# Electronic Supplementary Information:

# From crystallographic data to solution structure of photoreceptors through Molecular Dynamics and Spectroscopy Simulations: the case of AppA BLUF domain.

Shaima Hashem,[a] Veronica Macaluso,[a] Michele Nottoli,[a] Filippo Lipparini,[a] Lorenzo Cupellini,[a*] and Benedetta Mennucci[a*]

## S1 Supplementary Tables

Table S1 Equilibration protocols for MD simulations

| **Structures** | **MD Simulations** | | **System preparation/Heating/Equilibration** | | | |
|---|---|---|---|---|---|---|
| | | Residue at position 20 | Temperature (K) | Thermodynamic ensemble | Time (ps) | Harmonic restraints |
| $Trp_{in}$ | MD1/MD2 | Cys4 | 0-50 | NVT | 50 | protein backbone |
| | | | 50-300 | NPT | 250 | |
| | | | 300 | NPT | 750 | |
| | MD3/MD4 | Cys4 | 0-100 | NVT | 50 | protein backbone |
| | | | 100-315 | NPT | 250 | protein backbone except for loops (resid 95-103,108-111) |
| | | | 315-300 | NPT | 2000 | |
| | | | 300 | NPT | 1000 | |
| $Met_{in}$ | MD1/MD2 | Ser4 | 0-50 | NVT | 50 | protein backbone |
| | | | 50-300 | NPT | 250 | |
| | | | 300 | NPT | 750 | |
| | MD3/MD4-S20C | Cys4 | 0-50 | NVT | 50 | protein backbone |
| | | | 50-300 | NPT | 250 | |
| | | | 300 | NPT | 750 | |

[a] *Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via G. Moruzzi 13, 56124 Pisa, Italy*
[*] *Corresponding author. Email: lorenzo.cupellini@unipi.it benedetta.mennucci@unipi.it*

Table S2 Probabilities of hydrogen bond formation during MD simulations. Hydrogen bonds were calculated using CPPTRAJ on snapshots sampled every 100 ps using a cutoff of 3.2 Å on the donor-acceptor distance and a cutoff of 135° on the donor-hydrogen-acceptor angle.

| | | Gln63(NH)-FMN(O4) | Gln63(NH)-FMN(N5) | Tyr21(OH)-Gln63(O) | Tyr21(O)-Gln63(NH) | Asn45(NH)-FMN(O4) | Asn45(O)-FMN(N3H) | Trp104(H)-Gln63(O) |
|---|---|---|---|---|---|---|---|---|
| $Trp_{in}$ | MD1 | 14 | 29 | 69 | 7 | 62 | 57 | 7 |
| | MD2 | 2 | 47 | 13 | 7 | 91 | 77 | 1 |
| | MD3 | 23 | 19 | 86 | 4 | 60 | 38 | 5 |
| | MD4 | 18 | 33 | 53 | 2 | 92 | 81 | 1 |
| | Average | 14 | 32 | 55 | 5 | 76 | 63 | 4 |
| $Met_{in}$ | MD1 | 14 | 44 | 96 | 0 | 72 | 92 | 0 |
| | MD2 | 7 | 52 | 93 | 0 | 68 | 91 | 0 |
| | MD3-S20C | 4 | 48 | 56 | 0 | 59 | 91 | 0 |
| | MD4-S20C | 7 | 56 | 93 | 0 | 68 | 88 | 0 |
| | Average | 8 | 50 | 85 | 0 | 67 | 90 | 0 |

## S2 Supplementary Figures

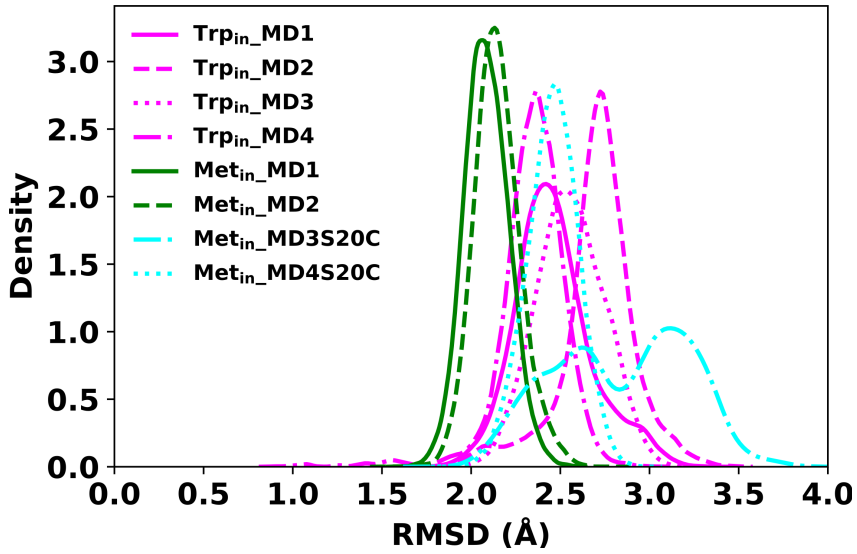


Fig. S1 Stability of $Met_{in}$ trajectories over $Trp_{in}$ ones. Distribution of RMSD values calculated between the $Trp_{in}$ (magenta), $Met_{in}$ (green) and $Met_{in}$-S20C (cyan) MD simulations, calculated from the respective crystal structures. All replicas are reported. RMSD values were calculated on backbone atoms coordinates of snapshots extracted from the production trajectory every 100 ps.

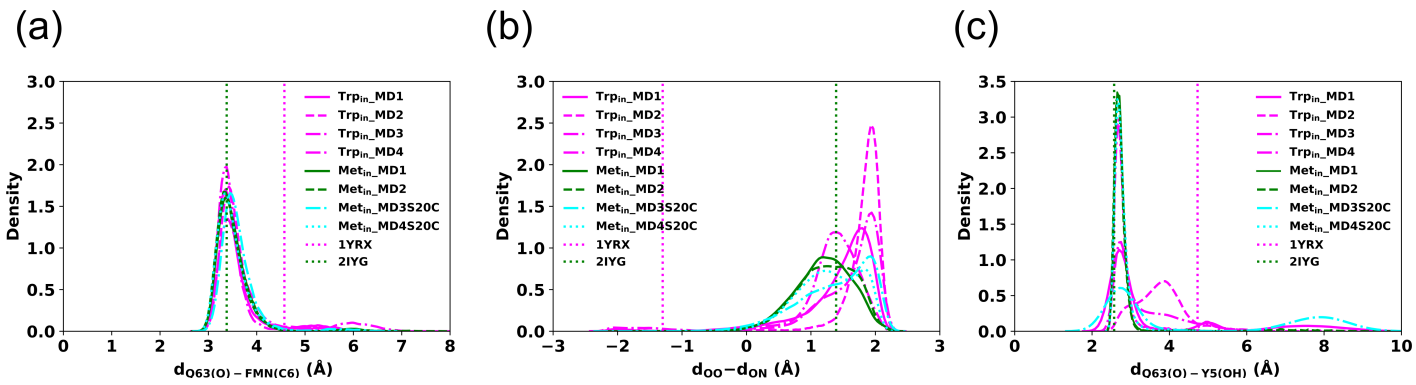


Fig. S2 Distributions of the distances detecting Gln 63 motion along the MD simulations. (a) Distribution of the distance between the oxygen of Gln63 and C6 of FMN. (b) Distribution of the difference between FMN(O4)-Gln63(O) distance denoted as $d_{OO}$ and FMN(O4)-Gln63(N) distance denoted as $d_{ON}$. (c) Distribution of the distance between the oxygen of Gln63 and the hydroxy group of Tyr21. Trp$_{in}$, Met$_{in}$, Met$_{in}$-S20C simulations are shown in magenta, green and cyan, respectively. The distances obtained from the Trp$_{in}$ (1YRX) (magenta) and Met$_{in}$ (2IYG) (green) crystal structures are represented as dotted lines.

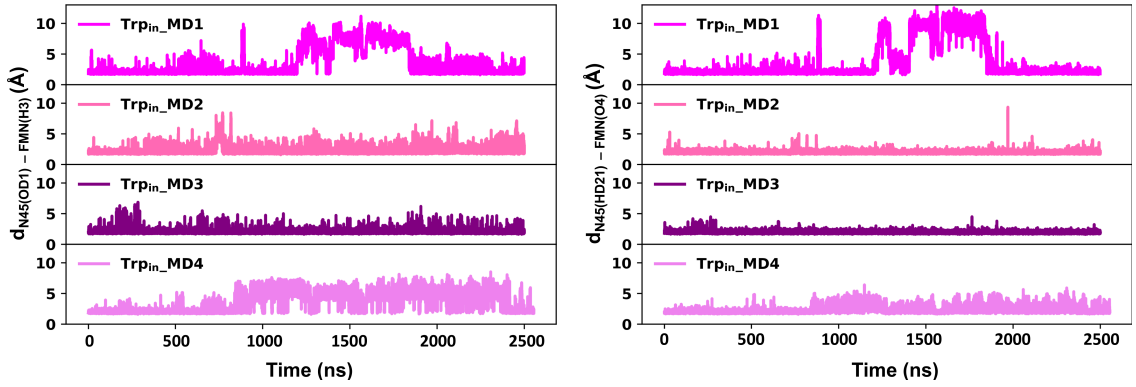


Fig. S3 Conformational dynamics of residues in the binding pocket during the Trp$_{in}$ simulations (pink hues). (a) Time evolution of the N45(OD1)-FMN(H3) distance. (b) Time evolution of the N45(HD21)-FMN(O4) distance.

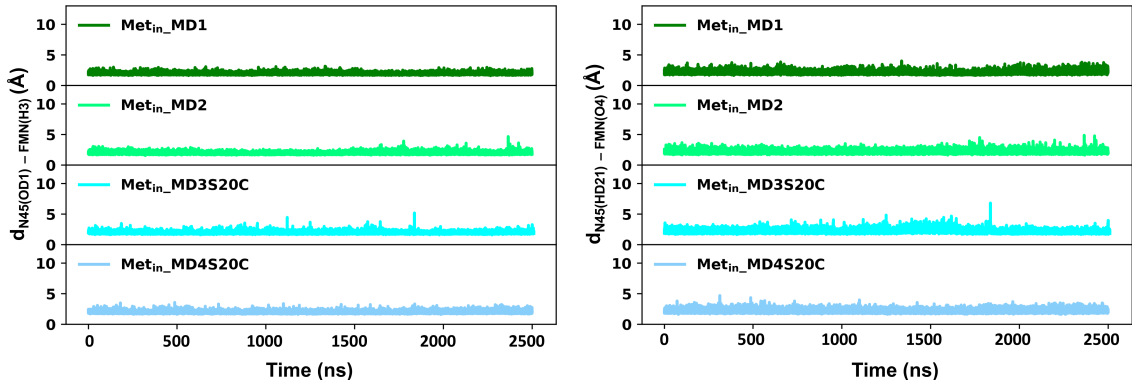


Fig. S4 Conformational dynamics of residues in the binding pocket during the Met$_{in}$ (green hues) and Met$_{in}$-S20C simulations (cyan hues). (a) Time evolution of the N45(OD1)-FMN(H3) distance. (b) Time evolution of the N45(HD21)-FMN(O4) distance.

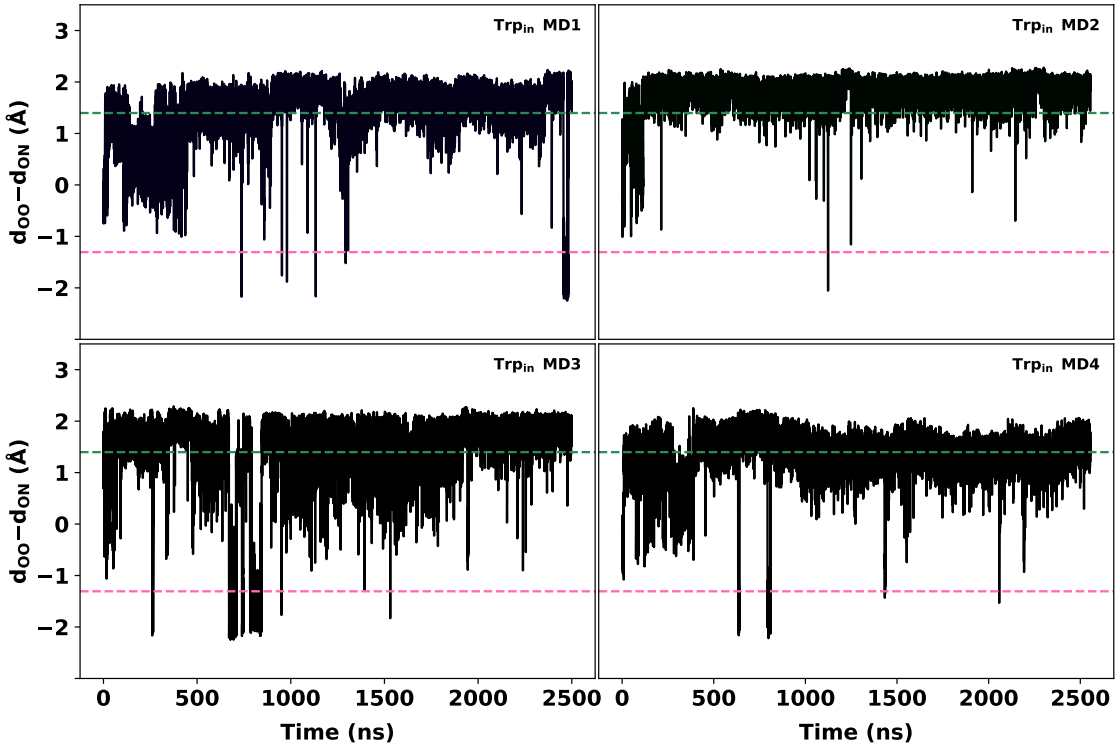


Fig. S5 Orientation of Gln63 along four Trp$_{in}$ replicas as measured by the difference between FMN(O4)-Gln63(O) distance $d_{OO}$ and FMN(O4)-Gln63(N) distance $d_{ON}$. The values corresponding to the two available crystal structures are shown as horizontal dotted lines (2IYG in green and 1YRX in pink).

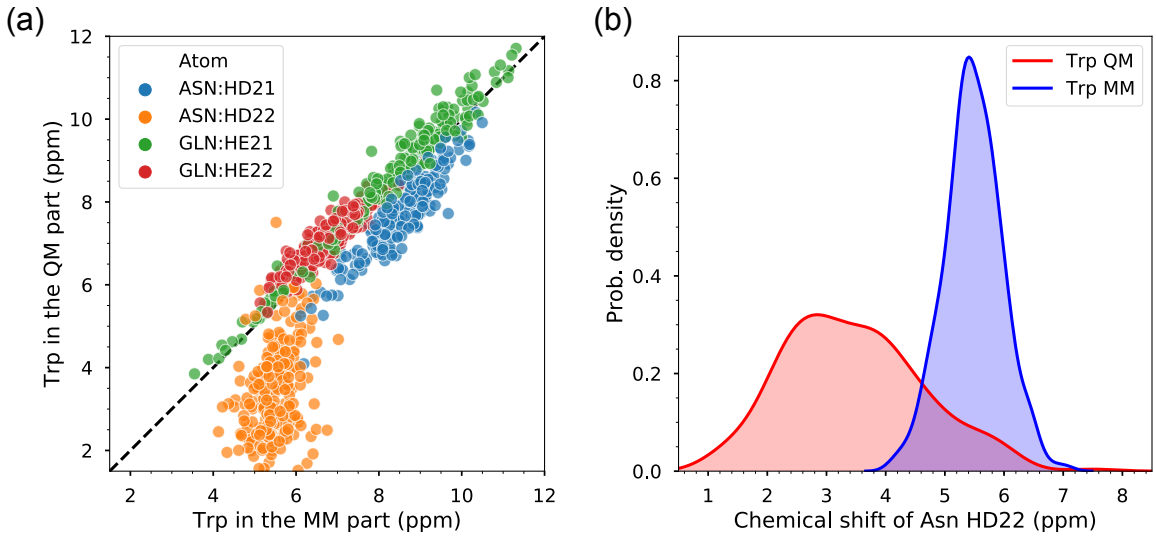


Fig. S6 Effect of Trp104 on the chemical shifts of Asn45 and Gln63 amide protons. (a) Scatter plot of amide protons chemical shifts calculated with the full-QM cluster and with Trp104 excluded from the QM part. The dashed $x = y$ line serves as a guide to the eye. (b) Distribution of chemical shifts for the Asn45 HD22 atom, with Trp104 included (red) or excluded (blue) from the QM part.

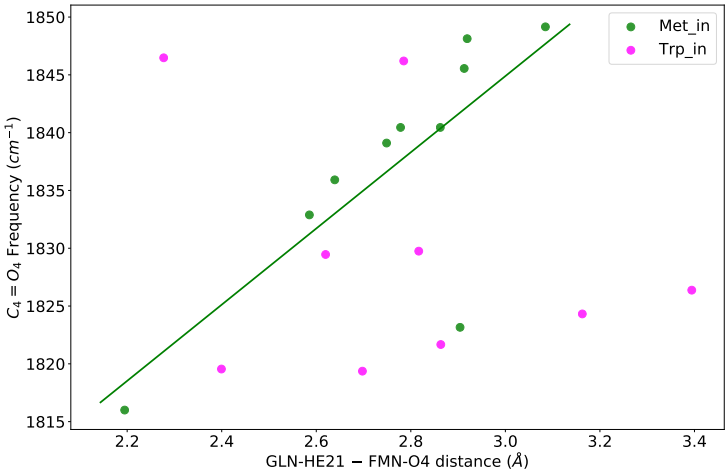


Fig. S7 Correlation between $C_4$=$O_4$ QM/MM harmonic frequencies and hydrogen bond distance from Gln63 amino hydrogen to the FMN $O_4$ atom.

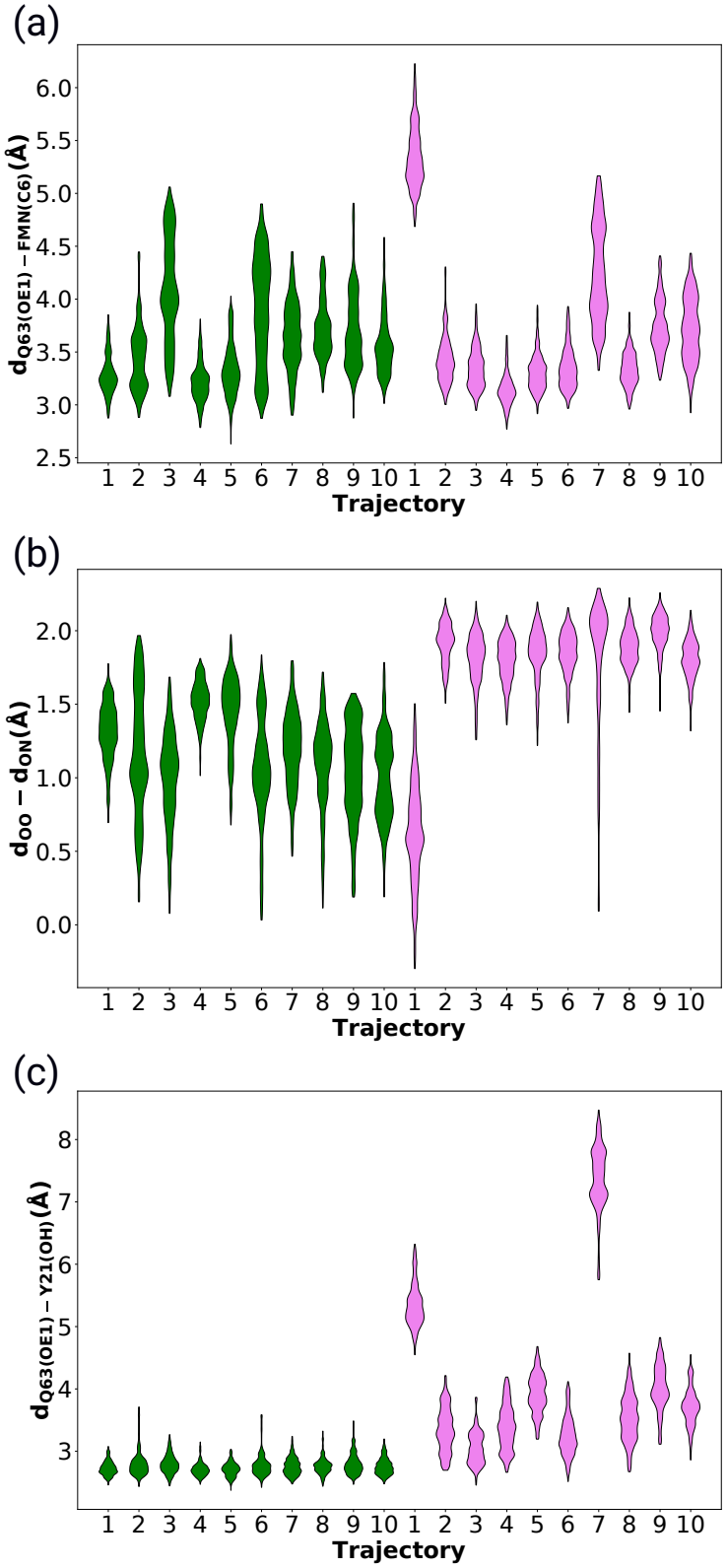


Fig. S8 Distributions of the distances detecting Gln63 motion along the $Trp_{in}$ (magenta) and $Met_{in}$ (green) QM/MMPol MD simulations. (a) Distribution of the distance between the oxygen of Gln63 and C6 of FMN. (b) Distribution of the difference between FMN(O4)-Gln63(O) distance denoted as $d_{OO}$ and FMN(O4)-Gln63(N) distance denoted as $d_{ON}$. (c) Distribution of the distance between the oxygen of Gln63 and the hydroxyl group of Tyr21. The atom numbering is reported in Fig.3

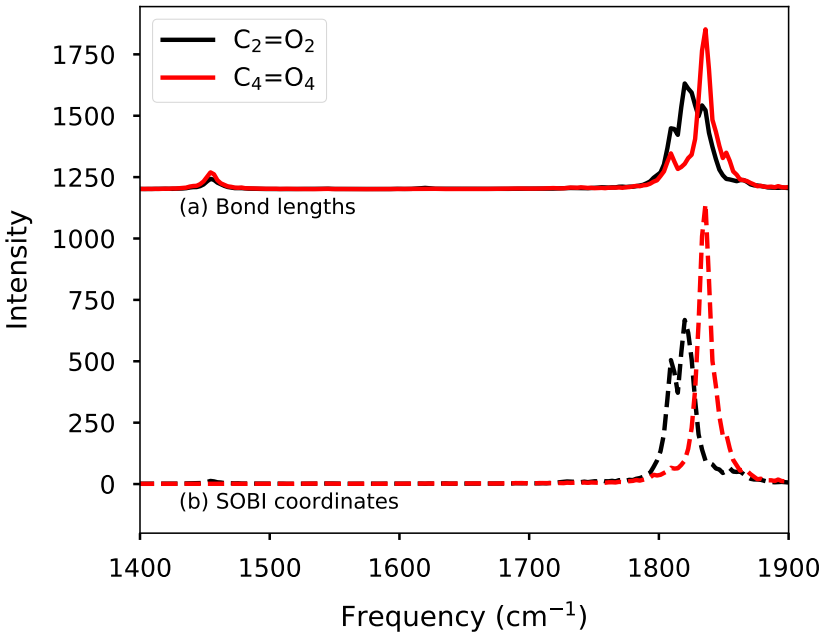


Fig. S9 Mode resolution with SOBI independent component analysis on one of the QM/MM MDs. The power spectra of the $C_2$=$O_2$ (black) and $C_4$=$O_4$ (red) bond lengths are shown as solid lines, and offset by 1200 on the $y$ axis for clarity. The power spectra of the SOBI coordinates with the largest contribution from $C_2$=$O_2$ (black) or $C_4$=$O_4$ (red) are shown as dashed lines.

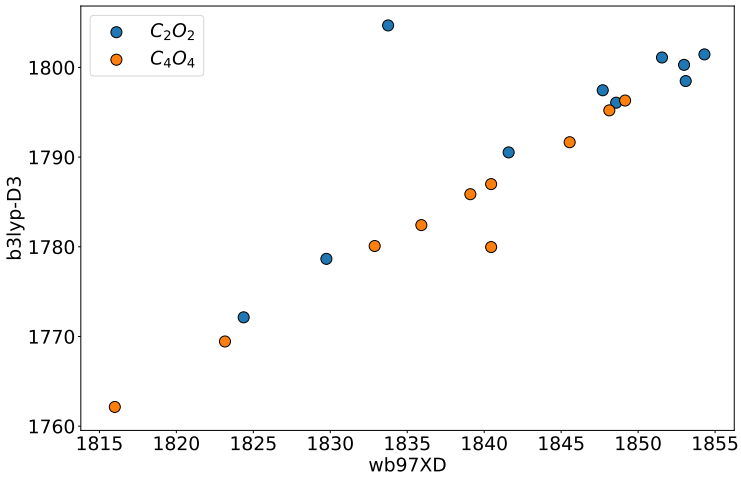


Fig. S10 Comparison between $C_2$=$O_2$ and $C_4$=$O_4$ harmonic IR frequencies calculated with b3lyp-D3 and wb97XD functionals. Frequencies were calculated on ten QM/MM optimized structures corresponding to $Met_{in}$ conformation. For all calculations we used the 6-31G(d) basis set.

## S3 QM/MMPol simulations: computational details

QM/MMPol simulations were carried out using an interface[1–3] between the molecular dynamics package Tinker[4,5] and the development version of Gaussian[6,7] which implements an embedding based on the AMOEBA force field.[8,9] For these simulations, initial configurations were extracted from the classical molecular dynamics trajectories: ten of them from the $Trp_{in}$ dynamics and ten of them from the $Met_{in}$ dynamics. For each of the selected frames, we selected a shell containing only the protein and solvent residues within 30 Å from the FMN residue to be used for the QM/MMPol simulations. The FMN isoalloxazine ring was treated using a QM method, whereas the remaining part of the system was described using the AMOEBA force field. As QM level of theory we chose the DFT $\omega$b96XD/6−31G(d)). This functional shows a similar trend (see figure S10 to b3lyp-D3 for $C_2O_2$ and $C_4O_4$ ONIOM QM/MM harmonic frequencies calculated on $Met_{in}$-conformation structures. Non periodic boundary conditions were included by freezing all the residues beyond 22 Å from the FMN, moreover, to speed up the calculation, we set to zero all the AMOEBA polarizabilities of the frozen residues. To generate suitable starting condition for the molecular dynamics simulations, on each of the structures, we first performed a QM/MMPol geometry optimization using the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm.[10] Then, QM/MMPol simulations in the NVT esemble were classically propagated for 10 ps, using an integration step of 0.5 fs and the Velocity Verlet algorithm. A constant temperature of 300 K was kept using the Bussi thermostat.[11]

## S4 Frequency extraction from QM/MM MD simulations

Vibrational spectra can be calculated from MD simulations via autocorrelation functions. Power spectra of atomic velocities, *i.e.* the Fourier transform of the velocity autocorrelation function, contain all the vibrational frequencies of the system[12]. While for simple systems the velocity power spectra can be analyzed and decomposed in terms of effective normal modes[13], our QM/MM simulations cannot be easily analyzed in this way. Instead, we used internal coordinates (bond lengths and angles) to compute power spectra that are directly related to specific normal modes. Example of power spectra for the $C_2$=$O_2$ and $C_4$=$O_4$ bond lengths are shown in Figure S9 (solid lines). Even though the power spectra are quite localized in frequency around 1800 $cm^{-1}$, a smaller contribution appears at ∼1450 $cm^{-1}$, due to the mixing of C=O modes with ring modes. In order to obtain effective normal modes, we employed linear combinations of internal coordinates.

Assuming that each internal coordinate evolves as a combination of independent signals (normal modes), we can use a SOBI, a signal-processing algorithm, to disentangle the contributions from various normal modes. The independent components $s(t)$ can be written as linear combinations of the internal coordinates $\mathbf{x}(t)$:

$$s(t) = A\mathbf{x}(t)$$

The SOBI algorithm relies on the property that independent signals have negligible cross-correlation functions at all time delays. Therefore, independent components should satisfy the relation

$$\langle s_i(t) s_j(t+\tau) \rangle = 0 \qquad i \neq j$$

for different time delays $\tau$. By defining the time-lagged correlation matrix,

$$R_\tau = \langle \mathbf{x}(t) \mathbf{x}(t+\tau)^T \rangle$$

the above condition is equivalent to imposing the diagonality of the matrix $AR_\tau A^T$. In practice, one needs to find $A$ such that $AR_\tau A^T$ is approximately diagonal for several values of $\tau$. We use a joint diagonalization algorithm based on Jacobi angles[14] in order to find the best transformation $A$. Correlation matrices were built at lag times 50, 100, ..., 500 fs and then averaged over all QM/MM MD trajectories. In this way, a single common transformation matrix was obtained for all trajectories. This ensures that exactly the same normal-mode coordinates are used to construct the power spectra.

The power spectra of the SOBI-transformed coordinates are shown at the bottom of Figure S9 (dashed lines) for an example trajectory. As it can be seen, the SOBI algorithm is able to resolve the two frequencies, and to remove the small contribution at ∼1450 $cm^{-1}$.

## References


1 D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2017, 13, 4025–4033.

2 M. Nottoli, B. Mennucci and F. Lipparini, *Phys. Chem. Chem. Phys.*, 2020, 22, 19532–19541.

3 M. Nottoli, M. Bondanza, F. Lipparini and B. Mennucci, *J. Chem. Phys.*, 2021, 154, 184107.

4 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, 14, 5273–5289.

5 L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder and J.-P. Piquemal, *Chem. Sci.*, 2018, 9, 956–972.

6 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, A. V. Marenich, M. Caricato, J. Bloino, B. G. Janesko, J. Zheng, 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. J. A. Montgomery, 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 Development Version, Revision J.13*, 2020, Gaussian, Inc., Wallingford CT, 2020.

7 F. Lipparini, *J. Chem. Theory Comput.*, 2019, 15, 4312–4317.

8 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, 107, 5933–5947.

9 J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, 114, 2549–2564.

10 D. C. Liu and J. Nocedal, *Math. Program.*, 1989, 45, 503–528.

11 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, 126, 014101.

12 M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, 15, 6608–22.

13 M. Martinez, M.-P. Gaigeot, D. Borgis and R. Vuilleumier, *J. Chem. Phys.*, 2006, 125, 144106.

14 J.-F. Cardoso and A. Souloumiac, *Siam J. Matrix Anal. Appl.*, 1996, 17, 161–164.