicates, *Int. J. Quantum Chem.* 110 (2) (2010) 406–415.
- [44] M.D. Halls, J. Velkovski, H.B. Schlegel, Harmonic frequency scaling factors for Hartree-Fock, S-VWN, B-LYP, B3-LYP, B3-PW91 and MP2 with the Sadlej pVTZ electric property basis set, *Theor. Chem. Acc.* 105 (6) (2001) 413–421.
- [45] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 21, August 2020, Editor: Russell D. Johnson III. <http://cccbdb.nist.gov/>.
- [46] S. Miertuš, E. Scrocco, J. Tomasi, Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects, *Chem. Phys.* 55 (1) (1981) 117–129.
- [47] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, Gaussian 16, Revision A 3 (2016).
- [48] Z. Konkoli, D. Cremer, A new way of analyzing vibrational spectra. I. Derivation of adiabatic internal modes, *Int. J. Quantum Chem.* 67 (1) (1998) 1–9.
- [49] Z. Konkoli, J.A. Larsson, D. Cremer, A new way of analyzing vibrational spectra. II. Comparison of internal mode frequencies, *Int. J. Quantum Chem.* 67 (1) (1998) 11–27.
- [50] W. Zou, Y. Tao, M. Freindorf, M. Makos, N. Verma, E. Kraka, LMODEA2020, Dallas, TX, 2020.
- [51] E. Kraka, W.L. Zou, Y.W. Tao, Decoding chemical information from vibrational spectroscopy data: local vibrational mode theory, *WIREs Comput. Mol. Sci.* 10 (5) (2020).
- [52] W. Zou, D. Cremer, C2 in a box: determining its intrinsic bond strength for the X(1)Sigma(+) ground state, *Chemistry* 22 (12) (2016) 4087–4099.
- [53] W.L. Zou, Y.W. Tao, M. Freindorf, D. Cremer, E. Kraka, Local vibrational force constants – from the assessment of empirical force constants to the description of bonding in large systems, *Chem. Phys. Lett.* 748 (2020).
- [54] R. Kalescky, W. Zou, E. Kraka, D. Cremer, Local vibrational modes of the water dimer – comparison of theory and experiment, *Chem. Phys. Lett.* 554 (2012) 243–247.
- [55] R. Kalescky, E. Kraka, D. Cremer, Local vibrational modes of the formic acid dimer–the strength of the double hydrogen bond, *Mol. Phys.* 111 (9–11) (2013) 1497–1510.
- [56] M. Freindorf, E. Kraka, D. Cremer, A comprehensive analysis of hydrogen bond interactions based on local vibrational modes, *Int. J. Quantum Chem.* 112 (19) (2012) 3174–3187.
- [57] Y.W. Tao, W.L. Zou, E. Kraka, Strengthening of hydrogen bonding with the push-pull effect, *Chem. Phys. Lett.* 685 (2017) 251–258.
- [58] E. Kraka, D. Cremer, Characterization of CF bonds with multiple-bond character: bond lengths, stretching force constants, and bond dissociation energies, *ChemPhysChem* 10 (4) (2009) 686–698.
- [59] E. Kraka, J.A. Larsson, D. Cremer, Generalization of the Badger rule based on the use of adiabatic vibrational modes, in: J. Grunenberg (Ed.), *Computational Spectroscopy*, Wiley, New York, NY, USA, 2010, pp. 105–149.
- [60] B. Rieff, G. Mathias, S. Bauer, P. Tavan, Density functional theory combined with molecular mechanics: the infrared spectra of flavin in solution, *Photochem. Photobiol.* 87 (3) (2011) 511–523.
- [61] Z. Konkoli, D. Cremer, A new way of analyzing vibrational spectra. III. Characterization of normal vibrational modes in terms of internal vibrational modes, *Int. J. Quantum Chem.* 67 (1) (1998) 29–40.
- [62] D. Cremer, J.A. Larsson, E. Kraka, New developments in the analysis of vibrational spectra On the use of adiabatic internal vibrational modes, *Theoretical and Computational Chemistry*, vol. 5, Elsevier, 1998, pp. 259–327.