Engineering Excitonically-Coupled Dimers in an Artificial Protein for Light Harvesting via Computational Modelling

Mariano Curti,[1,*] Valentin Maffeis,[1,§] Luís Gustavo Teixeira Alves Duarte,[1] Saeed Shareef,[1,2] Luisa Xiomara Hallado,[1,2] Carles Curutchet,[3,4] and Elisabet Romero[1,*]

[1] Institute of Chemical Research of Catalonia (ICIQ), Barcelona Institute of Science and Technology (BIST), Avda. Països Catalans 16, 43007 Tarragona, Spain.

[2] Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, 43007 Tarragona, Spain

[3] Departament de Farmàcia i Tecnologia Farmacèutica, i Fisicoquímica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Av. Joan XXIII 27-31, 08028 Barcelona, Spain.

[4] Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona (UB), Martí i Franqués 1, 08028 Barcelona, Spain.

* Corresponding authors:
Mariano Curti
mcurti@iciq.es · Avda. Països Catalans 16, 43007 Tarragona, Spain · Tel: (+34) 977 920 20
Elisabet Romero
eromero@iciq.es · Avda. Països Catalans 16, 43007 Tarragona, Spain · Tel: (+34) 977 920 20

[§] Current address: Univ. Lyon, ENS Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie, UMR 5182, 46 Allée d'Italie, 69364 Lyon, France

Running title: Excitonically-Coupled Dimers in an Artificial Protein

Number of manuscript pages: 26

Number of figures: 6

Number of tables: 1

Abstract

In photosynthesis, pigment – protein complexes achieve outstanding photoinduced charge separation efficiencies through a set of strategies in which excited states delocalisation over multiple pigments (‘excitons’) and charge-transfer states play key roles. These concepts, and their implementation in bioinspired artificial systems, are attracting increasing attention due to the vast potential that could be tapped by realising efficient photochemical reactions. In particular, *de novo* designed proteins provide a diverse structural toolbox that can be used to manipulate the geometric and electronic properties of bound chromophore molecules. However, achieving excitonic and charge-transfer states requires closely spaced chromophores, a non-trivial aspect since a strong binding with the protein matrix needs to be maintained. Here, we show how a general-purpose artificial protein can be optimised *via* molecular dynamics simulations to improve its binding capacity of a chlorophyll derivative, achieving complexes in which chromophores form two closely spaced and strongly interacting dimers. Based on spectroscopy results and computational modelling, we demonstrate each dimer is excitonically coupled, and propose they display signatures of charge-transfer state mixing. This work could open new avenues for the rational design of chromophore – protein complexes with advanced functionalities.

Keywords

De novo protein design · Artificial photosynthesis · Molecular dynamics · Cofactor binding

Statement

By forming complexes with pigments, proteins provide the key machinery for photosynthesis. Bioinspired systems based on artificial proteins and chromophores are thus a very attractive platform to develop solar powered processes. We report here the (computational) optimization of an artificial protein, in order to increase the affinity and uptake of chromophoric cofactors from two to four molecules. This leads to a system with strongly interacting chromophores, a feature that underpins the efficiency of photosynthetic complexes.

Introduction

In photosynthesis, Light-Harvesting Complexes (LHC) are responsible for photon absorption and subsequent energy transfer to Reaction Centres (RC), where charge separation takes place to initiate a chain of reactions that ultimately store solar energy as chemical bonds.[1,2] As their name implies, LHCs are specialized in light absorption, densely packing a large number of pigments within a protein matrix that maximizes the absorption cross section while avoiding concentration quenching.[3] Excitation energy is then funnelled to RCs, where, using a set of strategies termed "design principles of charge separation",[4] it leads to charge-separated states with almost perfect efficiency (i.e., near unitary quantum yield). The cornerstone of this efficiency (and of the *design principles*) is a very complex energy landscape,[4] finely tuned by millions of years of evolution. Among other features, excitonic interactions and charge-transfer states play preponderant roles in these systems.[4–7]

Modification of natural systems to fulfil new specific needs (for instance, to drive a chemical reaction), is evidently burdensome, if not impossible. As an alternative, the *de novo* design of artificial cofactor–protein complexes has seen substantial progress in recent years, resulting in bioinspired systems able to act, for instance, as ion channels[8] or multi-electron oxidases/reductases.[9–11] This strategy generally follows a bottom-up approach, where the complexes are generated from scratch following simple design rules, yet drawing inspiration from natural systems.[12,13] Excitingly, the growing knowledge on protein design could provide a vast array of tools to adjust the energy landscapes as necessary – for instance, to implement the *design principles* and reach high charge separation efficiencies. However, this will only be possible if the protein binds cofactors with a high packing density, in order for them to display excitonic interactions and low-lying charge-transfer states.

Among different possibilities, a family of structures termed "maquette proteins" provides relatively simple scaffolds, based on the four-alpha-helix bundle motif, amenable to cofactor loading.[14–20] These complexes have already shown basic light absorption and energy transfer functionalities.[19,21,22] For instance, the group of Noy studied the binding of a bacteriochlorophyll derivative into an artificial maquette originally designed to bind heme, and observed the uptake of two to three chromophore molecules per protein.[23] Although this was enough to observe closely separated chromophore dimers with strong excitonic

interactions, the protein had a nominal capacity of four chromophores, given by four histidine residues used to ligate them. In another remarkable example of the potential of *de novo* protein design, the DeGrado group has recently demonstrated a maquette designed with sub-angstrom accuracy, capable of binding a diphenyl Mn-porphyrin with redox functionality.[24] On the other hand, the design was performed in an *ad hoc* fashion with specific software, and the cofactor:protein mass ratio was very low (3.7%) if compared with LHCs (LH2 from *Rs. Molischianum*: 24%, only considering bacteriochlorophylls).[25]

Here, we analyse and optimise a 132 amino acids-long maquette protein developed by the Dutton group, which has been previously employed to bind different cofactors.[19,21] However, even though four histidines (His) are available for ligation, this protein has been used to bind at most two cofactors, that remained too far apart to show excitonic coupling. We perform here spectroscopic analyses that show that indeed this protein can strongly bind up to two molecules of a chlorophyll *a* derivative. In order to engineer excitonically coupled chromophore dimers, we perform a computational study to find point mutations that increase the number of bound cofactors and their binding affinities. Next, we produce two of the candidate structures and experimentally demonstrate full cofactor binding. Finally, we employ quantum mechanics calculations at the TD-DFT level in combination with circular dichroism and fluorescence spectroscopy to get insights into the excitonic and electronic properties of the complexes and corroborate the computational models.

Results

Binding to original structures

We start the analysis by considering the BT6 protein,[19,26–28] which has 4 histidine residues (Figure 1a), and two mutants in which either two[21] (2H→2A variant) or all four[29] (4H→4A) His were replaced by alanine (Ala, see sequences in Table 1). With light-harvesting functionality in mind, we have chosen Zn 13^2-OH-methylpheophorbide *a* (Figure 1b), a chlorophyll *a* derivative, as the chromophoric cofactor.

Although crystallization efforts on BT6 have so far been unsuccessful, protein structure prediction software[30–32] yields a model (Figure 1a) consistent with insights from previous studies.[19] The protein sequence (Table 1), following *de novo* design principles, is built around heptad repeats of the form *HPPHHPP*, where *H* are hydrophobic and *P* polar amino acids.[20]

This simple pattern, corresponding to two turns of an alpha helix, leads to a structure where the *H* amino acids are buried in a hydrophobic core while *P* amino acids are solvent exposed, conferring it water solubility and great thermal stability.[19] BT6 thus has a single-chain four alpha helix bundle topology, connected by three glycine- and serine-rich loops. In order to ligate chromophores, a His residue is placed in each of the helices at buried positions, forming two closely spaced pairs at the top and the bottom of the structure (Figure 1a). In the 2H→2A variant, His residues at positions 7 and 112 were replaced by alanine, leaving two remaining His, one at each side of the protein structure (top and bottom).

To assess chromophore binding, we perform titration experiments, in which a concentrated chromophore solution is added stepwise to a protein solution at pH 9 (for the detailed procedure see SI Section 1.5). As monitored by absorption spectroscopy (Figure 1c), chromophore addition to a solution of the 4H→4A variant simply results in chromophore dissolution and subsequent aggregation,[33] because this protein has no His residues to ligate it. The aggregation process is evinced by the progressive formation of a red shoulder at ~683 nm on the first electronic transition, termed Q_y and initially centred at ~664 nm. The Soret region (300 to 450 nm) shows a complex pattern with several transitions. A decrease in the extinction coefficient of (at least) two overlapping bands centred at ~413 and 435 nm is clear as chromophore concentration increases and aggregation proceeds. Control experiments in the absence of protein show that aggregation is unaffected by the presence of the 4H→4A variant (Figure S6).

The 2H→2A variant, with two His residues, is expected to bind two chromophores. Titration of this protein shows a narrower and stronger Q_y transition centred at 669 nm for low [C]:[P] (chromophore to protein) ratios (Figure 1d). After [C]:[P] exceeds 2, this band red-shifts and broadens, approaching then the behaviour of 4H→4A. The lower-energy side of the Soret region (400 – 450 nm) shows a different behaviour as well when compared to 4H→4A. For [C]:[P] ≤ 2, the band at ~442 nm is more intense than that at 420 nm; as the ratio increases, both shift towards the blue, and the latter becomes more intense than the former. Together with the width of the Q_y band, this indicates chromophore ligation for a [C]:[P] ratio up to two, followed by the formation of free and aggregated chromophore for larger ratios (cf. experiments in the absence of protein, Figure S6).

Given that the absorption spectral changes upon chromophore binding or aggregation are rather subtle, we employed singular value decomposition (SVD) and chemical equilibrium fitting (based on least-squares optimization) to disentangle the concurrent processes in the titration experiments. Whereas this analysis can only identify three absorbing species in titrations of the 4H→4A variant (namely, free chromophore, aggregated chromophore, and protein), an additional species is detected for 2H→2A, corresponding to protein-bound chromophore. The fact that only one extra species is observed for 2H→2A indicates that the absorption spectra for the complexes with either one or two bound chromophores are indistinguishable. We can then only model the binding process as two sequential steps with identical equilibrium constants:

$$P + C \overset{K_1}{\rightleftharpoons} PC$$

$$PC + C \overset{K_1}{\rightleftharpoons} PC_2$$

In these equations, P, C, PC, PC_2 and K_1 represent the protein, the chromophore, the protein with one or two bound chromophores, and the equilibrium constant, respectively. From the absorption spectra and SVD analysis, we determine the binding affinity to be around 160 nM. The SVD and fitting results are summarized in Table 1 and Figure S7 for all systems under study.

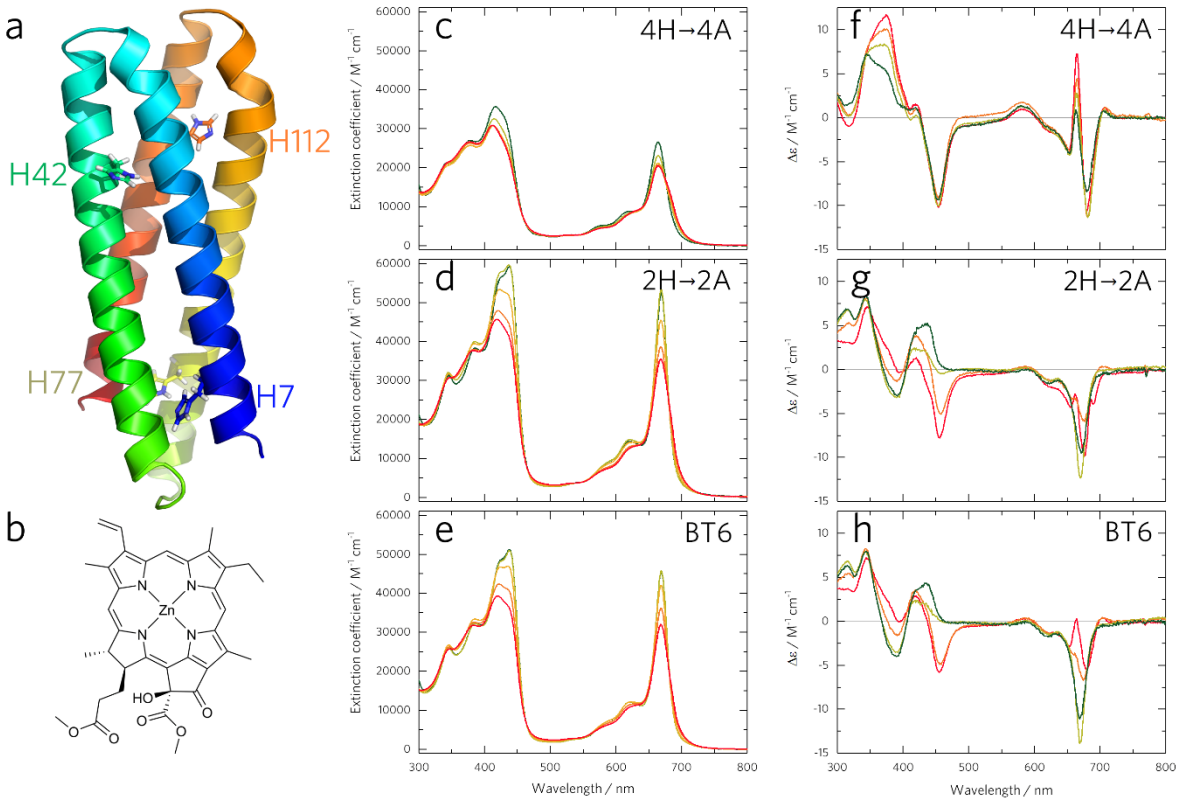

Figure 1: a) Protein structure for BT6 as predicted with trRosetta[32]. Histidine residues are shown in sticks representation. b) Chemical structure of Zn 13^2-OH-methylpheophorbide *a*. c-h) Chromophore to protein binding titration experiments monitored by absorption (centre) and circular dichroism (right) spectroscopy, for 4H→4A (c and f), 2H→2A (d and g) and BT6 (e and h). For absorption spectra, the chromophore:protein concentration ratios increase from 1 (green) to 5 (red), at a 5 µM protein concentration. For circular dichroism spectra, chromophore:protein ratios are shown from 1 (green) to 4 (red), at a 20 µM protein concentration. The proteins are in a buffer solution with 20 mM CHES at pH 9 and 150 mM KCl.

Moving on to the base BT6 design (with four His residues), the data is remarkably similar to that of the 2H→2A variant. Indeed, SVD indicates the same number of species, which hints at the difficulty of binding a third and fourth chromophore molecule (which should yield different spectra due to the expected close range interactions between the cofactor pairs at the top and bottom of the protein structure). The binding model used for the 2H→2A variant fits the titration dataset reasonably well, and produces the same binding constant (Table 1). However, a model where the protein binds three chromophores results in a comparable fit quality (with pK values of 7.2 for the first two chromophores and 6.2 for the third). Thus, although these fittings are not conclusive, it is likely that BT6 binds two chromophores relatively strongly, and then a third one more loosely (i.e., with a dissociation constant

comparable to the concentrations employed here, in the micromolar range), obscuring the observation of PC_3 species.

Analysis of circular dichroism (CD) spectra (Figure 1) leads to similar conclusions. For the 4H→4A variant, the formation of an excitonically-coupled species (i.e., chromophore aggregates) is clear at all [C]:[P] ratios. This is most evident in the Q_y region in the form of a conservative doublet with positive (negative) components at 663 nm (681 nm). The spectra of chromophore solutions in the absence of protein display a similar pattern (Figure S10). For the 2H→2A variant, spectra at [C]:[P] ≤ 2 show instead a single negative band at ~669 nm (cf. absorption spectra), indicative of the chiral nature of the chromophore and of the absence of excitonic interactions. Chromophore solutions in methanol, where it exists as monomers, display a similar CD spectrum (Figure S11).

Spectra for [C]:[P] ≥ 3 show complex signals arising from the simultaneous presence of proteins with two bound chromophores and chromophore aggregates. The base BT6 design shows a behaviour similar to the 2H→2A variant; it is clear that despite having four histidine residues, only two chromophores can bind to BT6 before aggregation occurs.

Table 1: Amino acid sequences of the artificial proteins employed in this study, along with the dissociation constants for cofactor binding. The two values, $K_{D,1}$ and $K_{D,2}$, correspond to the dissociation constants for the first and second pair of cofactors, respectively. The uncertainty in the pK values, obtained from duplicate experiments, is ~0.5.

Name	Sequence	$pK_{D,1}$ ($K_{D,1}$)	$pK_{D,2}$ ($K_{D,2}$)
BT6	G EIWKQHEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDALQKFEEALNQFEDLKQL	6.8 (158 nM)	--
BT6 - 2H→2A	G EIWKQAEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQL	6.8 (158 nM)	--
BT6 - 4H→4A	G EIWKQAEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQL	--	--

BT6 - 4E→4K	G EIWKQHKDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHKDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHKDALQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHKDALQKFEEALNQFEDLKQL	7.4 (40 nM)	6.6 (251 nM)
BT6 - 4L→4A	G EIWKQHEDAAQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDAAQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDAAQKFEEALNQFEDLKQL GGSGSGSGG EIWKQHEDAAQKFEEALNQFEDLKQL	8.2 (6 nM)	6.8 (158 nM)

Computational characterization

Computational simulations can shed light on the properties of these complexes, as it has been elegantly revealed by the pioneering and seminal work by Prof. Klaus Schulten[34–38] and co-workers for LH complexes[39–43] and even for more complex multi-component biological systems[44–47]. They offer, however, a particular challenge, since all attempts to crystallize BT6 have so far failed, and thus no structural model is available from experiments.[48] Nevertheless, the simplicity and strong foldability of maquette proteins render them good candidates for structure prediction software.[13] Indeed, all tested programs[32,49,50] predict four alpha helix bundles. These, however, only produce structures for apo-proteins; to generate holo forms, we developed a computational protocol (described in detail in the SI Section 2.3) in which we mimic the binding process with steered MD simulations, followed by relaxation to allow the complexes to find a stable conformation. Based on the BT6 sequence and on structural information from related systems (see Supplementary Note 1), we generated four complexes (Figure 2), labelled as Holo 1, Holo 2 *cis*, Holo 2 *trans*, and Holo 4. The numbers indicate how many chromophores are bound per protein, while *cis* and *trans* indicate whether the chromophore pair is bound to neighbouring or non-neighbouring His residues, respectively (see Figure 2). In Holo 2 *cis* chromophores are bound to H42 and H112, while Holo 2 *trans* has chromophores bound in positions 42 and 77, i.e., the same as for the 2H→2A variant. In turn, the Holo 4 structure is representative of fully-bound BT6. The fact that the four helices share the same sequence (Table 1) ensures that the chromophore environment is similar regardless of the His residue at which it is bound. This is corroborated by the fact that even though there are four possibilities for Holo 1 (with one His at each of the four helices), all the four different Holo 1 proteins display similar properties (data not shown).

From microsecond-long MD simulations, we first calculate chromophore binding free energies using the Quantum/Molecular Mechanics Generalized Born Surface Area (QM-MMGBSA) approach, at the PM6 theory level for the quantum part. Generally, cofactor binding energy calculations based on this type of methods offer reliable *relative* values, but inaccurate *absolute* values.[51] Nevertheless, some factors affecting the absolute energy values are worth discussing. The binding free energy ΔG_{bind} is calculated as:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand}$$

Ligand refers to the chromophore(s), receptor to the apo-protein, and complex to the holo-protein (i.e., including the chromophores). In terms of the energetic contributions, in the QM/MMGBSA method the binding free energy can be expressed as:

$$\Delta G_{bind} = \Delta E_{QM} + \Delta E_{MM} + \Delta G_{solv} - T\Delta S$$

Where ΔE_{QM} and ΔE_{MM} are the differences in the quantum mechanics and molecular mechanics energies (respectively), ΔG_{solv} is the solvation free energy, T temperature, and ΔS the entropy change.

In first place, as commonly done, we employ the *1-trajectory approach*, in which the holo-protein, apo-protein, and chromophore free energies are estimated from a single trajectory calculated for the complex. When considering binding of a cofactor to a receptor that has a well-defined pocket, the apo-structure can generally be well described by removing the cofactor from the holo-structure. Under these conditions, the 1-trajectory approach has good accuracy. In the present case, however, the apo-structure is predicted to be well-packed, without leaving any pocket to be filled by the chromophores, similar to a cryptic pocket.[52] Thus, removing the chromophores from the holo-structure leads to an empty pocket that does not exist in the apo-structure, which increases its surface area and leaves hydrophobic residues exposed to water. This results in an artificial destabilization of the receptor, and consequently on a binding energy that is too large.

When looking at the chromophore, the situation is similar: its structure is different in its free and bound states. Particularly, when bound to histidine, the Zn atom is displaced out of the porphyrin ring plane. Using one trajectory, then, leads to a binding energy that is too large as well. The magnitude of this error can be estimated more easily than for the receptor. To do

so, we re-parametrized the free chromophore using the same procedure but in absence of protein, and run a 1.0 μs production MD. Using this trajectory we calculated its average free energy using QM/MMGBSA. By comparison with the free energy calculated by isolating the chromophore from the protein, we estimate that G_{ligand} is 5.3 kcal mol^{-1} too low (and, correspondingly, ΔG_{bind} is too exergonic). A similar test performed over six snapshots of each trajectory (i.e., either the free chromophore, or structures taken from the complex trajectory) yielded an 11 kcal mol^{-1} free energy difference at the B3LYP-D3/def2-TZVP/PCM level.

From this discussion, a 3-trajectory approach would seemingly yield better results. However, in practice, the large fluctuations in the MM terms mask the free energy change from the binding process. It then leads to worse *absolute and relative* values of binding free energies. On the other hand, the 1-trajectory approach guarantees that the MM terms of complex, receptor, and ligand cancel out with each other. As already stated, it leads to wrong *absolute* values, but trustable *relative* values when evaluating a series of complexes.

Under these considerations, the binding energy for Holo 1 is -33.8 ± 1.0 kcal mol^{-1} at 298 K. Compared to the experimental value (-9.3 kcal mol^{-1} as calculated from the binding constant for BT6), it is clear that our calculations could not reproduce the *absolute* value. However, as explained above, it does not compromise the accuracy of *relative* values. The binding energy for Holo 2 *trans* is -33.9 ± 1.6 kcal mol^{-1} per chromophore, i.e., identical within error. On the other hand, Holo 2 *cis* yields a significantly lower binding energy per chromophore of -27.7 ± 1.0 kcal mol^{-1}. Although the decrease in binding energy between Holo 2 *trans* and Holo 2 *cis* is only around 20%, the fact that binding constants are exponentially related to binding free energies (through $K = e^{-\Delta G/RT}$) means that this difference has a large impact. Chromophore binding to neighbouring His thus proceeds with lower affinity, a fact that can be attributed to steric hindrance.

The binding energy for Holo 4, -27.8 ± 1.5 kcal mol^{-1}, is similar to that of Holo 2 *cis* (Figure 2), and explains the absence of full binding for BT6: while two chromophores can easily bind at non-neighbouring His residues, the binding of a third or fourth molecule would form closely packed dimers (i.e., as in Holo 2 *cis*) that are energetically unfavoured.

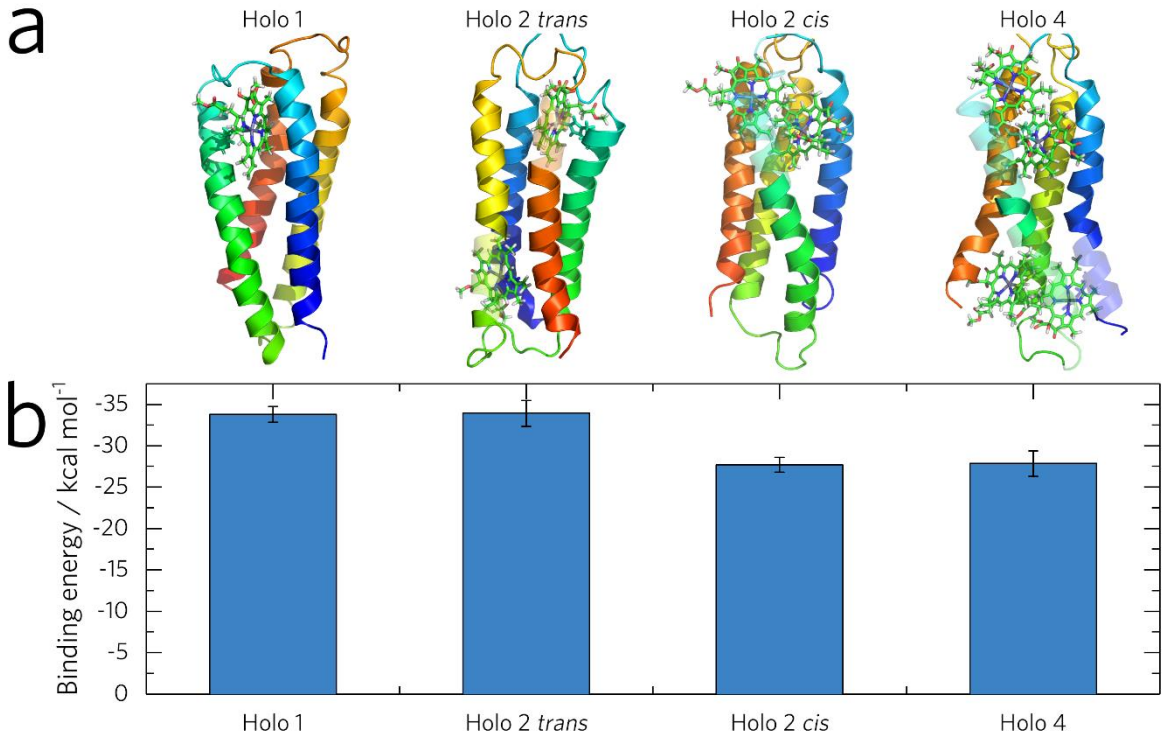

Figure 2: Representative structures (a) and binding free energies per chromophore molecule (b), as obtained from MD simulations for BT6. The structures differ on the number of bound chromophores, and in the case of Holo 2, whether the two chromophores are bound at neighbouring (*cis*) or non-neighbouring (*trans*) His residues. Binding free energies were obtained from 1.0 μs production runs, employing QM-MMGBSA.

We also performed excitonic QM calculations to model the CD spectra, which aid in the interpretation of the experimental data. In Holo 2 *trans*, the two chromophores are separated by ~2.0 nm (centre-to-centre), and thus excitonic interactions are negligible, as illustrated by the small electronic coupling between Q_y states (4 ± 3 cm^{-1}). As a result, its calculated CD spectrum (Figure S10) is given by the chirality of the chromophores. Indeed, the calculated CD spectral shape matches very well that for both BT6 and the 2H→2A variant at [C]:[P] ≤ 2 (Figure 1). On the other hand, the chromophores are closer together in Holo 2 *cis* (~1.2 nm), resulting in a larger electronic coupling (44 ± 11 cm^{-1}), and a conservative splitting in the Q_y region (Figure S10). The absence of this feature in the CD spectra of BT6 along the titration suggests that the two chromophores bind at non-neighbouring His residues, similarly to the 2H→2A variant. On the contrary, the Holo 4 model yields results similar to Holo 2 *cis*, and thus can be understood as two excitonically coupled dimers, with average couplings of 27 ± 12 cm^{-1} and 4 ± 2 cm^{-1} for neighbouring and non-neighbouring chromophore pairs.

Computational optimisation via Molecular Dynamics

With the aim of improving binding to BT6, we first decomposed the calculated binding energy into the contribution from each of the 14 residues closest (on average) to the chromophore pair in Holo 2 *cis* (Figure 3; see protein structure in Figure S5). All values are negative, indicating that each of these residues contributes favourably to binding. Largest contributions arise from hydrophobic residues in the protein interior, such as I38, F21, and W109, while residues in the loops (G28) or in the exterior of the structure (E113) have much smaller ones.

Next, we generated 40 structures by mutating these 14 residues to either Ala, Asp, or Lys (alanine, aspartate, or lysine), performed 0.7 μs long MD runs for each, and calculated the change in the binding free energy (Figure 3a). For most mutants, this change is small and within the uncertainty (< 1 kcal mol^{-1}). At the same time, a large fraction displays more negative binding energy values (i.e., a higher binding affinity) than Holo 2 *cis*, which has a simple explanation. Most of the residues under analysis are large and hydrophobic, and contribute to the tight packing of the protein core. Replacing them with the smaller Ala leaves more room for the chromophores to find an appropriate conformation, thus increasing the binding affinity. Conversely, their replacement with charged residues (Asp or Lys) induces structural distortion by which these residues, initially in the core, rotate to become solvent exposed (Figure S12). From the point of view of the chromophore pair, this has a similar effect of decongesting the binding site and decreasing steric hindrance.

To select the most promising candidates to improve chromophore binding affinity from these mutants, we defined a set of criteria. First, the improvement in affinity should be as large as possible. Second, the residue to mutate should not largely contribute to the binding energy in Holo 2 *cis*; otherwise, one risks disturbing favourable interactions. Third, there should exist a reasonable physical explanation behind the improvement. And finally, the geometrical relation between the mutated residue and the ligating His residues should be well defined. Regarding this point, under the assumption of a four alpha helix bundle, all residues within a given helix have a clear relation to the His in the same helix. However, the position relative to residues in other helices depends on the specific topology, which has in principle two possibilities for BT6 (see Section 2.1. *Apo-structures generation* in the SI). Two candidates for point mutation emerge from this description: L46A/L116A (symmetry related, Figure 3

(b)) and E113K (Figure 3 (c)). The remaining mutants can be discarded on the basis of the aforementioned criteria (as detailed in the Supplementary Note 2).

As the chromophores in Holo 2 *cis* are bound to H42 and H112, residues L46 and L116 are one helix turn below those His. Interestingly, all mutants replacing these Leu (i.e. by Ala/Asp/Lys) display a significantly better binding energy. This is rationalized from the above discussion on steric hindrance; the structure of Figure 3b illustrates how the smaller Ala induces a conformation change on the chromophore dimer. Given that these mutations are promising for further evaluation, and that charged residues may introduce unforeseeable distortions, we chose to replace these two Leu (and to keep symmetry, L11 and L81) by Ala, yielding the mutant BT6 – L11A L46A L81A L116A (4L→4A for short).

As a second candidate, we analysed the E113K mutant. In contrast with most other considered residues, the negative charge of E113 ensures this residue stays in the outer side of the structure. At the same time, the polar side of the chromophore, dominated by negative partial charge on oxygen atoms (Figure S3), occupies a similar position. Replacing E113 by Lys (K) then removes an unfavourable interaction with the chromophore, while providing a favourable electrostatic interaction with the positively charged Lys. These interactions are depicted in Figure 3 (c) (yellow dotted line). To maintain a protein structure in which all four alpha helices are equivalent, we thus selected the BT6 – E8K E43K E78K E113K mutant (4E→4K for short) for experimental corroboration and analysis, together with the above-described 4L→4A (their sequences are shown in Table 1).

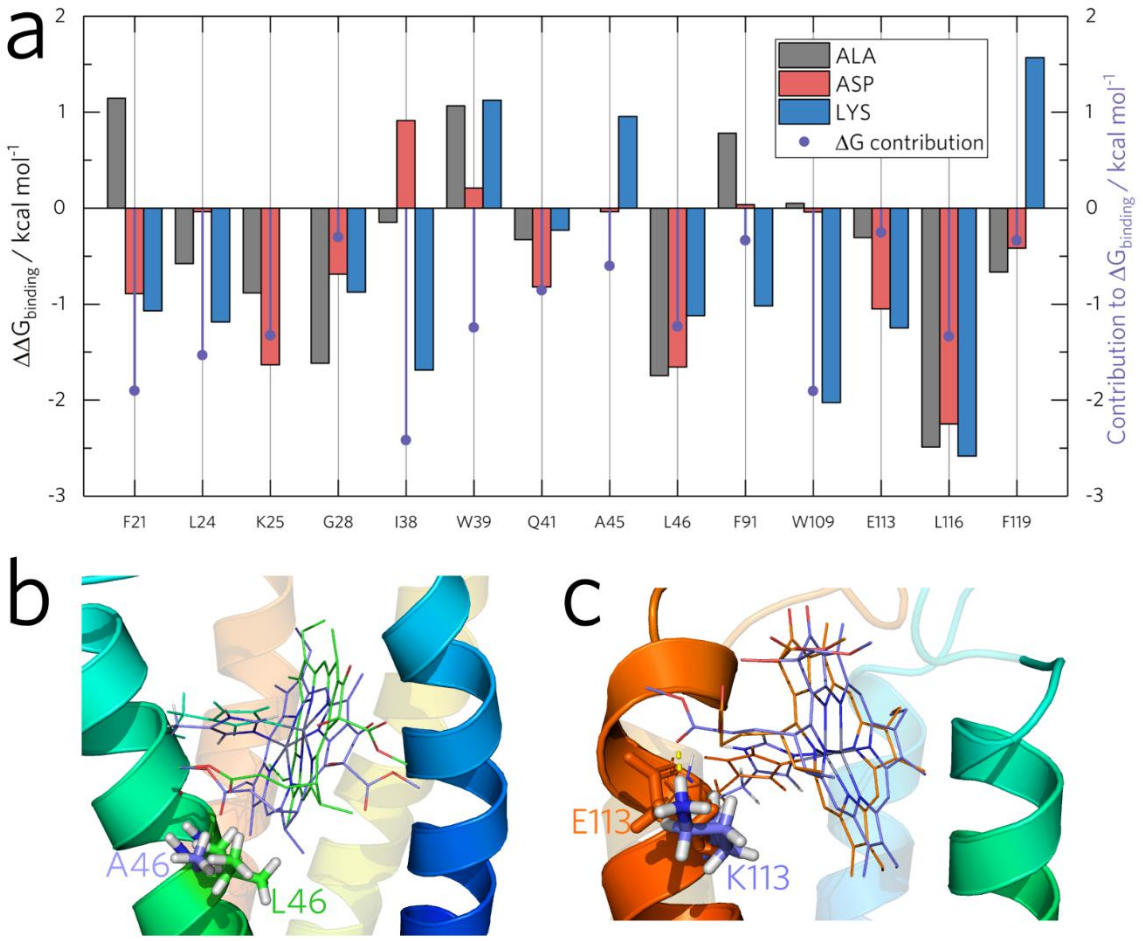

Figure 3: (a) Contribution to the binding energy of amino acids close to the chromophore pair in Holo 2 *cis*, together with the difference in binding energy after point mutation of each residue by Ala, Asp or Lys, as calculated from MD simulations. Binding energies were calculated with QM(PM6)-MMGBSA from a 0.5 µs trajectory for each mutant. (b) Representative structures for BT6 and the mutants L46A and E113K (c). Chromophore structures are shown in line representation (omitting hydrogens for clarity), and are coloured according to the parent structure: in b), green and blue correspond to a chromophore in BT6 and in the L46A mutant respectively, while in c) orange and blue represent a chromophore in BT6 and in E113K respectively.

Binding to optimised proteins and spectroscopic properties

To analyse chromophore binding, we performed titrations of the two newly proposed variants (4L→4A and 4E→4K) in identical conditions as for BT6 (Figure 4). An interesting observation is that the binding affinity for the first pair of chromophores is improved in these variants (Table 1) with respect to BT6 and 2H→2A: pK values increase from 6.8 for these two, up to 7.4 and 8.2 for 4E→4K and 4L→4A, respectively. Moreover, although the evolution of the spectral shapes is superficially similar, SVD analysis reveals an important

difference: in the titration of the new variants there is another species that appears at [C]:[P] ratios between 2 and 4, and thus can be confidently attributed to complexes containing 3 and 4 chromophores. Although the binding affinity for this pair is lower than for the first two chromophores, fitted dissociation constants are still sub-micromolar (Table 1), confirming a remarkably improved binding by the two new variants. The spectra extracted from SVD and assigned to PC_2 and PC_4 (Figure 4 and Figure S13) show a slight broadening and redshift of the latter, attributed to interactions now present in the closely packed chromophore dimers. Both variants (4L→4A and 4E→4K) show similar spectral features, indicating that the chromophore environments are comparable.

CD spectroscopy in the visible range indicates a similar behaviour than BT6 for [C]:[P] ≤ 2. However, while chromophore aggregates were immediately visible at higher ratios for BT6, 4L→4A and 4E→4K display gradual changes, with the formation of a positive feature at ~700 nm and a redshift and decrease of the negative band initially centred at 670 nm. A confounding factor in this series, however, is that under these experimental conditions (i.e., [P] = 20 μM) several species are present at any given [C] (see simulated concentration profiles in Figure S14). Since most of these contribute to CD spectra, we performed a SVD analysis similar to that for the absorption titrations, from which we obtained the specific contributions of PC_2 and PC_4 (Figure 4 and Figure S13). While the former has all the features derived from the chirality of the chromophores, the latter shows an intriguing pattern in the Q_y region, with an asymmetric doublet that is far from conservative.

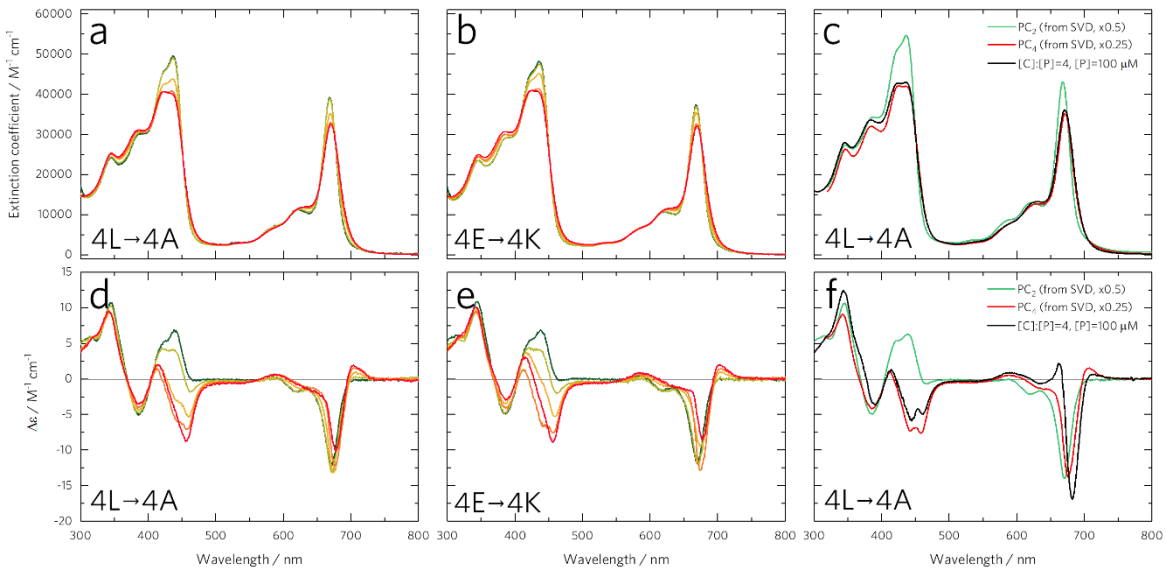

Figure 4: Chromophore to protein binding titration experiments monitored by absorption (a and b) and circular dichroism (d and e) spectroscopy, for the 4L→4A (a and d) and 4E→4K (b and e) variants. For absorption spectra the protein concentration is 5 µM, whereas for circular dichroism experiments it is 20 µM. In both cases chromophore:protein ratios increase from 1 (green) to 5 (red). The proteins are in a buffer solution with 20 mM CHES at pH 9 and 150 mM KCl. Absorption and circular dichroism spectra obtained from singular value decomposition of the 4L→4A titrations (c and f respectively), together with the spectra of a 100 µM solution of this variant at a chromophore:protein ratio of 4. In c and f the spectra are normalized to the number of chromophores to facilitate comparison.

To corroborate the spectra extracted from SVD analysis, we prepared the full complexes employing high protein concentrations ([P] = 100 µM) and stoichiometric amounts of chromophore (i.e., [C]:[P] = 4). Moreover, these samples were prepared from diluted solutions ([P] = 5 µM) in centrifugal concentrators. These conditions hinder chromophore aggregation, while shifting the equilibrium towards the formation of PC_4 (see speciation profiles in Figure S14). Both the absorption and CD spectra can be readily compared to those for PC_4 extracted from SVD, and indeed the main features coincide (Figure 4 c and f). An important difference is the appearance of a positive feature at ~662 nm in the CD spectrum at 100 µM protein concentration, that, by comparison with the 4H→4A variant (Figure 1), can be attributed to minor aggregate formation. Regardless of that, the experimental data taken at 100 µM protein concentration corroborates the accuracy of the spectra extracted from the titrations.

To analyse the properties of the excitonic excited states we performed steady-state fluorescence. From the SVD analyses, we chose two different excitation wavelengths: while 437 nm preferentially excites bound chromophores, 405 nm is more suitable to excite both free and aggregated chromophore molecules (Figure 5a). The obtained fluorescence spectra are displayed in Figure S15; most importantly, bound chromophore emission shows a maximum at 673 nm for all proteins, while a shoulder at ~640 nm rises as the concentrations of free and aggregated chromophore increase. We have thus chosen two different conditions to monitor fluorescence evolution as a function of the [C]:[P] ratio: excitation at 437 nm and emission at 673 nm to analyse bound chromophores (Figure 5b), and excitation at 405 nm with emission at 630 nm to assess unbound species (Figure 5c).

When looking at the fluorescence of bound chromophores (Figure 5b), the 2H→2A variant offers the simplest pattern: its intensity increases up to a [C]:[P] ratio of 2, reaching a plateau afterwards. The asymptotic value is approximately double than that at [C]:[P] = 1, indicating that both bound chromophore molecules emit with a similar quantum yield, and do not interact.

The rest of the chromophore binding proteins present a more complex behaviour: the emission intensity increases for [C]:[P] from 0 to 2, and then it promptly decreases for $2 \leq$ [C]:[P] ≤ 4. For the 4E→4K and 4L→4A variants, this behaviour can be readily ascribed to the formation of chromophore dimers inside the proteins. In fact, the fluorescence intensity profiles are strongly correlated to the calculated concentration profiles (see Figure S16): as the fraction of bound chromophore dimers increase, the fluorescence intensity concomitantly decreases (Figure 5b). We formulate two hypotheses: the excitonic interaction between chromophores could lead to H-type aggregates (with reduced emission[53]), and/or the first excitonic state could be admixed with a charge-transfer state. Both possibilities will be discussed in the following section.

A surprising result here is that chromophores bound to the BT6 protein show a comparable decrease in fluorescence intensity for $2 \leq$ [C]:[P] ≤ 4. Although the rest of the spectroscopic evidence points toward this protein binding two chromophore molecules, the fluorescence data clearly shows some intensity decrease, albeit not as strong as that in the case of the 4E→4K and 4L→4A variants. Assessing all available data, we conclude that BT6 strongly

binds two chromophores, and can incorporate a third one with lower affinity (i.e., with a dissociation constant above the micromolar range), giving rise to chromophore dimers in a fraction of the complexes, and an associated decrease in fluorescence intensity.

The fluorescence signal from free and aggregated chromophore shows, expectedly, a simpler pattern (Figure 5 c). Since all proteins can bind (at least) two chromophores, these signals are minimal for $0 \leq [C]:[P] \leq 2$. For larger ratios, the signal grows more strongly for BT6 and the 2H→2A variant, since they can tightly bind only two chromophores. On the other hand, the variants capable of binding four chromophores (4E→4K and 4L→4A) show a milder increase, that becomes stronger for $[C]:[P] \geq 2$ due to the lower binding affinity of the second chromophore pair. It is noteworthy that the observed intensities are correlated with the binding constants (Table 1), e.g., the 4L→4A variant shows both the lowest signal ascribed to free/aggregated chromophore and the strongest binding constants.

Quantum yield determinations on these systems are complicated by the fact that low complex concentrations lead to significant dissociation fractions (see Figure 5c and Figure S14). While employing larger concentration could minimized this issue, it has the undesired effect of increasing the optical densities of the solutions, thus exacerbating inner filter effects. Therefore, although precise measurements are not possible in these conditions, we estimate the fluorescence quantum yields to be around 20% for all proteins at low [C]:[P] ratios, which mainly corresponds to protein-bound, non-interacting chromophores. The values decrease to ~5% at [C]:[P] = 4 for BT6 and the 4L→4A and 4E→4K variants, situation in which most chromophore molecules are part of protein-bound dimers. These values are comparable to those observed in photosynthetic antennae. For instance, the LH-1 and LH-2 complexes of purple bacteria display fluorescence quantum yields of 8 – 10 %[54] (for monomeric bacteriochlorophyll in polar solvents, the value ranges from 11 to 20 %[55]).

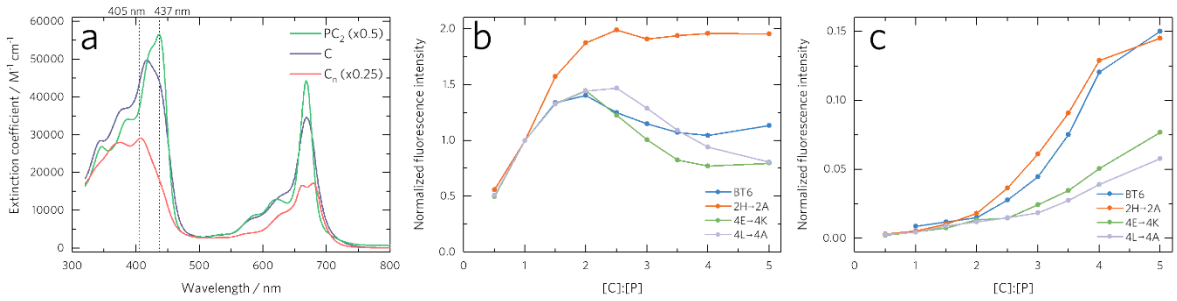

Figure 5: a) Absorption spectra obtained from singular value decomposition of the titrations (4L→4A variant). The vertical lines highlight the two excitation wavelengths, 405 and 437 nm. b) Fluorescence emission intensity at 673 nm, with an excitation wavelength of 437 nm. c) Fluorescence emission intensity at 630 nm, with an excitation wavelength of 405 nm. In b and c, profiles correspond to BT6 (blue), 2H→2A (yellow), 4E→4K (green) and 4L→4A (purple). Protein concentration was 1.25 μM and the buffer solution contained 150 mM KCl and 20 mM CHES at pH 9.

Modelling of the electronic properties

We resort again to computational modelling, to explain first the CD spectra of PC_4 in the 4L→4A and 4E→4K variants. Figure 6a shows the calculated spectra for Holo 2 *cis*, using either as-calculated or scaled electronic couplings. If the excitonic coupling between the chromophores is set to zero, we obtain signals exclusively related to their chirality, as explained before (indeed, the shape matches that of PC_2). If we include the full calculated coupling (i.e., 44 cm^{-1}), the CD signal in the Q_y range displays a perfectly conservative shape, as expected for strongly excitonically-coupled excited states. However, by scaling the coupling to a lower value (i.e., 40%), we are able to reproduce the pattern in the Q_y range, with a small positive feature for the low energy exciton component and a stronger negative signal for the high energy counterpart. The spectral signals in the Q_x and Soret regions qualitatively reproduce those observed in the spectra obtained by SVD. It can be concluded, then, that the excitonic coupling is smaller than 44 cm^{-1}, leading to a non-conservative signal that still is partly excitonic in origin.

This relatively small coupling is also evinced by the minor shift of the first absorption band displayed by complexes with four bound chromophores. Moreover, the experimental extinction coefficients for the 0–0 and 0–1 vibronic transitions provide information on excitonic interactions: a redistribution in oscillator strength from the 0–0 to the 0–1 transition is a characteristic signature of H-type aggregates.[56,57] In 4E→4K, for instance, the 0–0 / 0–1

intensity ratio (at ~669 nm and ~625 nm, respectively) decreases from 3.8 to 3.0 when comparing PC_2 to PC_4 (Figure S13), indicating relatively weak coupling.[57]

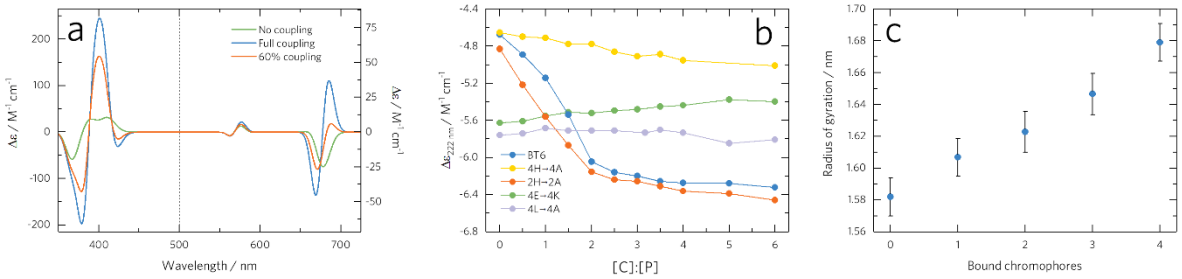

Figure 6: a) Simulated circular dichroism spectra of Holo 2 *cis*, using no (green), scaled (orange), or full (blue) coupling between cofactors. Spectra are the average of 20 calculations performed along 1.0 µs trajectories at the CAM-B3LYP/6-31G*/PCM level, and were shifted -0.25 eV to reproduce the experimental position of the Q_y band. To facilitate visualization, different *y*-scales are used for the ranges 350 – 500 nm and 500 – 750 nm. b) Circular dichroism signal at 222 nm, expressed as molar circular dichroism per residue, at different chromophore:protein ratios for BT6 (blue), 4H→4A (yellow), 2H→2A (orange), 4E→4K (green) and 4L→4A (violet). c) Calculated radii of gyration from MD simulations of chromophore – protein complexes including up to four cofactors.

We note that our initial coupling predictions could likely be overestimated due to the assumption of a continuum medium. Indeed, previous reports based on atomistic simulations pointed to strong screening effects in the strongly-coupled dimer of the PE545 photosynthetic complex of cryptophytes.[58] The heterogeneous polarizable environment of the protein leads to a ~31% attenuation of the coupling, compared to the continuum assumption.[58] Similar strong screening effects have also been observed for the central dimers of the PC630 and PC645 complexes.[59] Another explanation for the error incurred in the coupling calculations could be that transition dipole strengths might be overestimated, translating then into overestimated couplings. Finally, inaccuracies in the precise mutual arrangements of the chromophores could also impact this estimate.

The excitonic couplings are also relevant to explain the fluorescence experiments. From our computational modelling, we do observe H-type aggregates for the Holo 2 *cis* and Holo 4 structures. However, due to the relatively small value of the coupling, the oscillator strength is distributed in a 40% – 60% split between the low- and high-energy excitonic components.

It thus seems unlikely that the observed decrease in fluorescence intensity originates from excitonic interactions. Comparison with related systems, however, offers a possible explanation for the fluorescence intensity decrease: by employing Stark absorption and fluorescence spectroscopies, Noy *et al.* demonstrated excitonic – charge-transfer state mixing in both a chromophore – maquette protein complex and in the B820 subunit of the purple bacterial LH1 complex.[60] In addition, the maquette complex showed a decrease in fluorescence intensity similar to the present case,[23] while the B820 subunit displays a fluorescence quantum yield of 3%,[54] comparable to ours. By analogy, we thus propose that the decrease in fluorescence intensity observed in Figure 5b is related to a similar mixing of the low-energy excitonic state with a charge transfer state.

Structural characterization

Given that the chromophore:protein mass ratio in these complexes is remarkably high (around 18% for 4 cofactors per protein, comparable for instance to the 24% of LH2 in *Rs. Molischianum*),[25] it is informative to analyse structural changes upon binding. Several insights are available from circular dichroism spectroscopy in the UV range, where signals are related to both secondary and tertiary structures. All proteins under study display the characteristic pattern for alpha-helical secondary structures (Figure S17), with minima at *ca.* 208 nm and 222 nm. The evolution upon chromophore addition of both the signal at 222 nm and the ratio between 222 nm and 208 nm give related (but complementary) information.

From the intensity of the signal at 222 nm we can infer that, in the apo-state (Figure 6b), both new variants (4L→4A and 4E→4K) display a larger fraction of folded states than BT6 and the 4H→4A and 2H→2A variants. It has been determined that maquette proteins in general, and BT6 in particular, behave as molten globules in the apo-state.[15,17,19,61]At the same time, cofactor binding induces a strong structuring.[17,19] Indeed, both BT6 and the 2H→2A variant display a marked increase in the signal at 222 nm for [C]:[P] ≤ 2, remaining relatively constant for larger ratios. The 4H→4A variant, unable to bind cofactors, only shows a small increasing slope that we attribute to chromophore contributions to the CD signal. The 4L→4A and 4E→4K variants show a very different behaviour: as mentioned, their signals are stronger in absence of chromophore (due to a better structuring in the apo-states), and they slightly decrease in intensity with increasing [C]:[P] ratios (considering 4H→4A as the

baseline). This suggests that the apo-structures are fully folded, and thus chromophore binding slightly decreases the fraction of alpha-helical secondary structures.

The signal ratio between 222 nm and 208 nm has been empirically determined to be related to the coiled-coil character of helical bundles, with values larger than 1 indicating the presence of this feature.[62–64] All four-His containing proteins display an increase in this ratio for $0 \leq [C]:[P] \leq 2$ (Figure S18). The increase is much steeper for BT6 and the 2H→2A variant, attributed to a less structured apo-state, and a slightly more structured holo-state when compared with 4L→4A and 4E→4K. Interestingly, while the ratio at $[C]:[P] \geq 2$ remains constant for BT6 and 2H→2A, the 4L→4A and 4E→4K variants show a decrease in this range. Since the former strongly bind only two chromophore molecules, further additions do not modify the structures. On the contrary, 4L→4A and 4E→4K can each bind four chromophore molecules; we hypothesize that the uptake of the third and fourth molecules induces a distortion on the structure that reduces its coiled-coil character.

Structural information obtained from the simulations support these interpretations. The average fraction of residues adopting an alpha-helix secondary structure is 69 ± 3% for the apo structure, i.e. smaller than the canonical value of 80% derived from the sequence (Table 1), but expected from some fraying at the helices caps. This value is consistent with the experimental circular dichroism spectra in the UV range (Figure S17): fitting with the semi-empirical method SESCA[65] yields alpha helix fractions close to 70% for both new variants.

Chromophore uptake decreases the alpha-helix fraction: in the calculated Holo 4 structure, 61 ± 3% of the residues adopt a helical secondary structure (*versus* 69% in the apo-state). This explains why both 4L→4A and 4E→4K display a decrease in the 222 nm signal intensity concomitant with chromophore addition. On the contrary, for BT6 and the 2H→2A variant it can be concluded that the apo-structures are indeed molten globules, and their signal increase is due to an equilibrium shift towards the native-like state driven by chromophore binding.

The increase in the 222 nm to 208 nm ratio for $0 \leq [C]:[P] \leq 2$ (Figure S18) could be interpreted as the structures getting more compact, and thus acquiring a stronger coiled-coil character. However, data from calculations seem to disprove this notion: the size of the chromophores actually induces a separation of the helices, and the radius of gyration

undergoes a 6% increase on chromophore uptake (Figure 6c). Inspection of the protein structure indeed shows significantly larger deformations in the Holo 4 structure when compared to Holo 2 *trans* (Figure 2), explaining the decrease in the 222 nm to 208 nm ratio at [C]:[P] ≥ 2 for 4L→4A and 4E→4K.

Discussion

Decades of research on natural photosynthesis have led to a thorough understanding of its underlying mechanisms and of the strategies organisms employ to convert solar light into chemical energy.[4] Implementation of these concepts in artificial systems is, however, much harder, and the first fruitful attempts have been attained only recently.[19,22] Our work illustrates one possible pathway towards that goal, employing MD simulations to effectively optimise an artificial protein, and using them at the same time to obtain a wealth of information to understand its properties.

Given that we target light-harvesting applications, we use a zinc-based chlorin chromophore (Figure 1b) as cofactor. As a chlorophyll *a* derivative, it inherits several desired properties, such as a large absorption cross-section in a wide spectral range and tunable electronic properties.[66] Instead of designing a new protein in an *ad hoc* fashion, we instead employ BT6, a versatile four-alpha-helix structure that has already shown strong cofactor binding (up to two per protein) and photoinduced energy transfer.[19]

Our spectroscopic studies clearly show that, despite possessing four (chromophore-ligating) His residues, BT6 strongly binds only two chromophores; further additions lead to chromophore aggregates formation. Comparisons with a two-His variant (2H→2A) and computational modelling indicate that BT6 can easily accommodate two chromophores at non-neighbouring His residues (i.e., in opposites halves of the structure), but the incorporation of a third or fourth cofactor is sterically hindered. These results are comparable to those of Noy *et al.*, who employing a related maquette protein observed binding of two to three chromophores (depending on the protein concentration).[23]

Binding energy calculations on 40 hypothetical single-point mutants resulted in several candidates with improved cofactor affinity. We selected two of these candidates for experimental corroboration, and indeed both of them display a markedly improved affinity.

The dissociation constants are in the sub-micromolar range for the binding of four chromophore molecules per protein. In addition, spectroscopic studies in combination with data from simulations yield important structural information. While BT6 exists in the apo-state as a molten globule and chromophore binding induces its structuring, the two new variants are significantly better structured in the apo-state, and are largely unaffected by the binding process. However, the binding of a third and fourth chromophore molecules does lead to a reduction of the coiled-coil character in the new variants, as a consequence of the relatively large space occupied by chromophores with respect to the host structure.

Using the structural information generated from MD calculations, we performed quantum mechanical simulations at the TD-DFT level to calculate the CD spectra of the new variants, in order to assess the excitonic couplings. By comparing with experimental spectra, we find that the calculated excitonic coupling between the Q_y states of neighbouring chromophores (44 cm^{-1}) is likely overestimated. A 0.4 correction factor, on the other hand, leads to a remarkable agreement with the experimental spectra, confirming both the structural model and the computational scheme. Calculations nicely reproduce as well the spectra at low [C]:[P] ratios, confirming the notion that chromophore incorporation occurs first at non-neighbouring His residues, and later at neighbouring sites.

The effects of closely packing chromophore dimers inside the proteins are more evident on the emission properties of the complexes: both new variants display a marked fluorescence quenching for $2 \leq$ [C]:[P] ≤ 4. The relatively small excitonic coupling, together with prior Stark spectroscopy results on related systems[60] suggest that the origin of this decrease in fluorescence intensity is the admixture of the first excitonic state with a charge-transfer state. From this point of view, the complexes prepared here not only display a light-harvesting role, but could also conceivably integrate a charge separation function, which would represent a further step towards an artificial photosynthetic system.[4] Moreover, partial mixing with charge transfer states can increase static disorder, while avoiding quenching, an effect that has been shown to contribute to efficient light harvesting in the LH2 complex of purple bacteria.[67]

Software such as Rosetta are extremely powerful for their specific functions of protein design and structural prediction, but are significantly more limited in predicting chromophore-

protein complexes properties. Moreover, the lack of dynamical information in such methods can lead to a significant failure probability when dealing with protein design for cofactor binding.[68] In contrast, with an approach based on MD simulations, the same set of calculations (either by itself or in combination with other methods) can be used to gain useful information on different aspects,[69] illustrated here by binding affinity, excitonic couplings and spectroscopic properties determination, but not necessarily limited to those.

A limitation of our approach is the need for accurate structural models, usually provided by x-ray diffraction or nuclear magnetic resonance techniques. Since the proteins under study do not crystallize well, and their structures are relatively large and symmetric, these techniques are inadequate. On the other hand, protein structure prediction software, spearheaded by the AlphaFold[49] and RoseTTAFold[50] methods, have recently achieved remarkable accuracy, thus offering an attractive alternative to experimental determination.[70] Although these are still limited to apo-structures, we complemented them with an MD-based protocol that mimics the natural binding process. From this discussion, we conclude that our MD-based approach is not at odds with protein design and prediction software; instead, they form a synergistic combination. Finally, it is noteworthy that the excitonic signals observed in CD spectroscopy strongly depend on the relative orientation of the involved chromophores.[71] Therefore, the excellent match we obtained between calculated and experimental spectra can be taken as a confirmation that the structural model we present is accurate.

As mentioned above, we have observed that the two new variants are better structured than BT6 in the apo-state. We cannot thus rule out that this fact contributes to their improved binding affinity. It has been suggested before that structural stability could play an important role in cofactor binding: while a completely unstructured protein could hardly host cofactors, a protein that is too strongly packed will not be amenable for cofactor insertion.[48] An ideal protein then would probably display large flexibility, yet exhibit significant structuring in the apo-state. To unravel this issue, further studies could employ the new variants proposed here, and assess binding affinity after destabilizing the apo-state (for instance, by disrupting the hydrophobic core packing).

Besides protein-based scaffolds, controlled chromophore binding has recently been achieved *via* DNA origami.[57,72] However, exact control on the aggregates geometries proved to be difficult, as exemplified by inter-chromophore electrostatic interactions disrupting the DNA duplexes,[57] or by excitonic interactions that spanned only 2.5 chromophores despite the scaffolds providing a much larger number of binding sites.[72] Artificial proteins, on the other hand, can provide a more robust and tailorable scaffold for controlled chromophore binding.

All in all, through the computational optimisation of a maquette protein, we have achieved artificial complexes with record 18% chromophore-to-protein mass ratios. More importantly, this high packing density ensures the formation of strongly interacting chromophore dimer units. In the search for artificial photosynthetic systems, next steps will focus on the application of a similar computational approach to tune the energy landscape of these complexes, with particular attention on energy levels, excitonic couplings, and charge-transfer states.

Conclusions

In this work, we have tackled the realization of artificial chromophore – protein complexes, in which chromophores form strongly interacting dimeric units. Starting with an already designed maquette protein, we developed a computational protocol grounded on molecular dynamics simulations to both describe the properties of the base design, and to predict those of 40 mutants, with emphasis on the binding affinities towards a chlorophyll *a* derivative. While the initial design was able to strongly bind two chromophores at distant positions, two candidates from the computational protocol demonstrated binding of four chromophore molecules with a sub-micromolar affinity, as confirmed by experimental spectroscopic studies. The complexes reached a cofactor:protein mass ratio of 18%, approaching those of natural light-harvesting systems. Moreover, they contained two excitonically-coupled dimers, with charge-transfer state mixing signatures – essential features for efficient light conversion. Therefore, the complexes and protocols demonstrated here offer new ways for the realisation of artificial protein-based photosynthetic systems.

Data Availability

The data that support the findings of this study are available from the corresponding authors upon request. Moreover, all computational data has been deposited in the ioChem-BD database,[73] and is accessible through the DOI 10.19061/iochem-bd-1-240.

Supplementary Material Description

Detailed experimental and computational methods; Note on chromophore ligation; Note on protein mutants selection; Absorption and circular dichroism titrations in the absence of protein; Absorption spectra obtained from singular value decomposition of the titration experiments; Simulated circular dichroism spectra of Holo 2 *cis* and Holo 2 *trans*; Experimental and calculated circular dichroism spectra for Zn 13^2-OH-methylpheophorbide *a* in methanol; Absorption and circular dichroism spectra obtained from singular value decomposition of the 4E→4K titrations, and spectra of a 100 μM solution at [C]:[P] = 4; Speciation profiles for 4L→4A at three different concentrations; Normalized fluorescence spectra; Fitted fluorescence intensity profile for 4L→4A; circular dichroism spectra in the UV range for all studied proteins at different [C]:[P] ratios; Ratio in the circular dichroism signal at 222 nm over that at 208 nm for each protein at different [C]:[P] ratios.

Acknowledgements

M.C. acknowledges the funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801474. M.C. and E.R. acknowledge funding from the State Research Agency/Spanish Ministry of Science and Innovation (AEI/MICINN) through the Severo Ochoa Excellence Accreditation CEX2019-000925-S. L.G.T.A.D., S.S., L.X.H, and E.R. thank the European Research Council under the ERC starting grant agreement No. 805524 (BioInspired_SolarH2). V.M. thanks the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 754510. C.C. thanks the State Research Agency (AEI/10.13039/501100011033) through grants MDM-2017-0767 and PID2020-115812GB-I00. The authors thankfully acknowledge the computer resources at Pirineus and CTE-Power9, and the technical support provided by the Consorci de Serveis Universitaris de Catalunya and the Barcelona Supercomputing Center (QSB-2020-3-0002 and QSB-2021-1-0019). The authors thank Dominik Bäuerle for the initial fitting of some titration experiments.

Author contributions (CRediT statement)

Conceptualization and funding acquisition: M.C. and E.R.; Experimental investigation: V.M., L.G.T.A.D, S.S., and L.X.H; Computational investigation: M.C.; Formal analysis: M.C., V.M., and L.G.T.A.D; Visualization: M.C.; Writing – original draft: M.C. and V.M.; Writing – Review and Editing: M.C., V.M., L.G.T.A.D., C.C., and E.R.; Supervision: C.C. and E.R. All authors agreed to the final version of the manuscript.

Conflicts of interest

The authors declare no competing financial interest.

References

1. Blankenship RE Molecular Mechanisms of Photosynthesis. Wiley; 2002. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470758472
2. Scholes GD, Fleming GR, Olaya-Castro A, van Grondelle R (2011) Lessons from nature about solar light harvesting. Nat. Chem. [Internet] 3:763–774. Available from: http://www.nature.com/articles/nchem.1145
3. Beddard GS, Porter G (1976) Concentration quenching in chlorophyll. Nature 260:366–367.
4. Romero E, Novoderezhkin VI, Van Grondelle R (2017) Quantum design of photosynthesis for bio-inspired solar-energy conversion. Nature 543:355–365.
5. Lockhart DJ, Boxer SG (1988) Stark effect spectroscopy of Rhodobacter sphaeroides and Rhodopseudomonas viridis reaction centers. Proc. Natl. Acad. Sci. [Internet] 85:107–111. Available from: https://pnas.org/doi/full/10.1073/pnas.85.1.107
6. Frese RN, Germano M, de Weerd FL, van Stokkum IHM, Shkuropatov AY, Shuvalov VA, van Gorkom HJ, van Grondelle R, Dekker JP (2003) Electric Field Effects on the Chlorophylls, Pheophytins, and β-Carotenes in the Reaction Center of Photosystem II. Biochemistry [Internet] 42:9205–9213. Available from: https://pubs.acs.org/doi/10.1021/bi0273516
7. Romero E, Diner BA, Nixon PJ, Coleman WJ, Dekker JP, Van Grondelle R (2012) Mixed exciton-charge-transfer states in photosystem II: Stark spectroscopy on site-directed mutants. Biophys. J. 103:185–194.
8. Scott AJ, Niitsu A, Kratochvil HT, Lang EJM, Sengel JT, Dawson WM, Mahendran KR, Mravic M, Thomson AR, Brady RL, et al. (2021) Constructing ion channels from water-soluble α-helical barrels. Nat. Chem. 13:643–650.
9. Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. Proc. Natl. Acad. Sci. [Internet] 101:11566–11570. Available from: https://pnas.org/doi/full/10.1073/pnas.0404387101
10. Koebke KJ, Pecoraro VL (2018) Development of de Novo Copper Nitrite Reductases: Where We Are and Where We Need to Go. ACS Catal. 8:8046–8057.
11. Lombardi A, Pirro F, Maglio O, Chino M, Degrado WF (2019) De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. Acc. Chem. Res. 52:1148–1159.
12. Chalkley MJ, Mann SI, DeGrado WF (2022) De novo metalloprotein design. Nat. Rev. Chem. 6:31–50.
13. Naudin EA, Albanese KI, Smith AJ, Mylemans B, Baker EG, Weiner OD, Andrews DM, Tigue N, Savery NJ, Woolfson DN (2022) From peptides to proteins: coiled-coil tetramers to single-chain 4-helix bundles. Chem. Sci. [Internet] 13:11330–11340. Available from: http://xlink.rsc.org/?DOI=D2SC04479J
14. Hecht MH, Richardson JS, Richardson DC, Ogden RC (1990) De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. Science (80-.). [Internet] 249:884–891. Available from: https://www.science.org/doi/10.1126/science.2392678
15. Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'brien R, Sturtevant JM, Regan L (1996) What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Sci. 5:1584–1593.
16. Discher BM, Koder RL, Moser CC, Dutton PL (2003) Hydrophilic to amphiphilic design in redox protein

maquettes. Curr. Opin. Chem. Biol. 7:741–748.
17. Huang SS, Gibney BR, Stayrook SE, Leslie Dutton P, Lewis M (2003) X-ray structure of a maquette scaffold. J. Mol. Biol. 326:1219–1225.
18. Lichtenstein BR, Farid TA, Kodali G, Solomon LA, Anderson JLR, Sheehan MM, Ennist NM, Fry BA, Chobot SE, Bialas C, et al. (2012) Engineering oxidoreductases: maquette proteins designed from scratch. Biochem. Soc. Trans. 40:561–566.
19. Farid TA, Kodali G, Solomon LA, Lichtenstein BR, Sheehan MM, Fry BA, Bialas C, Ennist NM, Siedlecki JA, Zhao Z, et al. (2013) Elementary tetrahelical protein design for diverse oxidoreductase functions. Nat. Chem. Biol. 9:826–833.
20. Moser CC, Sheehan MM, Ennist NM, Kodali G, Bialas C, Englander MT, Discher BM, Dutton PL De Novo Construction of Redox Active Proteins. In: Methods in Enzymology. Vol. 580. Academic Press Inc.; 2016. pp. 365–388.
21. Kodali G, Mancini JA, Solomon LA, Episova T V., Roach N, Hobbs CJ, Wagner P, Mass OA, Aravindu K, Barnsley JE, et al. (2016) Design and engineering of water-soluble light-harvesting protein maquettes. Chem. Sci. 8:316–324.
22. Ennist NM, Zhao Z, Stayrook SE, Discher BM, Dutton PL, Moser CC (2022) De novo protein design of photochemical reaction centers. Nat. Commun. [Internet] 13:4937. Available from: https://www.nature.com/articles/s41467-022-32710-5
23. Cohen-Ofri I, Van Gastel M, Grzyb J, Brandis A, Pinkas I, Lubitz W, Noy D (2011) Zinc-bacteriochlorophyllide dimers in de novo designed four-helix bundle proteins. A model system for natural light energy harvesting and dissipation. J. Am. Chem. Soc. 133:9526–9535.
24. Mann SI, Nayak A, Gassner GT, Therien MJ, Degrado WF (2021) De Novo Design, Solution Characterization, and Crystallographic Structure of an Abiological Mn-Porphyrin-Binding Protein Capable of Stabilizing a Mn(V) Species. J. Am. Chem. Soc. 143:252–259.
25. Koepke J, Hu X, Muenke C, Schulten K, Michel H (1996) The crystal structure of the light-harvesting complex II (B800-850) from Rhodospirillum molischianum. Structure 4:581–597.
26. Tsargorodska A, Cartron ML, Vasilev C, Kodali G, Mass OA, Baumberg JJ, Dutton PL, Hunter CN, Törmä P, Leggett GJ (2016) Strong Coupling of Localized Surface Plasmons to Excitons in Light-Harvesting Complexes. Nano Lett. 16:6850–6856.
27. Lishchuk A, Kodali G, Mancini JA, Broadbent M, Darroch B, Mass OA, Nabok A, Dutton PL, Hunter CN, Törmä P, et al. (2018) A synthetic biological quantum optical system. Nanoscale 10:13064–13073.
28. Sutherland GA, Grayson KJ, Adams NBP, Mermans DMJ, Jones AS, Robertson AJ, Auman DB, Brindley AA, Sterpone F, Tuffery P, et al. (2018) Probing the quality control mechanism of the Escherichia coli twin-arginine translocase with folding variants of a de novo– designed heme protein. J. Biol. Chem. 293:6672–6681.
29. Sutherland GA, Polak D, Swainsbury DJK, Wang S, Spano FC, Auman DB, Bossanyi DG, Pidgeon JP, Hitchcock A, Musser AJ, et al. (2020) A Thermostable Protein Matrix for Spectroscopic Analysis of Organic Semiconductors. J. Am. Chem. Soc. 142:13898–13907.
30. Designed Research; J DBY (2020) Improved protein structure prediction using predicted interresidue orientations. 117:1496–1503.
31. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D (2020) Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. [Internet] 117:1496–1503. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1914677117
32. Du Z, Su H, Wang W, Ye L, Wei H, Peng Z, Anishchenko I, Baker D, Yang J (2021) The trRosetta server for fast and accurate protein structure prediction. Nat. Protoc. [Internet] 16:5634–5651. Available from: https://www.nature.com/articles/s41596-021-00628-9
33. Eichwurzel I, Stiel H, Röder B (2000) Photophysical studies of the pheophorbide a dimer. J. Photochem. Photobiol. B Biol. 54:194–200.
34. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, Yu H, Wu Z, Schulten K (2015) Molecular dynamics simulations of large macromolecular complexes. Curr. Opin. Struct. Biol. [Internet] 31:64–74. Available from: http://dx.doi.org/10.1016/j.sbi.2015.03.007
35. Amaro RE (2017) A Reflection on Klaus Schulten. J. Chem. Theory Comput. [Internet] 13:1–2. Available from: https://pubs.acs.org/doi/10.1021/acs.jctc.6b01111
36. Goh BC, Hadden JA, Bernardi RC, Singharoy A, McGreevy R, Rudack T, Cassidy CK, Schulten K (2016) Computational Methodologies for Real-Space Structural Refinement of Large Macromolecular Complexes.

Annu. Rev. Biophys. [Internet] 45:253–278. Available from: https://www.annualreviews.org/doi/10.1146/annurev-biophys-062215-011113
37. Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. Curr. Opin. Struct. Biol. [Internet] 11:224–230. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0959440X00001949
38. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K (2009) Discovery Through the Computational Microscope. Structure [Internet] 17:1295–1306. Available from: http://dx.doi.org/10.1016/j.str.2009.09.001
39. Strümpfer J, Schulten K (2009) Light harvesting complex II B850 excitation dynamics. J. Chem. Phys. [Internet] 131:225101. Available from: http://aip.scitation.org/doi/10.1063/1.3271348
40. Damjanović A, Ritz T, Schulten K (1999) Energy transfer between carotenoids and bacteriochlorophylls in light-harvesting complex II of purple bacteria. Phys. Rev. E [Internet] 59:3293–3311. Available from: https://linkinghub.elsevier.com/retrieve/pii/0006300259903282
41. Hu X, Schulten K (1998) Model for the Light-Harvesting Complex I (B875) of Rhodobacter sphaeroides. Biophys. J. [Internet] 75:683–694. Available from: http://dx.doi.org/10.1016/S0006-3495(98)77558-7
42. Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim. Biophys. Acta - Gen. Subj. [Internet] 1850:872–877. Available from: http://dx.doi.org/10.1016/j.bbagen.2014.10.019
43. Damjanović A, Kosztin I, Kleinekathöfer U, Schulten K (2002) Excitons in a photosynthetic light-harvesting system: A combined molecular dynamics, quantum chemistry, and polaron model study. Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 65.
44. Hu X, Ritz T, Damjanović A, Schulten K (1997) Pigment Organization and Transfer of Electronic Excitation in the Photosynthetic Unit of Purple Bacteria. J. Phys. Chem. B [Internet] 101:3854–3871. Available from: https://pubs.acs.org/doi/10.1021/jp963777g
45. Hu X, Damjanović A, Ritz T, Schulten K (1998) Architecture and mechanism of the light-harvesting apparatus of purple bacteria. Proc. Natl. Acad. Sci. [Internet] 95:5935–5941. Available from: https://pnas.org/doi/full/10.1073/pnas.95.11.5935
46. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K (2006) Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. Structure [Internet] 14:437–449. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0969212606000608
47. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, et al. (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature [Internet] 497:643–646. Available from: http://www.nature.com/articles/nature12162
48. Ennist NM (2017) Design, Structure, and Action of an Artificial Photosynthetic Reaction Centre.
49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.
50. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Dustin Schaeffer R, et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science (80-.). [Internet] 373:871–876. Available from: https://www.science.org/doi/10.1126/science.abj8754
51. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J. Chem. Inf. Model. 51:69–82.
52. Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A (2018) Cryptic binding sites on proteins: definition, detection, and druggability. Curr. Opin. Chem. Biol. [Internet] 44:1–8. Available from: https://doi.org/10.1016/j.cbpa.2018.05.003
53. Hestand NJ, Spano FC (2018) Expanded Theory of H- and J-Molecular Aggregates: The Effects of Vibronic Coupling and Intermolecular Charge Transfer. Chem. Rev. [Internet] 118:7069–7163. Available from: https://pubs.acs.org/doi/10.1021/acs.chemrev.7b00581
54. Monshouwer R, Abrahamsson M, van Mourik F, van Grondelle R (1997) Superradiance and Exciton Delocalization in Bacterial Photosynthetic Light-Harvesting Systems. J. Phys. Chem. B [Internet] 101:7241–7248. Available from: https://pubs.acs.org/doi/10.1021/jp963377t
55. Becker M, Nagarajan V, Parson WW (1991) Properties of the excited singlet states of bacteriochlorophyll a and bacteriopheophytin a in polar solvents. J. Am. Chem. Soc. [Internet] 113:6840–6848. Available from:

https://pubs.acs.org/doi/abs/10.1021/ja00018a020
56. Kasha M, Rawls HR, Ashraf El-Bayoumi M (1965) The exciton model in molecular spectroscopy. Pure Appl. Chem. [Internet] 11:371–392. Available from: https://www.degruyter.com/document/doi/10.1351/pac196511030371/html
57. Hart SM, Banal JL, Castellanos MA, Markova L, Vyborna Y, Gorman J, Häner R, Willard AP, Bathe M, Schlau-Cohen GS (2022) Activating charge-transfer state formation in strongly-coupled dimers using DNA scaffolds. Chem. Sci. [Internet]. Available from: http://xlink.rsc.org/?DOI=D2SC02759C
58. Curutchet C, Kongsted J, Muñoz-Losa A, Hossein-Nejad H, Scholes GD, Mennucci B (2011) Photosynthetic light-harvesting is tuned by the heterogeneous polarizable environment of the protein. J. Am. Chem. Soc. 133:3078–3084.
59. Corbella M, Cupellini L, Lipparini F, Scholes GD, Curutchet C (2019) Spectral Variability in Phycocyanin Cryptophyte Antenna Complexes is Controlled by Changes in the α-Polypeptide Chains. ChemPhotoChem 3:945–956.
60. Wahadoszamen M, Margalit I, Ara AM, van Grondelle R, Noy D (2014) The role of charge-transfer states in energy transfer and dissipation within natural and artificial bacteriochlorophyll proteins. Nat. Commun. [Internet] 5:5287. Available from: http://www.nature.com/articles/ncomms6287
61. Christensen H, Pain RH (1991) Molten globule intermediates and protein folding. Eur. Biophys. J. 19:221–229.
62. Lau SYM, Taneja AK, Hodges RS (1984) Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded α-helical coiled-coils. J. Biol. Chem. 259:13253–13261.
63. Zhou NE, Kay CM, Hodges RS (1992) Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded α-helical coiled-coils. J. Biol. Chem. 267:2664–2670.
64. López-García P, Goktas M, Bergues-Pupo AE, Koksch B, Varón Silva D, Blank KG (2019) Structural determinants of coiled coil mechanics. Phys. Chem. Chem. Phys. 21:9145–9149.
65. Nagy G, Igaev M, Jones NC, Hoffmann S V., Grubmüller H (2019) SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. J. Chem. Theory Comput. [Internet] 15:5087–5102. Available from: https://pubs.acs.org/doi/10.1021/acs.jctc.9b00203
66. Björn LO, Papageorgiou GC, Blankenship RE, Govindjee (2009) A viewpoint: Why chlorophyll a? Photosynth. Res. [Internet] 99:85–98. Available from: http://link.springer.com/10.1007/s11120-008-9395-x
67. Cupellini L, Caprasecca S, Guido CA, Müh F, Renger T, Mennucci B (2018) Coupling to Charge Transfer States is the Key to Modulate the Optical Bands for Efficient Light Harvesting in Purple Bacteria. J. Phys. Chem. Lett. 9:6892–6899.
68. Barros EP, Schiffer JM, Vorobieva A, Dou J, Baker D, Amaro RE (2019) Improving the Efficiency of Ligand-Binding Protein Design with Molecular Dynamics Simulations. J. Chem. Theory Comput. [Internet] 15:5703–5715. Available from: https://pubs.acs.org/doi/10.1021/acs.jctc.9b00483
69. Selvan D, Prasad P, Farquhar ER, Shi Y, Crane S, Zhang Y, Chakraborty S (2019) Redesign of a Copper Storage Protein into an Artificial Hydrogenase. ACS Catal. [Internet] 9:5847–5859. Available from: https://pubs.acs.org/doi/10.1021/acscatal.9b00360
70. Pearce R, Zhang Y (2021) Toward the solution of the protein structure prediction problem. J. Biol. Chem. 297.
71. Garab G, van Amerongen H (2009) Linear dichroism and circular dichroism in photosynthesis research. Photosynth. Res. 101:135–146.
72. Banal JL, Kondo T, Veneziano R, Bathe M, Schlau-Cohen GS (2017) Photophysics of J-Aggregate-Mediated Energy Transfer on DNA. J. Phys. Chem. Lett. [Internet] 8:5827–5833. Available from: https://pubs.acs.org/doi/10.1021/acs.jpclett.7b01898
73. Álvarez-Moreno M, De Graaf C, López N, Maseras F, Poblet JM, Bo C (2015) Managing the computational chemistry big data problem: The ioChem-BD platform. J. Chem. Inf. Model. 55:95–103.