How fine-tuned for energy transfer is the environmental noise produced by proteins around biological chromophores?

Kirsten Claridge, Daniele Padula,* Alessandro Troisi*

Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, U.K.

E-mail: DP dpadula@liverpool.ac.uk, AT atroisi@liverpool.ac.uk

Abstract

We investigate the role of the local protein environment on the energy transfer processes in biological molecules, excluding from the analysis the effect of intra-chromophore nuclear motions, and focussing on the exciton-phonon coupling. We studied three different proteins (FMO and two variants of the WSCP protein) with different biological functions but similar chromophores, to understand whether a classification of chromophores based on details of the environment would be possible, and whether specific environments enhance or suppress the coupling between exciton and protein dynamics. Our results show that despite the different biological role, there is no significant difference in the influence of the environment on the properties of the chromophores. Additionally, we show that the main role in influencing molecular properties is played by solvent molecules: the interaction occurs on a medium-range scale, and the solvent is kept in place by a strong H-bond network being free to rotate, suggesting a dipole-dipole interaction mechanism. Steric hindrance exerted by other moieties can help modulating the interactions and tuning the energy transfer process. Overall, considering also the relatively greater importance of intra-molecular nuclear motions, the protein environment around biological chromophores does not appear fine-tuned for a specific function.

Introduction

The light harvesting pigment protein complexes (PPCs) are a crucial component of the photosynthetic apparatus of plants, bacteria and algae.[1] Excitation energy transport (EET) of absorbed photon energy is conducted by these complexes with extremely high quantum efficiency.[2, 3] In addition, two-dimensional electronic spectroscopy (2D-ES) experiments have provided evidence of long-lived coherence in PPCs,[4-8] with more recent data apparently challenging this view.[9] It is however generally accepted that the interaction between the chromophores and protein surroundings is key in determining EET, as it controls the coherence of the quantum evolution of the system.[6, 10-12] Understanding the nature of the electron-nuclear interaction will elucidate the role of the PPC environment, and can help determine whether it is finely tuned to promote EET. It was in fact noted that a too strong or too weak coupling with the environment could both be detrimental for exciton transport[11, 13-15] and, identifying the presence of regions of the protein that amplify or reduce the exciton-nuclear coupling, would indicate whether the environment is adjusting the decoherence to some beneficial level. Conversely showing that no such interactions exist signifies the environment behaves as a glassy low dielectric medium as commonly assumed by biophysicists.[16, 17]

A typical method to study the exciton-environment interaction is by computing the spectral density of the chromophores. Thus, much work has focused on properly obtaining spectral densities of the pigments in various PPCs, for example the Fenna-Matthews-Olson complex(FMO),[18-20] phycoerythrin 545 (PE545)[14, 15] and light harvesting complex 2 (LH2).[21] A common approach is using molecular dynamics (MD) based methods in combination with quantum chemistry calculations to compute the spectral density. In this approach the autocorrelation function, C_i, of the excitation energy of chromophore *i*, is obtained by computing the fluctuation of the excitation energy, $\delta\varepsilon_i$, along the MD trajectory. A Fourier transform of the autocorrelation function then produces the spectral density, *J(ω)*.

$$C_i(t) = \langle \delta\varepsilon_i(t)\delta\varepsilon_i(0)\rangle \quad (1)$$

$$J(\omega) = \frac{\beta\omega}{\pi}\int_0^\infty C_i(t)\cos(\omega t)\,\mathrm{d}t \quad (2)$$

In this paper we propose the study of environmental effects on a *rigid* chromophore as an ideal method to explore the role of the environment and apply this method to all chromophores of a number of PPCs. This work was designed to identify specific local chemical environmental vibrations in LHC. However, we will show that such a specificity cannot be identified, at least for the presented cases. By freezing the intramolecular modes we achieve several advantages: firstly, the modes contributing to the total spectral density arising from intramolecular vibrations obscure those contributed by the environment and so by freezing the chromophore these obscuring vibrations are removed. In addition, a number of recent papers[22-25] have shown that for FMO the high frequency part (> 500 cm^{-1}) of the spectral density plays a small role in the exciton dynamics, as these modes have much larger vibrational energy than the excitonic transition energy. The motions characterised by this higher frequency part predominantly arise from the intramolecular vibrations. Furthermore, the intramolecular modes are a main source of inaccuracy in computed spectral densities.[9, 10, 13, 15, 18, 26] This is due to the problematic treatment of quantum vibrations in a classical framework[18, 25] and the so-called geometry mismatch,[26, 27] an error resulting from inconsistencies in the levels of theory used in the computation of the correlation function. The trajectory is obtained classically (typically) on the ground state surface, the resulting equilibrium geometries of which are governed by the force-field used, whilst the excitation energies are computed using quantum methods, including a classical atomistic description of the environment (*e.g.* point charges). Consequently, there is a mismatch between the geometries of the MD equilibrium structures and those of the quantum chemistry calculations. These inaccuracies can be removed by using refined approaches to parametrize force fields consistent with the DFT potential[28-30] or by treating vibrations in a quantum framework by normal modes analyses or sophisticated MD approaches.[25, 31-33] Other studies propose a separate treatment of intra- and inter-molecular vibrations, and finally they consider their interplay adopting projection techniques.[18] Here we propose, as a convenient alternative, to keep the chromophores frozen. This solution allows to isolate the effect of environmental motions on the fluctuation of excitation energies from the other main effect of changing the chromophores geometry and energy levels.[34] The separation is justified considering that, at room temperature, the RMS oscillations of the heavy atoms of a chromophore around their equilibrium positions, evaluated from normal modes computed in

a QM/MM scheme,[25] are lower than 0.20 Å and that the intramolecular modes have larger characteristic frequencies.[25] Keeping the chromophore frozen to isolate the effect of the environment is a procedure very common in the study of electron transfer[35-37] where it is equivalent to assuming additivity of internal and external reorganization energy.[38] The effect of the environment on the absolute (average) position of energy levels, instead, has been the object of several detailed studies.[13-15, 18, 19] Here, we aim at studying the effect of the environment exclusively on the fluctuations of the excitation energies with the aim to identify specific interactions that affect the low-frequency part of the spectral density.

Analysis of this spectral density calculated by utilising this frozen chromophore MD simulation gives information on the important modes of the environment in the modulation of excitation energies. Examining multiple chromophores embedded in locally different environments allows the determination of any difference/similarities of environmental fluctuations and thus, the ability to identify potentially important modes and/or structural features. Such details provide useful information for the engineering of materials with improved energy transport. Furthermore, through investigating PPCs with different functions it can be determined if there is any structure-function relationship present. The aim of this work is to provide chemical insight into the environment-chromophore interaction, by removing the intramolecular contributions and analysing the impact of environmental motions on the fluctuation of excitation energies. By freezing the chromophores, we also achieve an effective way to exclude errors in the spectral density resulting from the geometry mismatch, and to allow the comparison/classification of different chromophores based on the nature of their interaction(s) with the environment. To achieve this, we analysed a set of different proteins with similar chromophores, which thus result embedded in slightly different chemical environments. The analysis of a range of systems is advantageous because it helps ruling out accidental correlations that might be identified by analysing only a more modest amount of data, or to highlight patterns that were not arising frequently in other cases.

Systems

In this work we considered three systems, summarised in Table 1. These are the Fenna-Mattews-Olson complex (FMO) and two water soluble chlorophyll binding proteins, namely the chlorophyll-a binding protein (WSCP-a) and the chlorophyll-b binding protein (WSCP-b). FMO is one of the most studied proteins in this field, due to its crystal structure being available for many years[39] and it being the system for which long-lived coherence was initially observed.[4] Inclusion of the WSCPs in the study allows the investigation of a small range of chromophores. In addition, these proteins serve different functions: FMO is involved in light harvesting whilst the WSCPs are simply chlorophyll transporters, and so it is possible to comment on the structure-function relationships or determine if no such relationship exists.

The FMO protein is a trimer containing 24 chromophores, thus 8 per monomer unit. Each unit has 7 chromophores inside the monomer protein scaffold, with the eighth chromophore sandwiched between two adjacent monomer units. The WSCPs each contain 4 chromophores, organised as 2 adjacent parallel displaced pairs.

Table 1: Proteins and chromophores analysed in this work.

Protein	Chromophores	Abbreviation	# of Chromophores
FMO	Bacteriochlorophyll-A	BCL	8
WSCP-a	Chlorophyll-A	CLA	4
WSCP-b	Chlorophyll-B	CLB	4

Computational Details

We obtained the crystal structure of FMO and WSCP from the Protein Data Bank (PDB: 3BSD and 5HPZ, respectively).[40] Missing residues at the edges of the protein chain have been ignored, while we added missing atoms using the CHARMM-GUI website.[41] We prepared the systems for the simulations with the GROMACS 5.0.5 software.[42] The proteins were embedded in cubic boxes with side 120 Å and 110 Å for FMO and WSCP, respectively. Histidine residues have been assigned the appropriate protonation state to allow coordinating Mg atoms of chromophores, otherwise they have been assigned the proton in position ε. We added water molecules and we set the ionic strength of the starting box to 150 mM by adding the appropriate number of potassium and chloride ions. The system has been described using the TIP3P model for water molecules, the CHARMM36 force field for the protein,[43] and literature parameters for the chromophores.[44, 45] We initially minimised the system keeping all chromophores frozen, with 2000 steps of Steepest Descent. For all the following steps, we used an integration step of 2 fs and constrained all the bonds with the LINCS algorithm implemented in GROMACS.[46] We then equilibrated the system in two steps of 500 ps each, run in NVT (heating up to 300K) and NPT conditions (Berendsen barostat), respectively. Finally, we ran a separate equilibration of several nanoseconds for each frozen chromophore, keeping it frozen in the geometry obtained from X-ray experiments, which thus includes, at least to a first order, the effect of the surroundings on the geometry.

We used the last 16 ps of the trajectory to compute the excitation energies every 20 fs, using a QM/MM scheme within TDDFT linear response theory.[47] More in detail, the environment is treated as classical point charges that affect the QM system additively, *i.e.* its effect on the QM part is included in the Fock Matrix through an electrostatic contribution. We performed a preliminary study concerning the fluctuation of the excitation energies as a function of the boundary chosen for the MM region, using a set of 50 snapshots extracted from the last 5 ns of the trajectory at 100 ps intervals. We found that an MM radius of at least 30 Å around the chromophore is needed to neglect this uncertainty (see Fig. S1 in Supp. Information), which is consistent with similar considerations in the literature.[48] Additionally, we found a high correlation between 6-31G* and 3-21G* data, with a slope close to 1 ($R^2 \approx 0.98$, see Fig. S2 in Supp. Information), meaning that the standard deviations at the two levels of theory are comparable. The correlation between the excitation energies with the two basis sets is higher than the correlation between excitation energies computed using different MM radiuses ($R^2 > 0.98$ only for radiuses bigger than 45 Å, see Fig. S3 in Supp. Information), thus we decided to use the 3-21G* basis set as a good compromise between accuracy and computational cost. We also explored 4 different functionals (B3LYP,[49] CAM-B3LYP,[50] M06-2X,[51] ωB97X-D[52]), observing a consistent behaviour ($R^2 > 0.90$) among CAM-B3LYP, M06-2X, ωB97X-D, meaning that our results are independent of the choice of the functional (see Fig. S4 in Supp.

Information). Our final choice for the QM/MM calculations thus resulted in the ωB97X-D/3-21G* combination, including all the residues within a 35 Å radius around the QM chromophore, an approach similar to the ones used, for example, in refs [13, 15]. For QM/MM calculations, the number of atoms of the chromophores in the QM region has been reduced by inserting a link atom between the first and second Carbon atoms of the Phythyl chain. We used the QChem 4.2 software for the QM/MM calculations.[53] Fourier Transforms have been computed over 800 points, obtained on a 16 ps time window at regular intervals of 0.02 ps.

Results and Discussion

We studied the fluctuations of the excitation energy for the first excited state of each chromophore along a short MD trajectory. The objective is to retrieve the spectral density of each chromophore from the Fourier Transform of the autocorrelation function (ACF) of the excitation energy. As we sample the excitation energy for 16 ps intervals for each chromophore, the spectral density will represent the local configuration explored in that interval. In a recent work we showed that the spectral density evolves slowly over a nanosecond timescale – well beyond what would be computable as a time series of excitation energy – as the protein samples different local environments.[25] These motions are effectively stationary compared to the exciton dynamics and thus do not influence them. Instead the exciton dynamics are dictated by a specific spectral density within a limited time interval rather than the spectral density averaged over very long times.[54] Additionally, if interactions and/or environmental fluctuations are optimised specifically to influence exciton dynamics, then such interactions would be present in any sampled trajectory interval. For completeness, we report in the SI (see Fig. S8) the differences among spectral densities evaluated at different points of the trajectory.

Since we adopted a frozen chromophore scheme, the fluctuations are exclusively due to environmental motions. To identify differences among the various chromophores analysed, we report the statistics associated with our QM/MM calculations in Table 2, and the data processing, in terms of calculation of autocorrelation functions and their Fourier Transforms, in Figs. 1-2.

Table 2: QM/MM average excitation energies and their fluctuations along the MD trajectory for the chromophores belonging to proteins reported in Table 1.

Chromophore	Avg. (eV)	σ (eV)	Group
BCL367	1.7581	0.0119	1
BCL368	1.7908	0.0123	2
BCL369	1.7839	0.0088	2
BCL370	1.7613	0.0072	2
BCL371	1.7735	0.0073	1
BCL372	1.7972	0.0054	1
BCL373	1.7920	0.0089	1
BCL400	1.7710	0.0105	2
CLA1	2.0938	0.0067	2
CLA2	2.1316	0.0072	1
CLA3	2.0270	0.0071	2
CLA4	2.1145	0.0151	2
CLB1	2.1398	0.0070	2
CLB2	2.1130	0.0082	2
CLB3	2.0631	0.0126	1
CLB4	2.0901	0.0150	1

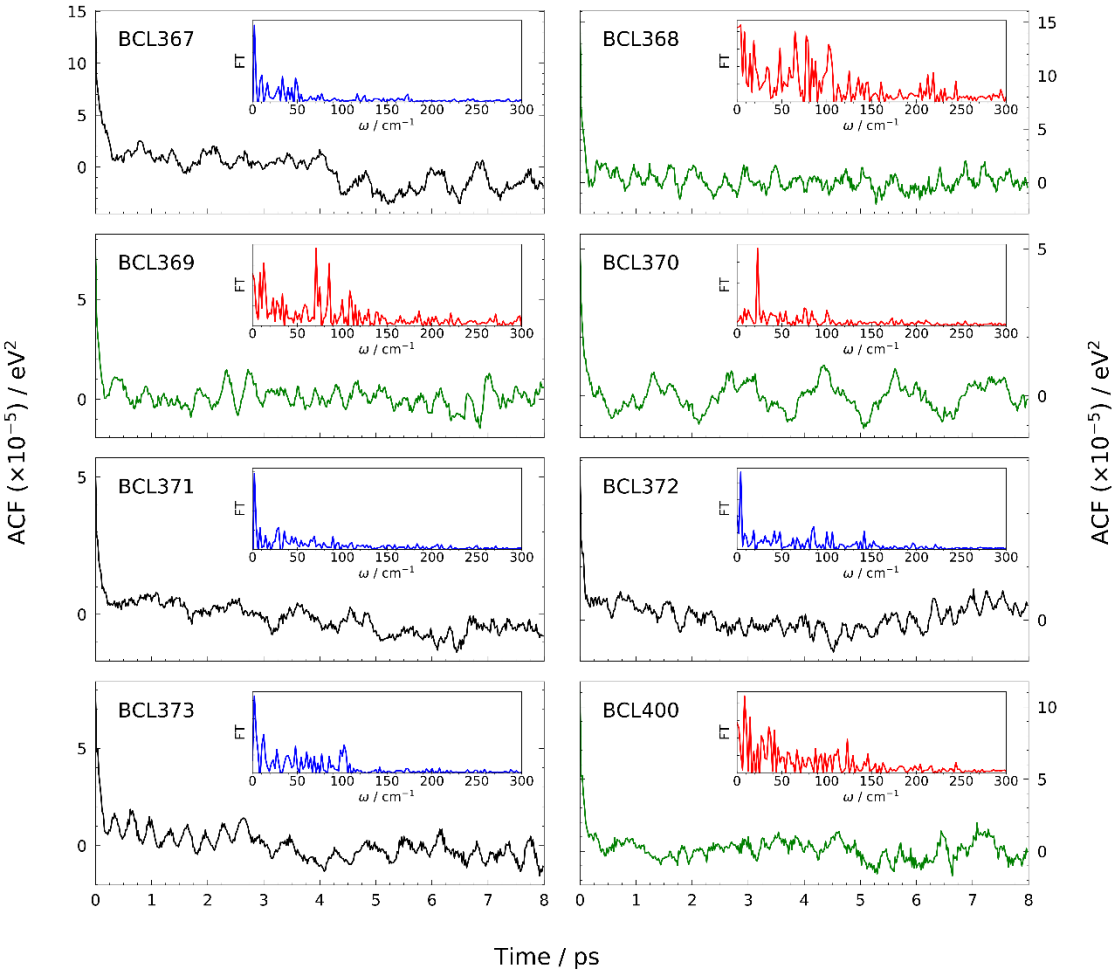


Fig. 1: Autocorrelation functions (black lines for Group 1, green lines for Group 2) and their Fourier Transforms (insets, blue lines for Group 1, red lines for Group 2) for FMO chromophores.

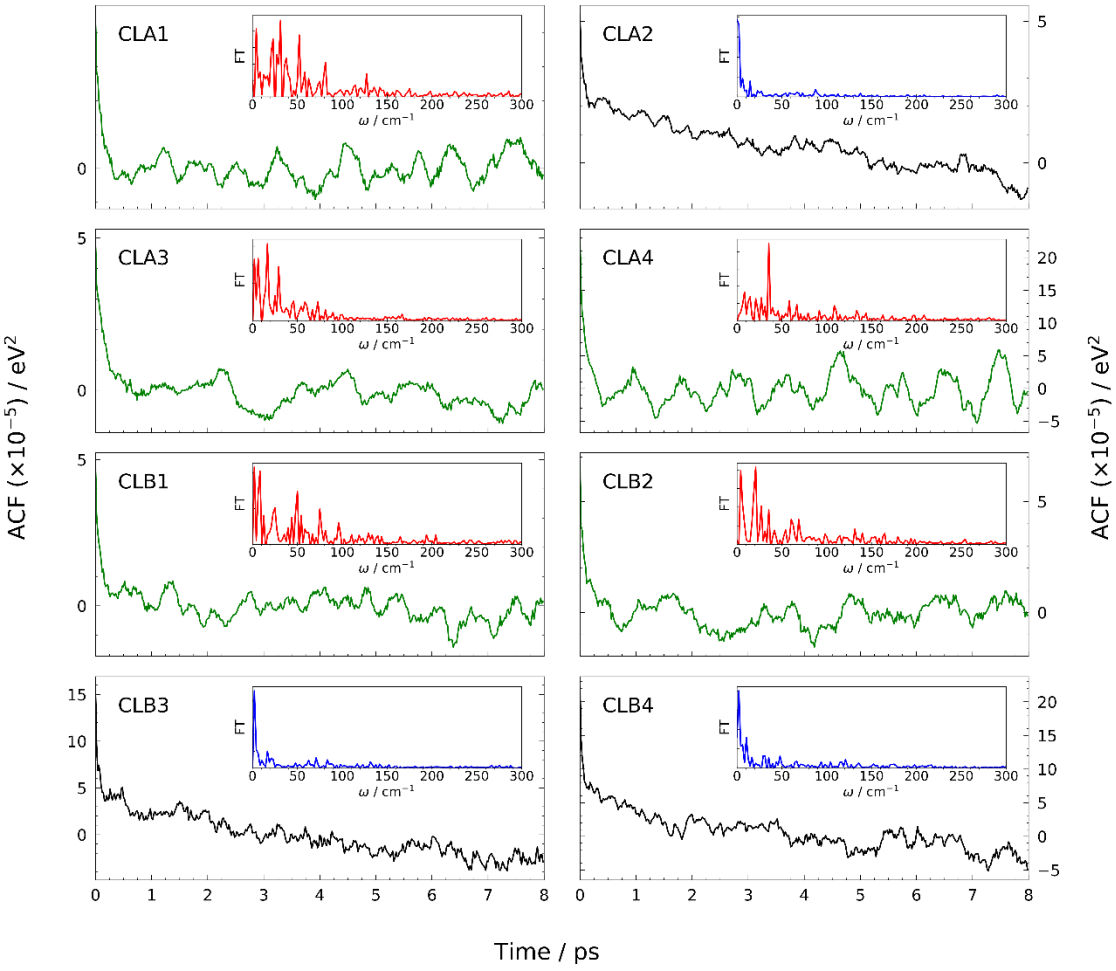


Fig. 2: Autocorrelation functions (black lines for Group 1, green lines for Group 2) and their Fourier Transforms (insets, blue lines for Group 1, red lines for Group 2) for WSCP chromophores.

Looking for patterns in the Fourier Transforms of the ACFs, we noticed that they can be classified mainly into two groups. The first group (Group 1) shows tails in the ACFs, possibly due to long-time protein dynamics that sometimes leads to not fully converged ACFs. For one of the chromophores of this group (BCL 367), we ran a longer simulation to compute a better ACF, but we noticed no significant difference obtained with such extension. For this reason, we did not repeat the calculations on a longer timescale for the other chromophores. The FTs show a strong peak only at very low frequencies (ω < 10 cm^{-1}). This group contains half of the chromophores of FMO (BCL 367, BCL 371, BCL 372, BCL 373), and chromophores CLA 2, CLB 3 and CLB 4 of the two variants of WSCP protein. The motions characterising this group are significantly slower than the exciton dynamics, and act as inhomogeneous broadening: a sample contains instances of the protein at different conformations, and to a good approximation, the experiment can be modelled assuming the system is static along these degrees of freedom. Note that, due to the considered time window, the reliability of vibrations at ω < 5 cm^{-1} might be questionable, as they oscillate over a larger time window than the one over which we computed the FTs. However, their role in terms of reorganisation energy is negligible, and they act, as said previously, as inhomogeneous broadening. It is worth noticing the discrepancy between the results for CLA 2 and the other CLA chromophores, which is quite surprising in light of the symmetry of the WSCP system. We assign this effect to a symmetry-breaking effect due to the specificity of local solvent interactions described in greater detail in Figs. S10-S11 of the Supp. Info. The importance of the solvent will be highlighted again later on.

The second group (Group 2) shows easily identifiable periodicities in the ACFs, which lead to one or more well defined strong peaks at frequencies on a significant time scale (ω > 10 cm^{-1}), in terms of lifetime of the exciton. It contains the remaining chromophores of FMO (BCL 368, BCL 369, BCL 370, BCL 400) and WSCP (CLA 1, CLA 3, CLA 4, CLB 1, CLB 2) proteins. These motions are coupled in a complex way with the exciton dynamics. The classification in these two groups is consistent with similar observations in the literature, where a neat separation among static and dynamic disorder has been identified.[28] We stress, however, that in the present case, the motions being investigated belong exclusively to the environment, due to the fact that the excited chromophore is frozen along the classical trajectory. The standard deviation of the excitation energy due to the *intra*-chromophore motion can be evaluated from the internal reorganization for exciton transfer. The results of ref. [25] concerning all chromophores of FMO indicate that the fluctuation of the excitation energy due to intra-chromophore mode is about 0.020 eV, i.e. almost twice the amplitude of the fluctuations due to the environment. Several studies suggest that fluctuations are necessary for an effective energy transfer to avoid an overly coherent propagation.[11] However, if the magnitude of such fluctuations is excessive, competitive processes like exciton localization would be favoured and reduce the efficiency of energy transfer that would then proceed via a series of incoherent hopping events. The amplitude of fluctuations presented here is comparable to other studies.[19, 28] However, the fact that such fluctuations are also comparable over various proteins with different biological functions, suggests that environmental noise is not specifically tuned to promote energy transfer between chromophores. Moreover, the magnitude of the fluctuations reported is comparable with the energy differences among

localised excitations. This observation makes clear how fluctuations could determine the order of energy levels and improve (or impair) level alignment.

We initially checked whether belonging to Group 1 or 2, *i.e.* with or without predominant slow motions, was correlated with the fluctuation of the excitation energy, *i.e.* the overall strength of the exciton-vibrational coupling. We used a Kruskal-Wallis statistical test,[55] to establish that the distribution of σ values observed for the two groups were likely drawn by the same distribution (Kruskal-Wallis-p = 0.60). In other words, there is no correlation between the magnitude of the fluctuations and belonging to a certain Group. This can be explained considering that, in a harmonic approximation framework, each mode contributes to σ proportionally to $\hbar\omega$, accounting for small contributions for slow motions. Another question that can be answered with analogous statistical tests is whether the magnitude of the fluctuations of the excitation energy are specific to a certain protein (and thus somehow related to its biological function), or if a generic proteic environment could be responsible for similar fluctuations. We first verified that, in agreement with our expectations, the fluctuations for the two instances of the WSCP protein belong to the same distribution (Kruskal-Wallis-p = 0.56). We then tested the fluctuations in FMO with the overall data set of WSCP: the result of the test (Kruskal-Wallis-p = 0.81) suggests that the observed fluctuations are not specific to FMO or WSCP, and thus unrelated to the biological function of these proteins. This observation can help to rule out hypotheses according to which these complex structures are the result of natural evolution aimed at optimising energy transport processes, at least in terms of interaction with the local environment.

The next step in our analysis consisted in identifying environmental motions correlated with the fluctuations of the excitation energy, to obtain chemical insight that could aid the design and engineering of materials with improved energy transport properties. Some authors have proposed that specific frequency of environmental motions can be important if they become resonant with specific energy level differences,[19] although there is still no consensus on this issue.[25] In any case, the identification of local molecular vibrations coupled with the chromophore excitation can address the issue of whether the local environment is providing a specific function. To identify specific residues able to modulate the excitation energy, we calculated the Pearson's correlation coefficient (r) between the TDDFT excitation energies of the chromophores and the coulombic interaction energies with each of the residues within 10 Å of the chromophore along the MD trajectory. To compute the coulombic interaction energy we represented the chromophore as a set of point charges obtained as the difference between ground and excited state atomic charges, computed within a Merz-Kollman scheme.[56, 57] As an example we report the results of this analysis for two chromophores, BCL 368 and CLA 3, in Table 3. In the table, we report the residues showing a correlation $r > 0.30$, together with the fluctuations of the coulombic interaction energy (σ_E), and the distance (d) of interacting atoms, together with its fluctuation (σ_d). Additionally, we report the atom names of the interacting atoms, and a classification of the interaction, assigning it to a particular component of the environment (solvent, S, protein, P, or chromophore, C). Solvent molecules cannot be assigned an order number, unlike other residues, and are formally identical and can exchange. However, we will refer to them using the order number of the

simulation. We report in the Supp. Information the complete series of tables (Tables S1-S3) containing the results of this analysis for all chromophores.

Table 3: Correlation analysis for residues surrounding chromophores.

Residue	r	σ_E (eV)	d (Å)	σ_d (Å)	Interaction	Class
			BCL 368			
SOL20272	0.373	2.89E-04	1.95	0.05	OW-MG	S
SOL16680	0.335	3.21E-04	2.19	0.28	HW1-ND	S
SER73	0.318	2.17E-04	1.77	0.19	HG-OBB	P
			CLA 3			
ALA34	0.612	2.06E-04	1.93	0.17	O-HE1	P
ALA33	0.516	1.15E-04	1.91	0.15	O-HN	P

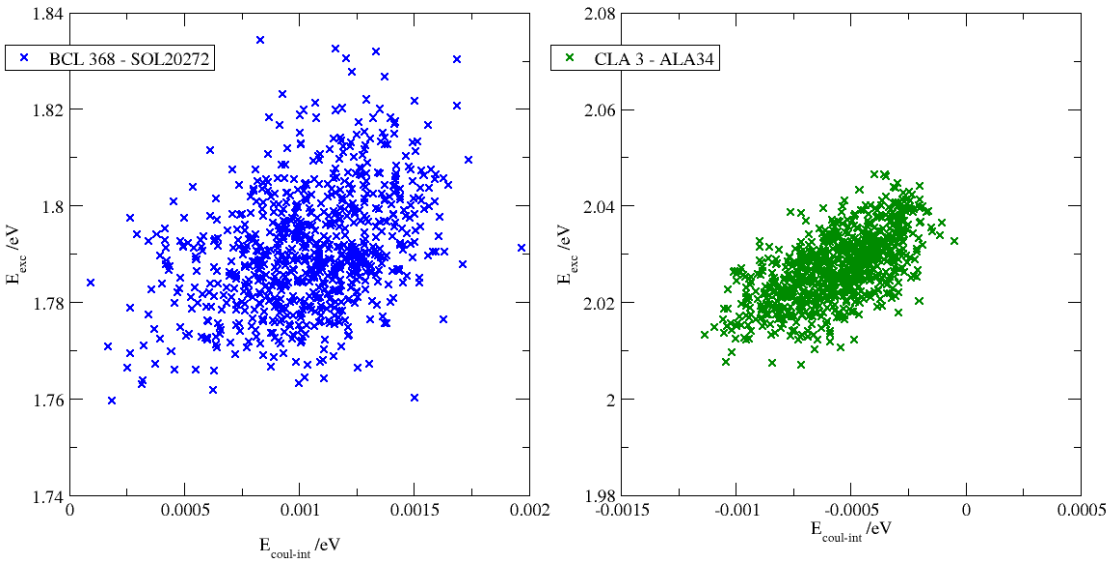


Fig. 3: correlation of coulombic interaction energy of residue and excitation energy of chromophore for SOL20272 and BCL 368 (left) and ALA34 and CLA 3 (right).

To highlight the significant difference between moderate ($0.30 < r < 0.50$) and good ($r > 0.50$) correlations, we report in Fig. 3 the scatter plot of two such correlations reported in Table 3. The residues identified by this correlation analysis play indeed an important role in the chemical environment surrounding the excited chromophore: a visual inspection of the MD trajectories confirms the proximity and involvement in important interactions (*e.g.* hydrogen bonds or coordination) with the chromophore (see Figs. 4 and 6). However, from Fig. 4 it is also clear how some residues located in the proximities of the excited chromophore do not play a significant role in terms of correlation with the fluctuation of the excitation energy. For example, Pro 32 is the coordinating residue of CLA 3, but it shows only a low correlation (r = 0.13) with the fluctuation of the excitation energy, which can be explained considering the conformational rigidity of this amino acid, preventing it from modulating the excitation energy despite its proximity. This demonstrates that the relative importance of the various components of the environment cannot be predetermined by vicinity alone.

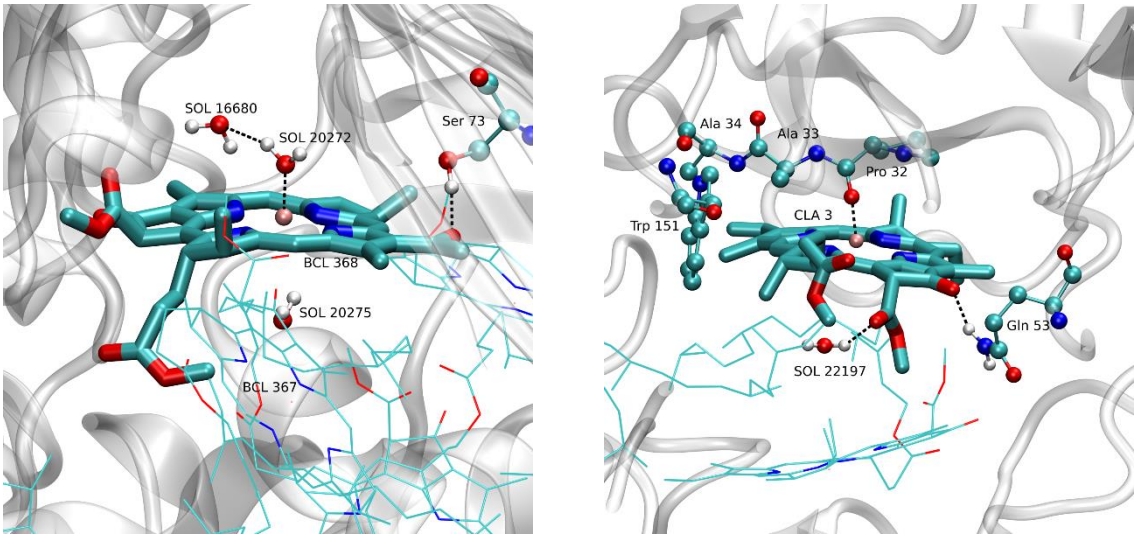


Fig. 4: location of the most correlated residues reported in Tables 3 (BCL 368 left, and CLA 3 right).

The data reported in Table 3 and in Tables S1-S3 in the Supp. Information are summarised in Fig. 5 (see also Fig. S5 in the Supp. Information). They reveal that the strongest influence on the excitation energy is in general due to the solvent, which is responsible for most of the correlations. In Fig. S5, it is evident how the protein is also responsible for a significant fraction of correlations, which are however less important than the ones assigned to the solvent. In Tables S1-3 of the SI, we notice that the most correlated residue is, in general, the one with the biggest fluctuation of the coulombic interaction energy.

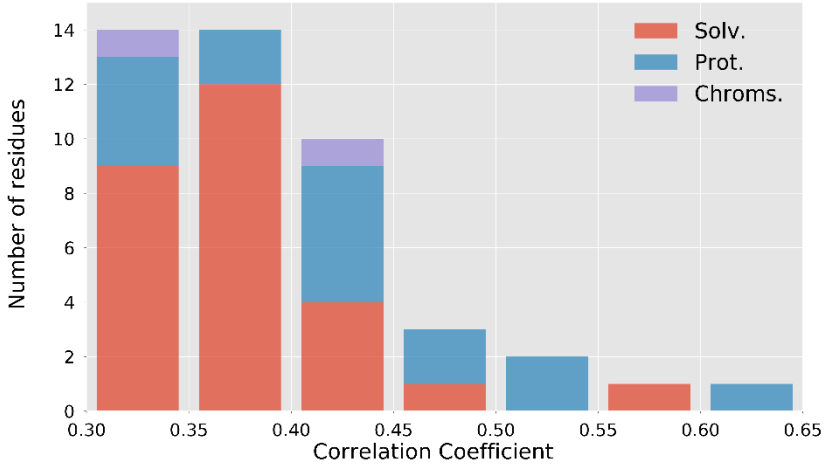


Fig. 5: Distribution of the correlation coefficients for chromophore/environment interactions. In blue we report interactions of the frozen chromophore with the protein, in red with the solvent, and in purple with other chromophores.

To determine if there is any relation between the proximity of the solvent to the chromophore and correlation strength, *i.e.* if the most correlated solvent molecules are those closest to the chromophore, we analysed the distance between the centres of mass of the chromophore and of solvent molecules within 10 Å. For the chromophores of the 2 WSCPs no such relation was found. However, in FMO the most correlated solvent is typically one of the 5 closest solvent molecules to the chromophore, although beyond the most correlated solvent there is no connection between proximity and correlation strength. The same analysis was applied

to the protein residues and there is no relation between proximity and correlation for any of the chromophores in any of the proteins. These facts confirm that the relative importance and effect of a residue cannot be simply determined by proximity to the chromophore alone.

Once we identified the residues most responsible for the fluctuations of the excitation energy, we attempted to identify specific motions of such residues: we computed the FT of the coulombic interaction energies (the same used to calculate *r*), and compared it to the ones coming from TDDFT data. We carried out this analysis for CLA 4 and BCL 370 because their TDDFT Fourier Transforms (see Figs. 1-2) show a well-defined peak in the frequency domain, which could in principle be assigned to a specific motion. However, this comparison proved scarcely useful for analysing specific motions, as the frequency resolved signal also contains the components of many protein motions when a small residue or solvent molecule is considered. The results can be found in the SI in Figs. S6-S7.

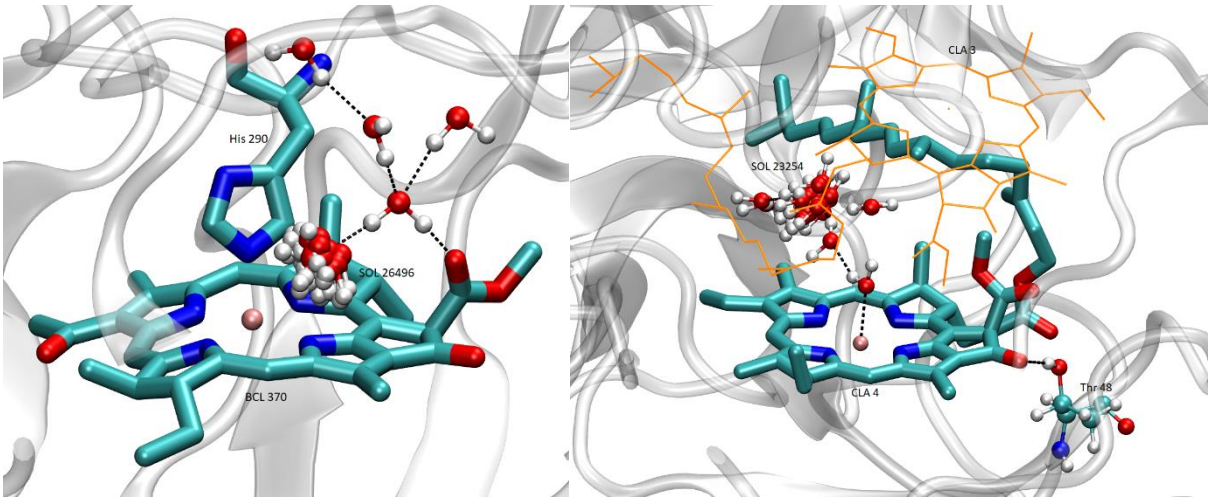


Fig. 6: Left: solvent chain and SOL 26496 motions modulating the excitation energy of BCL 370. Right: solvent chain, containing highly correlated solvent SOL 23254, mediating the interaction of the phytyl chains of CLA 3 (orange) and CLA 4.

For BCL 370 the most correlated residue (*r* = 0.48) is a solvent molecule (SOL 26496). This solvent molecule is involved in an H-bonded network of solvent molecules (see Fig. 6), which prevents it from getting far from the chromophore. However, such molecule is still free to rotate, and this motion modulates the excitation energy of BCL 370. For CLA 4 among the most correlated residue is again a solvent molecule (SOL 23254, *r* = 0.36). Also in this case it does not appear to be directly interacting with the chromophore or protein residues coordinated to the chromophore: instead it is part of an H-bonded chain of solvent molecules.

These two observations suggest that the interaction between the chromophore and solvent molecules is significant also on a medium-range scale. Moreover, despite the fact that the strongly correlated solvent molecules are kept in place by strong interaction networks, they are actually free to rotate, suggesting a dipole-dipole interaction mechanism. The rotation of solvent molecules (*i.e.* the dipole is free to change orientation) makes them able to modulate the excitation energy. For CLA 4, the solvent chain lies between the phytyl chains of CLA 3 and CLA 4, as shown in Fig. 6. While this arrangement is unique to CLA 4, phytyl chains of chromophores in WSCPs have been reported to play a fundamental role in the

photoprotection of the chromophores,[58] and their steric interaction with close solvent molecules may inspire new concepts for molecular design in related fields, such as photocatalysis.

We finally analysed the average composition of the environment around the chromophores, to understand the importance of being in a hydrophilic/hydrophobic pocket. We computed the radial distribution function between the centre of mass of the chromophore and the centre of mass of surrounding residues of a certain type and integrated it to obtain the average number of that particular type of residues along the MD trajectory. The results of this analysis, reported in Table 4, show a homogeneous behaviour among the chromophores. The only exception is BCL400, which is located at the edge of the protein skeleton of FMO and thus more exposed to solvent, but such a difference was expected beforehand. The results of this analysis do not change if we check the composition at different values of the shell radius (see Table S4 in Supp. Information). The most variable component in the environment surrounding the chromophores is the solvent. However, we found no strong correlation between the average fraction of volume due to the solvent in a 10 Å sphere around the chromophore and the fluctuation of the excitation energy along the trajectory (see Fig. S9 in Supp. Information), confirmed by additional correlation tests (Pearson's r = 0.22, Spearman's ρ = 0.30).

Table 4: average number and type of residues in a 10 Å shell from the chromophore.

	σ (eV)	Chrom	Prot	Solv	Tot
BCL367	0.0119	1.00	16.56	5.82	23.38
BCL368	0.0123	0.00	12.88	11.31	24.18
BCL369	0.0088	1.00	15.14	8.19	24.33
BCL370	0.0072	0.00	15.30	13.97	29.27
BCL371	0.0073	0.00	13.06	13.65	26.71
BCL372	0.0054	0.00	15.22	3.26	18.49
BCL373	0.0089	1.00	14.78	9.11	24.89
BCL400	0.0105	0.00	18.15	32.73	50.88
CLA1	0.0067	1.00	14.96	9.91	25.86
CLA2	0.0072	1.00	15.96	11.71	28.68
CLA3	0.0071	1.00	15.43	9.62	26.06
CLA4	0.0151	1.00	16.11	12.41	29.53
CLB1	0.0070	1.00	17.20	7.19	25.39
CLB2	0.0082	1.00	17.72	12.45	31.17
CLB3	0.0126	1.00	16.65	8.28	25.94
CLB4	0.0150	1.00	16.98	14.59	32.57

Conclusions

In this work we studied the fluctuations of the excitation energy of a set of similar chromophores (BCL, CLA, and CLB) embedded in different proteic environments (FMO, WSCP). We adopted an approach based on classical Molecular Dynamics, followed by QM/MM calculations on a number of snapshots extracted from the trajectories. However, to isolate the role of the environment, we kept one chromophore frozen in each simulation. This also allowed us to avoid inaccuracies due to the well-known geometry mismatch problem in a convenient way and to assign the observed fluctuations in the QM/MM excitation energies only to environmental rearrangements. We computed the autocorrelation functions for the QM/MM excitation energies and their Fourier Transforms, identifying two groups of

chromophores: one showing only low frequency (ω < 10 cm^{-1}) contributions (Group 1), and the other showing contributions also at frequencies that can be coupled in a complex way with the exciton (Group 2).

Statistical tests performed on the data obtained showed a lack of correlation between the magnitude of the fluctuations of the excitation energy and belonging to a particular Group. Furthermore, they suggested that such fluctuations are not specific to any of the proteins analysed, but rather a general value that could be obtained from any protein. Extending this preliminary observation to a wider range of substrates might have important consequences in ruling out hypotheses concerning the natural evolution of such complex proteic systems.

By examining the correlation between the excitation energy of the chromophore and the coulombic interaction energy with the residues, we showed that solvent molecules are predominantly responsible for the strongest correlations. Further investigation of the separation between the chromophore and surrounding molecules showed that the importance of a residue or solvent cannot be solely determined by proximity. A more detailed analysis of correlations and the MD trajectory revealed a medium-range interaction between chromophores and solvent molecules. Their relative arrangement and the distance suggests that the main interaction mechanism is dipole-dipole in nature. Solvent molecules strongly modulating the fluctuation of the excitation energy are kept in place by a strong interaction network (involving H-bonds), but are able to rotate, thus continuously modifying the extent of the interaction, resulting in a significant modulation. For WSCP proteins, we identified a steric interaction between the scaffold maintaining in place the relevant solvent molecules and the phytyl chains, which, as reported in the literature, play a fundamental role in photoprotection mechanisms.

Finally, an analysis of the composition of the environment around each chromophore ruled out obvious correlations between the magnitude of the fluctuations of the excitation energy and the number of residues of a certain type in the proximities of the chromophore. This observation is possible only upon analysis of a range of similar chromophores appearing in different proteins: by only looking at one protein (*e.g.* FMO), one can be lead to identify correlations between solvent proximity and magnitude of fluctuations. However, by considering more systems, we found similar fluctuations for similar chromophores embedded in a different environment, which make such correlation more modest.

Our analysis on three substrates seems to suggest that the nature of the protein environment is rather “unspecific” and independent of the biological function. Extending our findings to other substrates might help ruling out hypotheses on the target of natural evolution. Overall, it appears that the only obvious mean to decrease the exciton-environment coupling is the reduction of solvent accessible pockets in the vicinity of the chromophores. Other complementary studies[59, 60] have considered the role of the relative geometric arrangement of chromophores (at fixed chromophore-environment interaction) finding that in that case there is a degree of optimisation found in the naturally occurring light harvesting complexes. Moreover, if one considers that the exciton is more strongly coupled with motions of the chromophore itself[25] than protein/solvent motions, one may argue that the dynamics of the excited states cannot be affected too dramatically by the chemical detail of the surrounding protein environment. In summary, these results indicate that the most effective way to control the energy transfer process (either in artificial systems or by natural evolution) is by

controlling the chromophore-chromophore interaction[61] rather than the chromophore-environment interaction.

Acknowledgements

This work was supported by the ERC through Grant No. 615834. We are grateful to Dr. Ana Damjanović (Johns Hopkins, USA) for the CHARMM parameters of BCL chromophores.

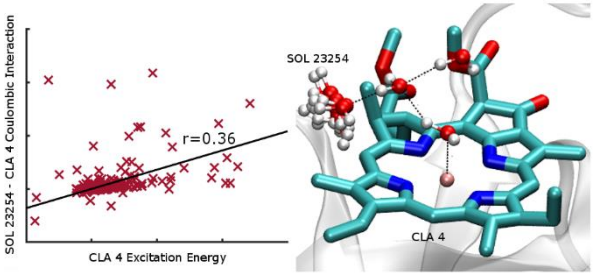
References

1. G. Renger, Royal Society of Chemistry, Cambridge, 2007, DOI: 10.1039/9781847558152-00001, pp. 5-35.
2. G. D. Scholes, G. R. Fleming, A. Olaya-Castro and R. van Grondelle, *Nat. Chem.*, 2011, **3**, 763-774.
3. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*, Blackwell Science Ltd, Oxford, UK, 2002.
4. G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mancal, Y.-C. Cheng, R. E. Blankenship and G. R. Fleming, *Nature*, 2007, **446**, 782-786.
5. C. Y. Wong, R. M. Alvey, D. B. Turner, K. E. Wilk, D. A. Bryant, P. M. G. Curmi, R. J. Silbey and G. D. Scholes, *Nat. Chem.*, 2012, **4**, 396–404.
6. A. Ishizaki, T. R. Calhoun, G. S. Schlau-Cohen and G. R. Fleming, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7319-7319.
7. E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, **463**.
8. G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship and G. S. Engel, *PNAS*, 2010, **107**, 12766-12770.
9. H. G. Duan, V. I. Prokhorenko, R. J. Cogdell, K. Ashraf, A. L. Stevens, M. Thorwart and R. J. D. Miller, *PNAS*, 2017, **114**, 8493-8498.
10. A. Ishizaki and G. R. Fleming, *Annu. Rev. Condens. Matter Phys.*, 2012, **3**, 333-361.
11. A. Chenu and G. D. Scholes, *Annu. Rev. Phys. Chem.*, 2015, **66**, 69-96.
12. S. Jang and Y.-C. Cheng, *Wiley Interdiscip Rev Comput Mol Sci*, 2013, **3**, 84-104.
13. S. Chandrasekaran, M. Aghtar, S. p. Valleau, A. Aspuru-Guzik and U. Kleinekathöfer, *J. Phys. Chem. B*, 2015, **119**, 9995-10004.
14. L. Viani, M. Corbella, C. Curutchet, E. J. O'Reilly, A. Olaya-Castro and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16302-16311.
15. M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2014, **5**, 3131-3137.
16. S. Hofinger and T. Simonson, *J. Comput. Chem.*, 2001, **22**, 290-305.
17. H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science*, 1991, **254**, 1598-1603.
18. M. K. Lee and D. F. Coker, *J. Phys. Chem. Lett.*, 2016, **7**, 3171-3178.
19. C. Olbrich, J. Strümpfer, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2011, **2**, 1771-1776.
20. T. Renger, A. Klinger, F. Steinecker, M. Schmidt Am Busch, J. Numata and F. Müh, *J. Phys. Chem. B*, 2012, **116**, 14565-14580.
21. A. Anda, L. De Vico and T. Hansen, *J. Phys. Chem. B*, 2017, **121**, 5499-5508.
22. F. Vaughan, N. Linden and F. R. Manby, *J. Chem. Phys.*, 2017, **146**, 124113.
23. P. Maly, O. J. Somsen, V. I. Novoderezhkin, T. Mancal and R. van Grondelle, *ChemPhysChem*, 2016, **17**, 1356-1368.
24. A. Chenu, N. Christensson, H. F. Kauffmann and T. Mancal, *Sci. rep.*, 2013, **3**, 2029.
25. D. Padula, M. H. Lee, K. Claridge and A. Troisi, *J. Phys. Chem. B*, 2017, **121**, 10026-10035.
26. C. W. Kim, J. W. Park and Y. M. Rhee, *J. Phys. Chem. Lett.*, 2015, **6**, 2875-2880.
27. C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294-343.

28. C. W. Kim, B. Choi and Y. M. Rhee, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3310-3319.
29. O. Andreussi, I. G. Prandi, M. Campetella, G. Prampolini and B. Mennucci, *J. Chem. Theory Comput.*, 2017, **13**, 4636-4648.
30. H. Do and A. Troisi, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25123-25132.
31. S. M. Blau, D. I. G. Bennett, C. Kreisbeck, G. D. Scholes and A. Aspuru-Guzik, *Proc Natl Acad Sci U S A*, 2018, **115**, E3342-E3350.
32. M. K. Lee, K. B. Bravaya and D. F. Coker, *J. Am. Chem. Soc.*, 2017, **139**, 7803-7814.
33. A. M. Rosnik and C. Curutchet, *J Chem Theory Comput*, 2015, **11**, 5826-5837.
34. D. Bednarczyk, O. Dym, V. Prabahar, Y. Peleg, D. H. Pike and D. Noy, *Angew. Chem. Int. Ed. Engl.*, 2016, **55**, 6901-6905.
35. V. Vaissier, P. Barnes, J. Kirkpatrick and J. Nelson, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4804-4814.
36. D. P. McMahon and A. Troisi, *The Journal of Physical Chemistry Letters*, 2010, **1**, 941-946.
37. K. A. Sharp, *Biophys. J.*, 1998, **73**, 1241-1250.
38. P. F. Barbara, T. J. Meyer and M. A. Ratner, *J. Phys. Chem.*, 1996, **100**, 13148-13168.
39. R. E. Fenna, B. W. Matthews, J. M. Olson and E. K. Shaw, *J. Mol. Biol.*, 1974, DOI: 10.1016/0022-2836(74)90581-6.
40. X-ray resolution 2.3 Å for FMO and 1.96 Å for WSCP.
41. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859-1865.
42. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
43. J. Huang and A. D. MacKerell, Jr., *J. Comput. Chem.*, 2013, **34**, 2135-2145.
44. A. Damjanović, I. Kosztin, U. Kleinekathöfer and K. Schulten, *Phys. Rev. E*, 2002, **65**, 031919.
45. F. Guerra, S. Adam and A. N. Bondar, *J. Mol. Graph. Model.*, 2015, **58**, 30-39.
46. B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116-122.
47. M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, 1998, **108**, 4439-4449.
48. C. K. Olbrich, Ulrich *J. Phys. Chem. B*, 2010, **114**, 12427-12437.
49. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372-1377.
50. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51-57.
51. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **120**, 215-241.
52. J. D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.
53. Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T.

Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184-215.
54. R. Zwanzig, *Acc. Chem. Res.*, 1990, **23**, 148-152.
55. W. H. Kruskal and W. A. Wallis, *J. Am. Stat. Assoc.*, 1952, **47**, 583-621.
56. B. H. Besler, K. M. J. Merz and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431-439.
57. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129-145.
58. A. Agostini, D. M. Palm, F. J. Schmitt, M. Albertini, M. D. Valentin, H. Paulsen and D. Carbonera, *Sci. rep.*, 2017, **7**, 7504.
59. G. C. Knee, P. Rowe, L. D. Smith, A. Troisi and A. Datta, *J. Phys. Chem. Lett.*, 2017, **8**, 2328-2333.
60. L. A. Baker and S. Habershon, *J. Chem. Phys.*, 2015, **143**, 105101.
61. R. P. Fornari, P. Rowe, D. Padula and A. Troisi, *J Chem Theory Comput*, 2017, **13**, 3754-3763.

TOC Graphics



Analysis of intermolecular motions of pigment-protein complexes shows no significant difference in influence of local environment despite different biological functions.