A polarisable QM/MM description of NMR chemical shifts of a photoreceptor protein

Shaima Hashem, Lorenzo Cupellini, Filippo Lipparini and Benedetta Mennucci

Department of Chemistry, University of Pisa, Pisa, Italy

ABSTRACT
We present a polarisable QM/MM investigation of NMR chemical shifts of a photoreceptor protein belonging to the Blue Light-Using Flavin family. Two different structures have been proposed for this photoreceptor which show a large variability in terms of the position and orientation of the protein residues around the flavin chromophore. Here, the two structures have been investigated with our multiscale approach using both DFT and MP2 level of theory. The picture that comes out comparing the \(^1\)H chemical shifts of the flavin and the most strongly interacting protein residues with the available experimental data, indicates a different behaviour of the two structures, with one showing a better correlation with NMR measurements. This shows that hybrid quantum chemical-classical simulations of NMR chemical shifts can indeed become a valuable tool to investigate the structure of complex biosystems.

1. Introduction
Quantum chemical simulations of NMR chemical shifts have been largely used to elucidate structures of molecules [1–5], and to investigate solvent effects when combined with continuum solvation models or MM descriptions [6–12]. As a matter of fact, QM calculations of NMR properties could also represent an invaluable tool to study complex biological systems [13–15]. So far, however, the applications in this field are scarce mainly due to the high computational cost and the difficulty of a correct comparison with available experimental data.

In this study, we present an example that shows the potential of this strategy applied to photoreceptive proteins in particular. In these cases, in fact, the quantum chemical description can be focused on the embedded chromophore and the residues forming the binding pocket, which allows to successfully use hybrid quantum mechanical-classical approaches. Among the many possible photoreceptor proteins, we have here selected a member of the Blue Light-Using Flavin (BLUF) family. BLUF photoreceptors are a family of proteins controlling important physiological processes like phototaxis and photosynthetic gene expression in bacteria. These proteins non-covalently bind flavin chromophores that absorb blue light and transmit signals that lead to switching the proteins from their dark resting state to a light-induced active signalling state [16, 17]. The primary transformations, which are sensed by neigh-
bouring amino acids and lead to signal propagation, do not involve any intramolecular rearrangement of the chromophore upon excitation but changes in its hydrogen-bond environment \[18, 19\]. Flavins in fact contain two hydrogen-bond acceptor C(2,4)O sites and one donor N(3)H site in the pyrimidine moiety. The H-bond changes around the flavin in the light-activated state are revealed by a characteristic 10–15 nm redshift of its first $\pi \rightarrow \pi^*$ transition and a 20 cm$^{-1}$ downshift of the C(4) = O stretching vibration compared to the dark-adapted state. Despite numerous experimental \[17, 19–21\] and computational \[22–27\] studies, the exact nature of the signalling state and the mechanism of photoactivation are not well understood and, in some aspects, controversial. What is commonly accepted is that the light-activation mechanism includes two essential amino acids, a conserved glutamine and a conserved tyrosine (Figure 1). The hydrogen bond responsible for spectroscopic changes is most likely donated by the conserved glutamine residue, which may either rotate or tautomerize in the light-adapted state to facilitate the new hydrogen bond \[18, 24, 25\].

In this study, we focus on one of the most widely studied proteins of the BLUF family, AppA, which acts as a transcriptional antirepressor of photosynthesis gene expression in *Rhodobacter sphaeroides*, which regulates cellular responses to light and oxygen. The dark state structure of AppA has been investigated with both X-ray \[28, 29\] and nuclear magnetic resonance (NMR) techniques \[30\] but many questions remain about the position and orientation of the residues forming the binding pocket for the chromophore (here a flavin mononucleotide). The first crystal structure (PDB: 1YRX) \[28\] showed a tryptophan (TRP104) located close to flavin in the so-called Trp$_{in}$ conformation (Figure 1(A)), whereas a subsequent crystal structure (PDB: 2IYG) \[29\] presented a different arrangement where the tryptophan is pointing away from the flavin while a methionine (MET106) is close to it (Met$_{in}$ conformation in Figure 1(B)). To further complicate the picture, solution NMR structures have shown a majority of snapshots in the Trp$_{in}$ conformation, but also some in the Met$_{in}$ conformation \[30\]. The different position of tryptophan and methionine corresponds to a different H-bond network and a different conformation of the other residue strongly interacting with the chromophore, the glutamine residue (GLN63). In particular, the Met$_{in}$ structure suggests that, in the dark, the glutamine side chain is oriented with its carbonyl oxygen near the OH group of the tyrosine residue (TYR21) and its amide nitrogen close to the flavin O4: in response to light this orientation is flipped by $\sim 180$ degrees. This interpretation is exactly opposite to that given on the basis of the Trp$_{in}$ structure, where instead the dark state is characterised by the GLN63 carbonyl group oriented toward TRP104 (see Figure 1(A)). It has, however, to be noted that an unambiguous assignment of the conformation of the glutamine side chain is difficult to be obtained through X-ray diffraction data, because the electron densities of the nitrogen atom of NH$_2$ amino and oxygen atom of carbonyl of the GLN63 side chain are similar. As a matter of fact, the solution NMR investigation shows a highly flexible side chain for the glutamine.

![Figure 1](image_url). Representation of the flavin mononucleotide and the protein residues in the binding pocket of the two available crystal structures. A: Trp$_{in}$ (PDB: 1YRX) \[28\]; B: Met$_{in}$ (PDB: 2IYG) \[29\].
The research presented here aims at clarifying the structure of the dark state of AppA by calculating NMR chemical shifts of both flavin and the most interacting residues in the two different crystal structures available (Trp\textsubscript{im} and Met\textsubscript{im}) and comparing them with the available experimental data. A multiscale approach is used in which quantum chemical calculations are integrated with a Molecular Mechanics (MM) model in such a way that the two descriptions are fully coupled through a polarisable MM embedding based on induced dipoles (from now indicated as MMPol) [31]. NMR calculations have been performed at both the B3LYP and second-order Møller-Plesset (MP2) perturbation theory level so to have a comparison of two conceptionally different schemes thus providing some indication for the accuracy and reliability of the computational results. As far as we know, this is the first time that a MP2/MMPol approach is used to compute NMR properties, and as such, a short summary of the main aspects of the MP2/MMPol implementation and of its extension to nuclear magnetic shieldings is reported before presenting the application to AppA.

2. Methods

2.1. The QM/MMPol approach

In the MMPol model, the environment is represented by endowing each MM atom \(i\) with a point charge \(q_i\) and with an isotropic polarisability \(\alpha_i\), so that the electric field produced by the overall charge distribution induces a point dipole \(\mu_i\) at the classical atoms. Within this framework, the MMPol energy reads

\[
P_{\text{MMPol}} = \sum_i \left[ q_i V_i^{\text{QM}} + \sum_{j \neq i} q_j T_{ij}^{00} q_j + \frac{1}{2} \left( \alpha_i^{-1} \mu_i^2 + \sum_{j \neq i} s_{ij}^{dd} \mu_i T_{ij}^{11} \mu_j \right) - \mu_i (E_i^{\text{MM}} + E_i^{\text{QM}}) \right].
\]

(1)

In Equation (1), we have introduced a generalised \(T_{ij}^{L1.L2}\) and a modified \(T_{ij}^{L1.L2}\) Coulomb interaction [32, 33], namely:

\[
[T_{ij}^{L1.L2}]_{\alpha\beta\ldots}^{\alpha'\beta'\ldots} = s_{ij}^{qq} \frac{\partial^{L1}}{\partial r_i^{\alpha} \partial r_i^{\beta} \ldots} \frac{\partial^{L2}}{\partial r_j^{\alpha'} \partial r_j^{\beta'} \ldots} \frac{1}{|r_i - r_j|},
\]

(2)

where \(s_{ij}^{qq}\) is a scaling factor [34], that can be set to zero to exclude an interaction, for instance between two bonded atoms, or to an intermediate value to scale it, according to the specific force field exclusion rules. Moreover, the damping function \(\lambda_{1}(|r_i - r_j|)\) ensures that the Coulomb interaction between two point multipoles does not diverge when the two get too close, and is a fundamental ingredient of polarisable force fields, as it allows them to avoid the so-called polarisation catastrophe. Within the same formalism, the electrostatic potential and the electric field due to the QM charge density at the polarisable sites, become

\[
V_i^{\text{QM}} = \sum_j T_{ij}^{00} Z_j - \sum_{\mu \nu} P_{\mu \nu} \langle \chi_{\mu}(r) | \hat{T}_i^{1} | \chi_{\nu}(r) \rangle, \quad E_i^{\text{QM}} = - \sum_j T_{ij}^{00} Z_j + \sum_{\mu \nu} P_{\mu \nu} \langle \chi_{\mu}(r) | \hat{T}_i^{1} | \chi_{\nu}(r) \rangle,
\]

(3)

where \(Z_j\) are the nuclear charges, \(\chi_{\mu}\) and \(\chi_{\nu}\) atomic orbital basis functions, \(P\) the one-body density matrix, and we have used the generalised Coulomb operator:

\[
\hat{T}_i^{1}(r) = - \frac{\partial}{\partial r_i^\alpha} \frac{1}{|r - r_i|}.
\]

(4)

In Equation (1), \(E_i^{\text{MM}}\) is the field due to the MM charges at the polarisable sites

\[
E_i^{\text{MM}} = - \sum_j s_{ij}^{dd} T_{ij}^{11} \mu_j
\]

(5)

and the induced dipoles solve the following linear system of equations

\[
\alpha_i^{-1} \mu_i + \sum_{j \neq i} s_{ij}^{dd} \mu_j = E_i^{\text{MM}} + E_i^{\text{QM}}.
\]

(6)

At the SCF level, the Hartree-Fock and MMPol equations are solved self-consistently. This is achieved by solving Equation (6) at each SCF iteration using a QM field computed with the current SCF density and then updating the Fock operator with the MMPol contribution

\[
V_{ij}^{\text{pol}} = \sum_i q_i \langle \chi_{\mu}(r) | \hat{T}_i^{0} | \chi_{\nu}(r) \rangle - \sum_i \mu_i \langle \chi_{\mu}(r) | \hat{T}_i^{1} | \chi_{\nu}(r) \rangle.
\]

(7)

The MP2 energy is computed using the molecular orbitals (MO) and orbital energies \(\epsilon_p, \epsilon_p\) obtained by
solving the coupled HF/MMPol equations:

$$E^\text{MP2/MMPol} = E^\text{HF/MMPol} + \frac{1}{4} \sum_{ijab} t_{ij}^{ab} \langle ab||ij \rangle,$$

where $i, j, \ldots$ denote occupied molecular orbitals and $a, b, \ldots$ virtual ones

$$t_{ij}^{ab} = \frac{\langle ij||ab \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}.$$

We note that no explicit MMPol contribution is included in the MP2 energy: this coupling scheme, which in the continuum solvation literature is known as Perturbation to the Energy (PTE), is consistent with a second-order perturbative expansion of the energy.

### 2.2. MP2/MMPol nuclear shieldings

MP2/MMPol NMR shielding tensors can be computed as analytical second derivatives of the MP2/MMPol energy with respect to the nuclear magnetic moment and an external magnetic field. As MP2 analytical second derivatives [35–37], as well as the calculation of magnetic properties with a polarisable embedding at the SCF level [15, 38], have been discussed extensively elsewhere, we limit the discussion to the specific MMPol contribution to the MP2 corrections to the shielding tensors. In the following, we assume that the AO basis functions are field-dependent London Orbitals [39, 40], which ensure the Gauge-origin invariance of the computed results.

By first differentiating with respect to the nuclear magnetic moment of the $j$-th nucleus $m_j$, we get

$$E_{m_j} = \frac{dE^\text{MP2/MMPol}}{dm_j} = \sum_{\mu \nu} \frac{p_{\mu \nu} m_{\mu j}}{h_{\mu \nu}},$$

where $h_{\mu \nu}^{m_j}$ is the derivative of the one-electron integrals with respect to the nuclear moments and $p^{\text{MP2}} = p^{\text{HF}} + p^{(2)}$ is the MP2 relaxed density matrix. In the MO basis, the second-order correction to the density is obtained from the MP2 amplitudes (occupied-occupied and virtual-virtual blocks) and by solving a set of perturbation independent, coupled-perturbed HF (CPHF) equations (occupied-virtual block), as reported by Gauss [35, 41]. Such equations are usually referred to as to the Z-vector equations [42]. In the presence of a polarisable embedding, the CPHF equations include an explicit contribution from the environment. Using Casida’s notation [43], the CPHF equations read

$$\sum_{jb} (\hat{A} + \hat{B})_{ai,bj} p^{(2)}_{bj} = -L_{ai},$$

where

$$\hat{A}_{ai,bj} = (\epsilon_a - \epsilon_i) \delta_{ab} \delta_{ij} + \langle aj||ib \rangle$$

$$+ J_{ai,bj}^{\text{MMPol}},$$

$$\hat{B}_{ai,bj} = \langle ij||ab \rangle + J_{ai,bj}^{\text{MMPol}},$$

$$J_{ai,bj}^{\text{MMPol}} = -\sum_{kl} E_{k,ai}(T_{kl}^{11})^{-1} E_{l,bj}$$

and $L$ is the so-called MP2 Lagrangian

$$L_{ai} = \sum_{jk} p_{jk}^{(2)} (\langle ij||ak \rangle + J_{jk,ai}^{\text{MMPol}})$$

$$+ \sum_{cd} p_{cd}^{(2)} (\langle ic||ad \rangle + J_{cd,ai}^{\text{MMPol}})$$

$$+ \frac{1}{2} \sum_{jcd} (cd||aj) t_{ij}^{cd} - \frac{1}{2} \sum_{jk} (ic||jk) t_{ij}^{jk}. (12)$$

By further differentiating with respect to an external magnetic field, we get

$$\sigma_j = \frac{d^2E^\text{MP2/MMPol}}{dm_j dB} = \sum_{\mu \nu} \left( p_{\mu \nu}^{\text{MP2,2}} h_{\mu \nu} + p_{\mu \nu}^{\text{MP2,2}} h_{\mu \nu}^{m_j} \right),$$

where $h_{\mu \nu}^{m_j}$ is the second derivative of the one-electron integrals with respect to the nuclear moment $m_j$ and the external field $B$, $p_{\mu \nu}^{\text{MP2,2}} = p_{\mu \nu}^{\text{HF,B}} + p_{\mu \nu}^{\text{MP2,2}}$, and

$$p_{\mu \nu}^{(2),B} = \sum_{pq} \left( C_{pq}^{B} p_{pq}^{(2)} C_{pq} + C_{pq}^{B} p_{pq}^{(2),C} C_{pq} \right.$$  

$$+ C_{pq}^{B} p_{pq}^{MMP2} C_{pq}^{B} \right).$$

In equation (14), $C_{pq}^{B}$ are the MO coefficients and

$$C_{pq}^{B} = \sum_q U_{pq}^{B} C_{q\mu}$$

their derivatives, that are parametrised in terms of the orbital rotations $U_{pq}$. The latter are obtained from the MO orthonormality conditions (occupied-occupied and virtual-virtual blocks) and from the solution to the magnetic CPHF equations (occupied-virtual block):

$$\langle \hat{A} - \hat{B} \rangle_{ai,bl} U_{bj}^{B} = 2Q_{ai}^{B},$$

where

$$Q_{ai}^{B} = h_{ai}^{B} + C_{ai}^{B}(p^{\text{HF}}) + \sum_k q_k V_{k,ai}^{B}$$

$$- \sum_k \mu_k E_{k,ai}^{B} - G_{ai}(S_{ov}^{B}) - (FS)_{ai}^{B} (17)$$

$$+ \sum_{jcd} (cd||aj) t_{ij}^{cd} - \frac{1}{2} \sum_{jk} (ic||jk) t_{ij}^{jk}. (12)$$

By further differentiating with respect to an external magnetic field, we get

$$\sigma_j = \frac{d^2E^\text{MP2/MMPol}}{dm_j dB} = \sum_{\mu \nu} \left( p_{\mu \nu}^{\text{MP2,2}} h_{\mu \nu} + p_{\mu \nu}^{\text{MP2,2}} h_{\mu \nu}^{m_j} \right),$$

where $h_{\mu \nu}^{m_j}$ is the second derivative of the one-electron integrals with respect to the nuclear moment $m_j$ and the external field $B$, $p_{\mu \nu}^{\text{MP2,2}} = p_{\mu \nu}^{\text{HF,B}} + p_{\mu \nu}^{\text{MP2,2}}$, and

$$p_{\mu \nu}^{(2),B} = \sum_{pq} \left( C_{pq}^{B} p_{pq}^{(2)} C_{pq} + C_{pq}^{B} p_{pq}^{(2),C} C_{pq} \right.$$  

$$+ C_{pq}^{B} p_{pq}^{MMP2} C_{pq}^{B} \right).$$

In equation (14), $C_{pq}^{B}$ are the MO coefficients and

$$C_{pq}^{B} = \sum_q U_{pq}^{B} C_{q\mu}$$

their derivatives, that are parametrised in terms of the orbital rotations $U_{pq}$. The latter are obtained from the MO orthonormality conditions (occupied-occupied and virtual-virtual blocks) and from the solution to the magnetic CPHF equations (occupied-virtual block):

$$\langle \hat{A} - \hat{B} \rangle_{ai,bl} U_{bj}^{B} = 2Q_{ai}^{B},$$

where

$$Q_{ai}^{B} = h_{ai}^{B} + C_{ai}^{B}(p^{\text{HF}}) + \sum_k q_k V_{k,ai}^{B}$$

$$- \sum_k \mu_k E_{k,ai}^{B} - G_{ai}(S_{ov}^{B}) - (FS)_{ai}^{B} (17)$$
being G the usual two-electron contribution to the Fock operator, which contains the MMPol contribution in equation (7), \( G^0 \) the same, but evaluated with derived integrals, and \( S \) the overlap matrix.

As the magnetic Hessian does not include explicit MMPol contributions [15, 38], the magnetic CPHF equations in the presence of the polarisable environment only require a modified right-hand side, as in equation (17). The same applies for the MP2 coupled-perturbed equations that are needed in order to evaluate the MP2 density derivatives. As these are very cumbersome, we do not report all the expressions, which can be found in Ref. [35]. We only point out here that, again, the polarisable environment does not modify the MP2 response kernel, but gives a contribution to the right-hand side of the response equations, i.e. the Lagrangian derivative, which only involves the induced dipoles and is of the form

\[
- \sum_{\mu \nu} C_{\mu \alpha} C_{\nu \beta} \sum_{kl} E_{\mu \nu}^B \sum_{\rho \sigma} (T_{kl}^{11})^{-1} E_{l,\rho \sigma} P_{\rho \sigma}^{(2)},
\]

(18)

where the \( E_{\mu \nu}^B \) integrals are electric field integrals computed with differentiated London orbitals. All the other contributions are assembled using information available from previous intermediate computations, i.e. the HF/MMPol Fock matrix, MO coefficients, Magnetic CPHF right-hand side and MP2 relaxed density.

On a more technical note, we remark that all the purely MMPol quantities, including the solution to the MMPol equations, are computed with linear scaling computational cost using our FMM-based implementation [34, 44], so that the overall cost of adding a polarisable environment is negligible with respect to the cost of the QM calculation.

2.3. Computational details

2.3.1. Structures

The crystal structures of Trp\textsubscript{in} (PDB: 1YRX) and Met\textsubscript{in} (PDB: 2IYG) AppA were refined by adding Hydrogen atoms using the LeAP module of AmberTools. The protein was protonated at pH 7 using the H++ web server [45]. Hydrogen atoms were minimised with Amber at the MM level using the Amber ff14SB force field [46]. In order to optimise the bias coming from the unrefined crystal structure, for each crystal structure we optimised the FMN molecule at the B3LYP/6-31G(d) level of theory with a fixed protein environment described at MM level, in the ONIOM scheme [47]. In these optimisations, the tail of the flavin mononucleotide was kept mobile but was not included in the QM part. The MM part was described with the Amber ff14SB force field.

2.3.2. Chemical shifts

Nuclear shieldings were calculated using B3LYP/6-311+G(d,p) and MP2/6-311G(d,p) levels of theory. All the MP2 calculations were performed without frozen core orbitals. The QM part consisted of the isalloxazine ring and the first carbon of the tail, as shown in Figure 2. In order to account for quantum mechanical effects of the most strongly interacting residues, we also added to the QM part the residues shown in Figure 2, which form a hydrogen-bonding network with the flavin. In particular, we considered the effect of ASN45, which is hydrogen bonded to the H3 atom of flavin, GLN63, which is hydrogen bonded to N5 and presents different conformations in the Trp\textsubscript{in} and Met\textsubscript{in} structures, and TRP104 or MET106, respectively, for the two structures. Finally, we also considered TYR21 which does not form directly a hydrogen bond with the flavin, but with GLN63. \(^1\)H chemical shifts are reported with respect to TMS.

Figure 2. Representation of the QM subsystem used in the Cluster model for the Trp\textsubscript{in} (A) and Met\textsubscript{in} (B) structures. The flavin atoms are numbered according to Ref. [48], whereas for the protein atoms we used the standard PDB atom names.
3. Results

In this section, we summarise the main results of our NMR calculations for AppA. In this analysis, DFT(B3LYP) and MP2 results are presented and, where available, compared with available experimental data. The presentation of the data is divided into parts: one focused on the $^1$H, $^{15}$N, and $^{17}$O of the flavin chromophore and the second on $^1$H of the protein residues forming the flavin binding pocket and most strongly interacting with it.

3.1. Flavin

In Table 1, we report the $^1$H chemical shifts obtained at B3LYP and MP2 level for flavin in the two refined crystal structures of AppA, namely Trp$_{in}$ and Met$_{in}$ (see Computational details). In particular, four different models have been tested and compared. The gas-phase one considers an isolated flavin molecule but using the structure optimised within each of the two protein environments. The other models use the same structures but include the effects of the protein also in the calculation of the chemical shifts. In particular, the Full-MMPol model describes all the protein residues at classical level while the other models extend the QM subsystem to the H-bonded asparagine residue (ASN) or to the four residues more strongly interacting with the flavin (Cluster): see Figure 2 for the representation of the QM clusters for the two investigated crystal structures. In both ASN and Cluster models, the protein residues not directly included in the QM subsystem are described at MMPol level. Because of the large computational cost, the Cluster model at MP2 level has been reduced removing from the QM subsystem the tyrosine residue (TYR21) which is not directly connected to the flavin. To check the effect of this change, the B3LYP calculations have been repeated also for the reduced clusters, finding that the effects of including tyrosine in the QM subsystem are negligible on the NMR chemical shifts (see Table 1).

As it can be seen from the values reported in Table 1, the two QM descriptions give very similar results when moving from one proton to the other and, for the same proton, when changing crystal structure. This shows that the description of the different chemical and protein environment is reproduced in the same way with the two different levels of theory. By comparing the chemical shifts computed in gas phase on the Trp$_{in}$ and Met$_{in}$ structures, we can also see that the internal geometry of flavin has a small effect in determining differences between the two structures.

As a further analysis, in Figure 3 we report the change in nuclear shieldings for nitrogens and oxygens when moving from the gas-phase to Full-MMPol and the Cluster models (the complete set of nuclear shieldings is reported in Supplemental online material). As it can be seen from the graphs, the two QM levels give a similar description of the effects of the protein and also a similar behaviour with respect to the selected structure. We note, however, that MP2 data seem to show a larger sensitivity to the selected crystal structure, especially for nitrogens.

Once we have verified the similarity between DFT and MP2 results, we can move to a detailed analysis of the single nuclei and compare with experimental data. In this analysis we focus on DFT data, possibly employing MP2 results when the two methods give significant differences.

It is known that the chemical shift of N(3)H reacts quite sensitively to hydrogen-bonding interaction; in aprotic solvents (e.g. in CHCl$_3$), the resonance appears at $\sim 8$ ppm, whereas in DMSO, a hydrogen-bond acceptor, it appears at $\sim 11$ ppm [48]: this corresponds to a $\Delta \delta$ value of $\sim 3$ ppm. In the AppA, a value of 11.16 ppm has been measured for such a proton which clearly indicates a strong H-bond interaction with the close-by asparagine.

### Table 1. B3LYP and MP2 chemical shifts (ppm) of selected protons of flavin as obtained by the different models.

|         | Trp$_{in}$ Gas phase | Full-MMPol | ASN | Cluster | Trp$_{in}$ Gas phase | Full-MMPol | ASN | Cluster |
|---------|----------------------|------------|-----|---------|----------------------|------------|-----|---------|
| EXP N(3)H | 11.16 | 7.80 | 8.75 | 11.2 | 10.88(10.89) | 7.67 | 8.98 | 11.13 | 11.26(11.25) |
| C(6)H | 8.08 | 7.57 | 7.56 | 8.42(8.13) | 8.34 | 8.15 | 8.17 | 8.64(8.60) |
| C9(H) | 7.99 | 7.15 | 8.18 | 8.20 | 8.21(8.19) | 7.16 | 7.71 | 7.93(7.78) |
| 7CH3 | 1.22 | 2.39 | 2.48 | 2.47 | 2.70(2.56) | 2.31 | 2.39 | 2.32 | 2.64(2.36) |
| 8CH3 | 1.33 | 2.32 | 2.43 | 2.44 | 2.54(2.45) | 2.31 | 2.43 | 2.45 | 2.48(2.43) |

### Met$_{in}$

|         | Gas phase | Full-MMPol | ASN | Cluster |
|---------|-----------|------------|-----|---------|
| EXP N(3)H | 11.16 | 7.38 | 8.30 | 10.88 | 10.82 | 7.38 | 9.55 | 12.23 | 13.28 |
| C(6)H | 8.05 | 6.75 | 7.23 | 7.87 | 8.10 | 4.98 | 9.56 | 8.71 |
| C9(H) | 7.99 | 7.55 | 8.70 | 8.74 | 8.44 | 10.53 | 9.54 | 8.08 |
| 7CH3 | 1.22 | 2.50 | 2.85 | 2.80 | 2.50 | 1.16 | 1.50 | 2.75 |
| 8CH3 | 1.33 | 2.52 | 3.14 | 2.89 | 2.52 | 3.97 | 3.40 | 3.06 |

Note: The values in parenthesis refer to the cluster lacking the tyrosine residue. All values are with respect to TMS.
residue. Our calculations well reproduce these observations. In fact, N(3)H shows the largest shift when moving from the gas-phase to the Cluster model: $\Delta \delta$ is $\sim 3$ ppm at B3LYP level for both structures and $3$–$5$ ppm at MP2 level. It has to be noted that this shift due to H-bonds effects is only partially reproduced by a fully classical description: the $\Delta \delta$ value moving from gas-phase to Full-MMPol is around $1$ ppm at B3LYP level and $1$–$2$ ppm at MP2. The differences between Full-MMPol and cluster models for the N(3)H clearly show that charge-transfer effects between flavin and the ASN residue are possible.

It has also been shown that the chemical shift of N(3)H increases by $\sim 0.6$ ppm moving from the dark state to the light-induced signalling state [49]. Either a change in the electronic structure of the flavin isovaloxazine ring, or a change in the local residue environment has been invoked to explain the shift. As reported in the Introduction, a possible photo-induced change in the protein around flavin is the rotation of glutamine with the consequent change of H-bonding network around it. This change is indeed present in the two investigated structures (see Figure 1): we can thus use the differences calculated between them to compare with the dark-to-light changes. Looking at our calculated values, we see that in the B3LYP description, the change is $\sim 0.4$ ppm when using the Cluster model while it becomes half if the Full-MMPol model is used. At MP2 level, instead, a large change ($\sim 1.2$ ppm) is found with the Full-MMPol model already, and it even doubles when the cluster model is used. This seems to show that an accurate evaluation of the chemical shift of the N(3)H proton is extremely challenging due to its high sensitivity to both the QM level and the local environment; nonetheless, our data confirm that a change in the local residue environment could explain the observed change in the chemical shift of N(3)H after illumination.

### 3.2. Protein residues

In addition to the flavin nuclei, the NMR measurements on AppA also allowed to quantify the chemical shifts of the protons of the most interacting residues [30]. As in the Cluster model we have access to these data in both the Trp$_{in}$ and Met$_{in}$ structures, we can extend the comparison with experiments also to the protein residues which most strongly interact with flavin. This comparison is shown in Figure 4 where we report a correlation plot between the calculated and measured $^1$H chemical shifts for GLN63, TYR21, ASN45, TRP104 and MET106. In the same plot, we have also reported the values for the flavin protons.

In Trp$_{in}$, an evident outlier is represented by the HD21 of asparagine (see Figure 2), which in our calculations presents a too small chemical shift, $3.53$ ppm vs the measured $6.97$ ppm. A much better agreement is instead found for the same proton when the Met$_{in}$ structure is used. We note that this behaviour is further confirmed by the MP2 description. In parallel, we note that the other proton in the same amino group (HD22), which is hydrogen bonded to O4, shows a higher chemical shift, $9.16$ ppm in Trp$_{in}$ than in Met$_{in}$, $7.99$ ppm and the latter is much closer to the experimental value ($7.43$ ppm). This overestimated value in Trp$_{in}$ can be explained by

**Figure 3.** Change in nuclear shieldings for $^{15}$N and $^{17}$O nuclei when moving from gas phase to Full-MMPol (full bars) and Cluster (striped bars). In grey we report the results obtained in the Trp$_{in}$ structure and in black, the analogs for the Met$_{in}$ structure. Upper panel refers to B3LYP calculations and lower panel to MP2 ones.
the formation of a stronger hydrogen bond in this structure with respect to the Met\textsubscript{in}. In fact, we expect that the stronger hydrogen bonding there is, the more the proton is deshielded and the higher its chemical shift will be. Looking at the H-bond distance between O4 and HD22 we indeed find $\sim 1.84 \text{ Å}$ in Trp\textsubscript{in} while it is around $2 \text{ Å}$ in Met\textsubscript{in}. It is certainly true that the experimental values are the result of a thermal average over the protein motions which are instead neglected in our calculations. We expect that the HD21 and HD22 would show less pronounced differences if we took into consideration the fluctuations in the hydrogen bond lengths. However, the very different findings obtained in the two structures for HD21 can only be explained in terms of the different local environment consisting of TRP104 and MET106 in Trp\textsubscript{in} and Met\textsubscript{in} structures, respectively. Another interesting observation concerns the chemical shifts of the amino protons (HE21 and HE22) in GLN63: the values calculated for the Trp\textsubscript{in} structure are too large, thus indicating an excessive deshielding. The analogous chemical shifts calculated for the Met\textsubscript{in} structure, instead, are in good agreement with experiments. This time, however, the B3LYP behaviour is not confirmed by MP2 which, for these two protons, give similar accuracy for the two structures.

We further note that thermal fluctuations of the protein could also open channels inside the structure, which facilitate water solvation of the binding pocket, and in turn affect the chemical shifts of the flavin and residue side chains. There are no water molecules, in either structure, close to the flavin. However, in the Met\textsubscript{in} structure, there is a closed channel, adjacent to MET106, that could mediate the entry of water molecules into the flavin binding pocket (Figure S1B). This channel was not found in the Trp\textsubscript{in} structure, because of the bulkier tryptophan residue in place of the methionine (Figure S1A).

Finally, still in the Trp\textsubscript{in} structure, our calculations give an overestimated chemical shift for Hz2 in TRP104. To better understand this result, we have repeated the calculations of the chemical shift of TRP104 in the Met\textsubscript{in} structure: to do that we have enlarged the cluster so to include also this residue that, we recall, in this structure is placed much farther away with respect to flavin. The results obtained for this enlarged cluster in the Met\textsubscript{in} structure indeed improve the agreement with experiment. Combining the results of this and the previous section, we can conclude that our QM/MMPol investigation seems to show a more accurate agreement between calculations and experiments for the Met\textsubscript{in} structure: this is particularly evident if we use a B3LYP description, whereas MP2 reduces the differences between the two structures.

4. Summary

In this study, we have combined a polarisable MM model with both DFT and MP2 descriptions to investigate NMR chemical shifts of the BLUF photoreceptor AppA. This system has been selected because there is still a large debate in the literature not only about the molecular mechanism involved in its photoactivation but also on the real structure of the dark state. The available

Figure 4. Correlation between calculated (left: B3LYP, right: MP2) and measured $^1$H chemical shifts (ppm) of the five residues around flavin (GLN63, TYR21, ASN45, TRP104 and MET106). The grey triangles refer to the Trp\textsubscript{in} structure whereas the black diamonds to the Met\textsubscript{in} structure. The empty circles refer to calculation on an enlarged clustere performed on the Met\textsubscript{in} structure in which the QM subsystem is extended to include also the TRP104 residue. The empty triangles/diamonds are the values of the flavin protons in the two structures.
Acknowledgments

The authors would like to thank Jürgen Gauss not only for his many seminal contributions to theoretical chemistry, including the ones that made this work possible, but also for the many moments spent together at conferences and during his visits to the group in Pisa.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors acknowledge funding from the European Research Council under the grant ERC-AdG-786714 (LIFETimeS).

ORCID

Lorenzo Cupellini  http://orcid.org/0000-0003-0848-2908
Filippo Lipparini  http://orcid.org/0000-0002-4947-3912
Benedetta Mennucci  http://orcid.org/0000-0002-4394-0129

References

[1] A. Bagno and G. Saielli, Theor. Chem. Acc. 117 (5-6), 603–619 (2006).
[2] T. Metzroth, M. Lenhart, and J. Gauss, Appl. Magn. Reson. 33 (4), 457–467 (2008).
[3] R. Bast, U. Ekström, B. Gao, T. Helgaker, K. Ruud, and A.J. Thorvaldsen, Phys. Chem. Chem. Phys. 13 (7), 2627–2651 (2011).
[4] J.C. Facelli, Prog. Nucl. Magn. Reson. Spectrosoc. 58 (3–4), 176–201 (2011).
[5] M.W. Lodewyk, M.R. Siebert, and D.J. Tantillo, Chem. Rev. 112 (3), 1839–1862 (2012).
[6] B. Mennucci, J. Am. Chem. Soc. 124 (7), 1506–1515 (2002).
[7] D. Sebastiani and U. Rothlisberger, J. Phys. Chem. B 108, 2807–2815 (2004).
[8] A. Bagno, F. Rastrelli, and G. Saielli, Chem. – A Eur. J. 12 (21), 5514–5525 (2006).
[9] M. Pavone, O. Crescenzi, G. Morelli, N. Rega, and V. Barone, Theor. Chem. Acc. 116 (4–5), 456–461 (2006).
[10] K. Aidas, A. Mogelhøj, C.B. Nielsen, K.V. Mikkelsen, K. Ruud, O. Christiansen, and J. Kongsted, J. Phys. Chem. A 111 (20), 4199–4210 (2007).
[11] D. Flaig, M. Beer, and C. Ochsenefeld, J. Chem. Theory. Comput. 8 (7), 2260–2271 (2012).
[12] R.M. Gester, C. Bistafa, H.C. Georg, K. Coutinho, and S. Canuto, Theor. Chem. Acc. 133 (1), 95–8 (2013).
[13] A. Frank, I. Onila, H.M. Möller, and T.E. Exner, Proteins: Struct. Funct. Gen. 79 (7), 2189–2202 (2011).
[14] K.V.J. Jose and K. Raghavachari, J. Chem. Theory. Comput. 3 (3), 1147–1158 (2017).
[15] S. Caprasecca, L. Cupellini, S. Jurinovich, D. Loco, F. Lipparini, and B. Mennucci, Theor. Chem. Acc. 137, 84 (2018).
[16] S.Y. Park and J.R.H. Tame, Biophys. Rev. 9 (2), 1–8 (2017).
[17] T. Fujisawa and S. Masuda, Biophys. Rev. 10 (2), 1–11 (2018).
[18] S. Masuda, K. Hasegawa, and T.a. Ono, Biochemistry 44 (4), 1215–1224 (2005).
[19] T. Iwata, T. Nagai, S. Ito, S. Oseogawa, M. Iseki, M. Watanabe, M. Unno, S. Kitagawa, and H. Kandori, J. Am. Chem. Soc. 140 (38), 11982–11991 (2018).
[20] A.A. Gil, S.P. Laptenok, J.N. Juliano, A. Lukacs, A. Verma, C.R. Hall, G.E. Yoon, R. Brust, G.M. Greetham, M. Towrie, J.B. French, S.R. Meech, and P.J. Tonge, J. Am. Chem. Soc. 139 (41), 14638–14648 (2017).
[21] K. Karadi, S.M. Kapetanaki, K. Raics, I. Pecsi, R. Kapronczai, Z. Fekete, J.N. Juliano, J.T. Collado, A.A. Gil, J. Orban, M. Nyitrai, G.M. Greetham, M.H. Vos, P.J. Tonge, S.R. Meech, and A. Lukacs, Sci. Rep. 10, 2061 (2020).
[22] M.G. Khrenova, T. Domratcheva, I. Schlichting, B.L. Grigorenko, and A.V. Nemukhin, Photochem. Photobiol. 87 (3), 564–573 (2010).
[23] Y.W. Hsiao, J.P. Götzte, and W. Thiel, J. Phys. Chem. B 116 (28), 8064–8073 (2012).
[24] T. Mathies and J.P. Götzte, Front. Mol. Biosci. 2, 7998–14 (2015).
[25] T. Domratcheva, E. Hartmann, I. Schlichting, and T. Kotke, Sci. Rep. 6 (1), 1–14 (2016).
[26] P. Goyal and S. Hammes-Schiffer, Proc. Natl. Acad. Sci. 114 (7), 1480–1485 (2017).
[27] E.R. Sayfutyarova, J.J. Goings, and S. Hammes-Schiffer, J. Phys. Chem. B 123 (2), 439–447 (2019).
[28] S. Anderson, V. Dragnea, S. Masuda, J. Ybe, K. Moffat, and C. Bauer, Biochemistry 44 (22), 7998–8005 (2005).
[29] A. Jung, J. Reinstein, T. Domratcheva, R.L. Shoeman, and I. Schlichting, J. Mol. Biol. 362 (4), 717–732 (2006).
[30] J.S. Grinstead, S.T.D. Hsu, W. Laan, A.M.J.J. Bonvin, K.J. Hellingwerf, R. Boelens, and R. Kaptein, ChemBioChem 7 (1), 187–193 (2005).
[31] C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G.D. Scholes, and B. Mennucci, J. Chem. Theory Comput. 5 (7), 1838–1848 (2009).
[32] J. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. Hsieh, H. Lei, R. Luo, and Y. Duan, J. Phys. Chem. B 115, 3100 (2011).
[33] F. Lipparini, L. Lagardère, B. Stamm, E. Cancès, M. Schnieders, P. Ren, Y. Maday, and J.P. Piquemal, J. Chem. Theory Comput. 10 (4), 1638–1651 (2014).
[34] F. Lipparini, J. Chem. Theory Comput. 15 (8), 4312–4317 (2019).
[35] J. Gauss, Chem. Phys. Lett. 191 (6), 614–620 (1992).
[36] R.J. Harrison, G.B. Fitzgerald, W.D. Laidig, and R.J. Bartelt, Chem. Phys. Lett. 124 (3), 291–294 (1986).
[37] J.A. Pople, R. Krishnan, H.B. Schlegel, and J.S. Binkley, Int. J. Quantum Chem. 16 (S13), 225–241 (1979).
[38] F. Lipparini, C. Cappelli, and V. Barone, J. Chem. Phys. 138, 234108 (2013).
[39] R. Ditchfield, Mol. Phys. 27 (4), 789–807 (1974).
[40] K. Wolinski, J.F. Hinton, and P. Pulay, J. Am. Chem. Soc. 112 (23), 8251–8260 (1990).
[41] J. Gauss, J. Chem. Phys. 99 (5), 3629–3643 (1993).
[42] N.C. Handy and H.F. Schaefer, J. Chem. Phys. 81 (11), 5031–5033 (1984).
[43] M.E. Casida, C. Jamorski, K.C. Casida, and D.R. Salahub, J. Chem. Phys. 108 (11), 4439–4449 (1998).
[44] S. Caprasecca, S. Jurinovich, L. Lagardère, B. Stamm, and F. Lipparini, J. Chem. Theory Comput. 11 (2), 694–704 (2015).
[45] R. Anandakrishnan, B. Aguilar, and A.V. Onufriev, Nucl. Acids Res. 40 (W1), 537–541 (2012).
[46] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling, J. Chem. Theory Comput. 11 (8), 3696–3713 (2015).
[47] L.W. Chung, H. Hirao, X. Li, and K. Morokuma, Wiley Interdiscipl. Rev.: Comput. Mol. Sci. 2 (2), 327–350 (2011).
[48] F. Müller, Methods Mol. Biol. (Clifton, NJ) 1146, 229–306 (2014).
[49] J.S. Grinstead, M. Avila-Perez, K.J. Hellingwerf, R. Boelens, and R. Kaptein, J. Am. Chem. Soc. 128 (47), 15066–15067 (2006).