

Engineering excitonically coupled dimers in an artificial protein for light harvesting via computational modeling

Mariano Curti¹  | Valentin Maffei^{1,2}  |

Luís Gustavo Teixeira Alves Duarte¹  | Saeed Shareef^{1,3}  |

Luisa Xiomara Hallado^{1,3}  | Carles Curutchet^{4,5}  | Elisabet Romero¹ 

¹Institute of Chemical Research of Catalonia (ICIQ), Barcelona Institute of Science and Technology (BIST), Tarragona, Spain

²Laboratoire de Chimie, UMR 5182, ENS Lyon, CNRS, Université Lyon 1, Lyon, France

³Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, Tarragona, Spain

⁴Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Barcelona, Spain

⁵Institut de Química Teòrica i Computacional (IQTUB), Universitat de Barcelona (UB), Barcelona, Spain

Correspondence

Mariano Curti and Elisabet Romero, Avda. Països Catalans 16, 43007 Tarragona, Spain.

Email: mcurti@iciq.es and eromero@iciq.es

Funding information

European Research Council, Grant/Award Numbers: 754510, 801474, 805524; State Research Agency/Spanish Ministry of Science and Innovation, Grant/Award Numbers: MDM-2017-0767, PID2020-115812GB-I00, CEX2019-0009; Red Española de Supercomputación, Grant/Award Numbers: QSB-2020-3-0002, QSB-2021-1-0019

Abstract

In photosynthesis, pigment–protein complexes achieve outstanding photoinduced charge separation efficiencies through a set of strategies in which excited states delocalization 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 realizing 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 optimized 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 modeling, 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

artificial photosynthesis, cofactor binding, de novo protein design, molecular dynamics

Reviewing Editor: Nir Ben-Tal

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

In photosynthesis, Light-Harvesting Complexes (LHC) are responsible for photon absorption and subsequent energy transfer to Reaction Centers (RC), where charge separation takes place to initiate a chain of reactions that ultimately store solar energy as chemical bonds (Blankenship, 2002; Scholes et al., 2011). 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 (Beddard & Porter, 1976). Excitation energy is then funneled to RCs, where, using a set of strategies termed “design principles of charge separation”, (Romero et al., 2017) 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 (Romero et al., 2017), finely tuned by millions of years of evolution. Among other features, excitonic interactions and charge-transfer states play preponderant roles in these systems (Frese et al., 2003; Lockhart & Boxer, 1988; Romero et al., 2012; Romero et al., 2017).

Modification of natural systems to fulfill 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 (Scott et al., 2021) or multi-electron oxidases/reductases (Kaplan & DeGrado, 2004; Koebke & Pecoraro, 2018; Lombardi et al., 2019). 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 (Chalkley et al., 2022; Naudin et al., 2022). 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 (Discher et al., 2003; Farid et al., 2013; Hecht et al., 1990; Huang et al., 2003; Lichtenstein et al., 2012; Moser et al., 2016; Munson et al., 1996). These complexes have already shown basic light absorption and energy transfer functionalities (Ennist et al., 2022; Farid et al., 2013; Kodali et al., 2016). 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 (Cohen-Ofri et al., 2011). 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 (Mann et al., 2021). 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. Molischanium*: 24%, only considering bacteriochlorophylls) (Koepke et al., 1996).

Here, we analyze and optimize a 132 amino acids-long maquette protein developed by the Dutton group, which has been previously employed to bind different cofactors (Farid et al., 2013; Kodali et al., 2016). 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.

2 | RESULTS

2.1 | Binding to original structures

We start the analysis by considering the BT6 protein (Farid et al., 2013; Lishchuk et al., 2018; Sutherland et al., 2018; Tsargorodska et al., 2016), which has four histidine residues (Figure 1a), and two mutants in which either two (Kodali et al., 2016) (2H → 2A variant) or all four (Sutherland et al., 2020) (4H → 4A) His were replaced by alanine (Ala, see sequences in Table 1). With light-harvesting functionality in mind, we have chosen Zn 13²-OH-methylpheophorbide *a* (Figure 1b), a chlorophyll *a* derivative, as the chromophoric cofactor.

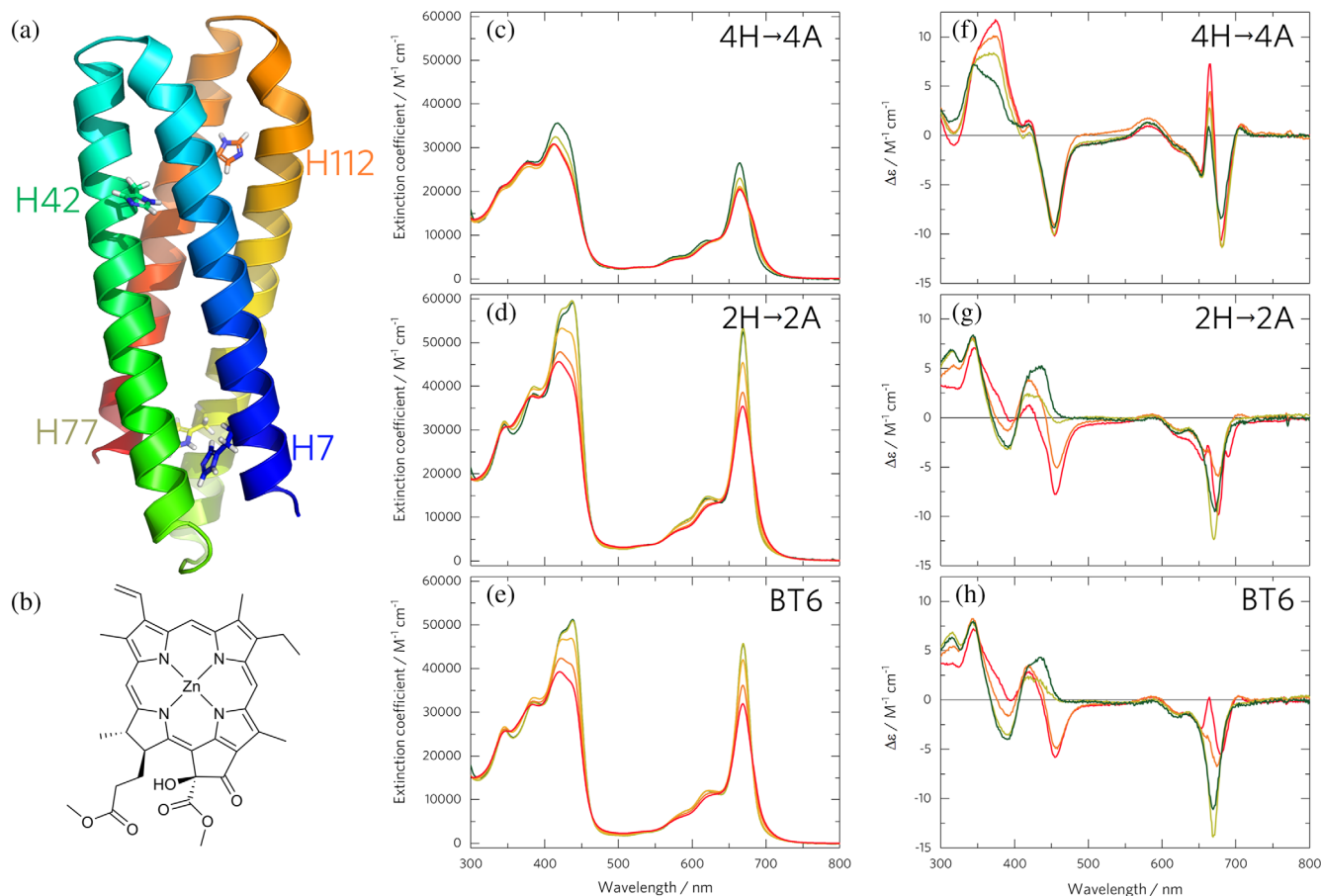


FIGURE 1 (a) Protein structure for BT6 as predicted with trRosetta (Du et al., 2021). Histidine residues are shown in sticks representation. (b) Chemical structure of Zn 13^2 -OH-methylpheophorbide *a*. (c-h) Chromophore to protein binding titration experiments monitored by absorption (center) and circular dichroism (right) spectroscopy, for 4H \rightarrow 4A (c and f), 2H \rightarrow 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 μ M protein concentration. For circular dichroism spectra, chromophore:protein ratios are shown from 1 (green) to 4 (red), at a 20 μ M protein concentration. The proteins are in a buffer solution with 20 mM CHES at pH 9 and 150 mM KCl.

Although crystallization efforts on BT6 have so far been unsuccessful, protein structure prediction software (Du et al., 2021; Yang et al., 2020) yields a model (Figure 1a) consistent with insights from previous studies (Farid et al., 2013). 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 (Moser et al., 2016). 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 (Farid et al., 2013). 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 \rightarrow 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 Section S1.5). As monitored by absorption spectroscopy (Figure 1c), chromophore addition to a solution of the 4H \rightarrow 4A variant simply results in chromophore dissolution and subsequent aggregation (Eichwurz et al., 2000), because this protein has no His residues to ligate it. The aggregation process is evinced by the progressive formation of a red shoulder at \sim 683 nm on the first electronic transition, termed Q_y , and initially centered at \sim 664 nm. The Soret region (300–450 nm) shows a complex pattern with several transitions. A decrease in the extinction coefficient of (at least) two overlapping bands centered at \sim 413 and 435 nm is clear as chromophore concentration increases and aggregation proceeds. Control experiments in the absence of protein

TABLE 1 Amino acid sequences of the artificial proteins employed in this study, along with the dissociation constants for cofactor binding

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)

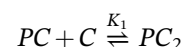
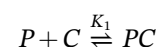
Note: The two values, $K_{D,1}$ and $K_{D,2}$, correspond to the dissociation constants for the first and second pair of cofactors, respectively. The uncertainty in the pK values, obtained from duplicate experiments, is ~ 0.5 .

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 centered 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 behavior of 4H → 4A. The lower-energy side of the Soret region (400–450 nm) shows a different behavior as well when compared with 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:



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.

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 behavior 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.

2.2 | 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 (Amaro, 2017; Goh et al., 2016; Isralewitz et al., 2001; Lee et al., 2009; Perilla et al., 2015) and co-workers for LH complexes (Bernardi et al., 2015; Damjanović et al., 1999; Damjanović et al., 2002; Hu & Schulten, 1998; Strümpfer & Schulten, 2009) and even for more complex multi-component biological systems (Freddolino et al., 2006; Hu et al., 1997; Hu et al., 1998; Zhao et al., 2013). 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 (Ennist, 2017). Nevertheless, the simplicity and strong foldability of maquette proteins render them good candidates for structure prediction software (Naudin et al., 2022). Indeed, all tested programs

(Baek et al., 2021; Du et al., 2021; Jumper et al., 2021) 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 Section S2.3) 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 Note S1), we generated four complexes (Figure 2), labeled 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 neighboring or non-neighboring 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, that is, 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 (Hou et al., 2011). Nevertheless, some factors affecting the absolute energy values are worth discussing. The binding free energy ΔG_{bind} is calculated as:

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

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:

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

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-

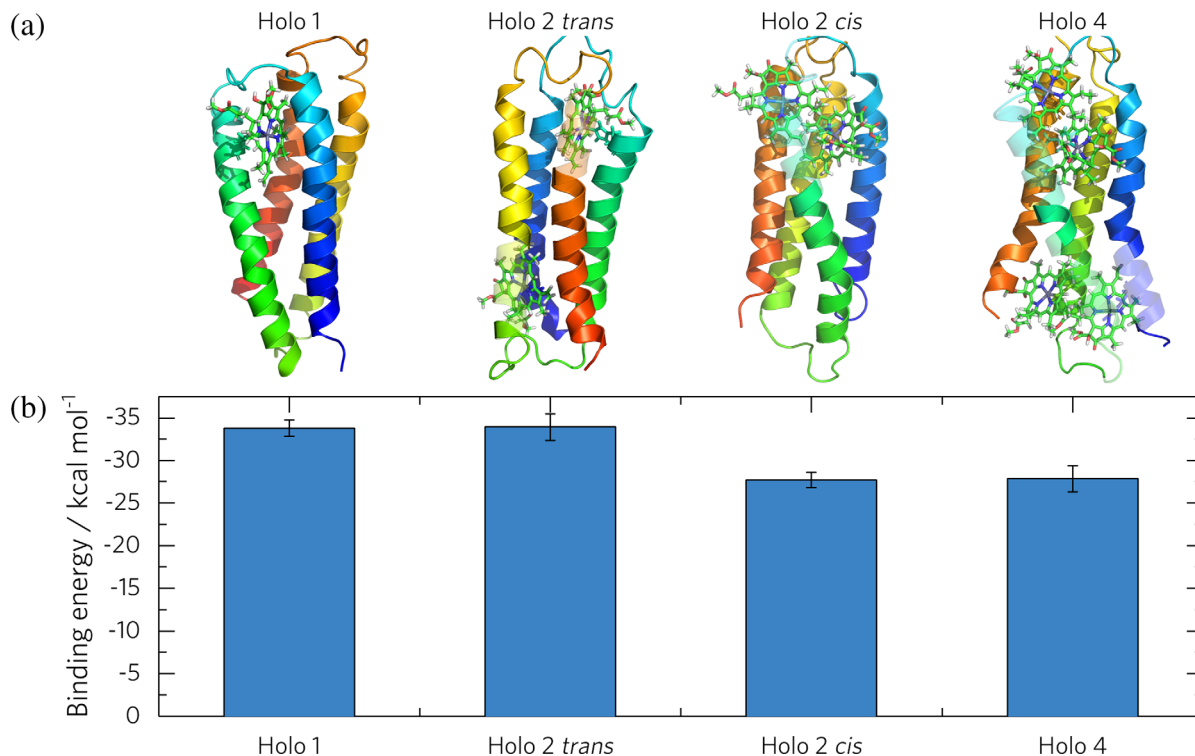


FIGURE 2 Representative structures (a) and binding free energies per chromophore molecule (b), as obtained from MD simulations for BT6. The structures differ on the number of bound chromophores, and in the case of Holo 2, whether the two chromophores are bound at neighboring (*cis*) or non-neighboring (*trans*) His residues. Binding free energies were obtained from 1.0 μ s production runs, employing QM(PM6)-MMGBSA.

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 (Hou et al., 2011). 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 reparametrized the free chromophore using the same procedure but in absence of protein, and run a 1.0 μ s

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 with 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 neighboring 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-neighboring 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 unfavored.

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 (center-to-center), 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 \rightarrow 2A variant at $[C]:[P] \leq 2$ (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-neighboring His residues, similarly to the 2H \rightarrow 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 neighboring and non-neighboring chromophore pairs.

2.3 | Computational optimization 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 favorably 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 μ s 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 favorable 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 S2.1: Apo-structures generation). Two candidates for point mutation emerge from this description: L46A/L116A (symmetry related, Figure 3b) and E113K (Figure 3c). The remaining mutants can be discarded on the basis of the aforementioned criteria (as detailed in the Note S2).

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

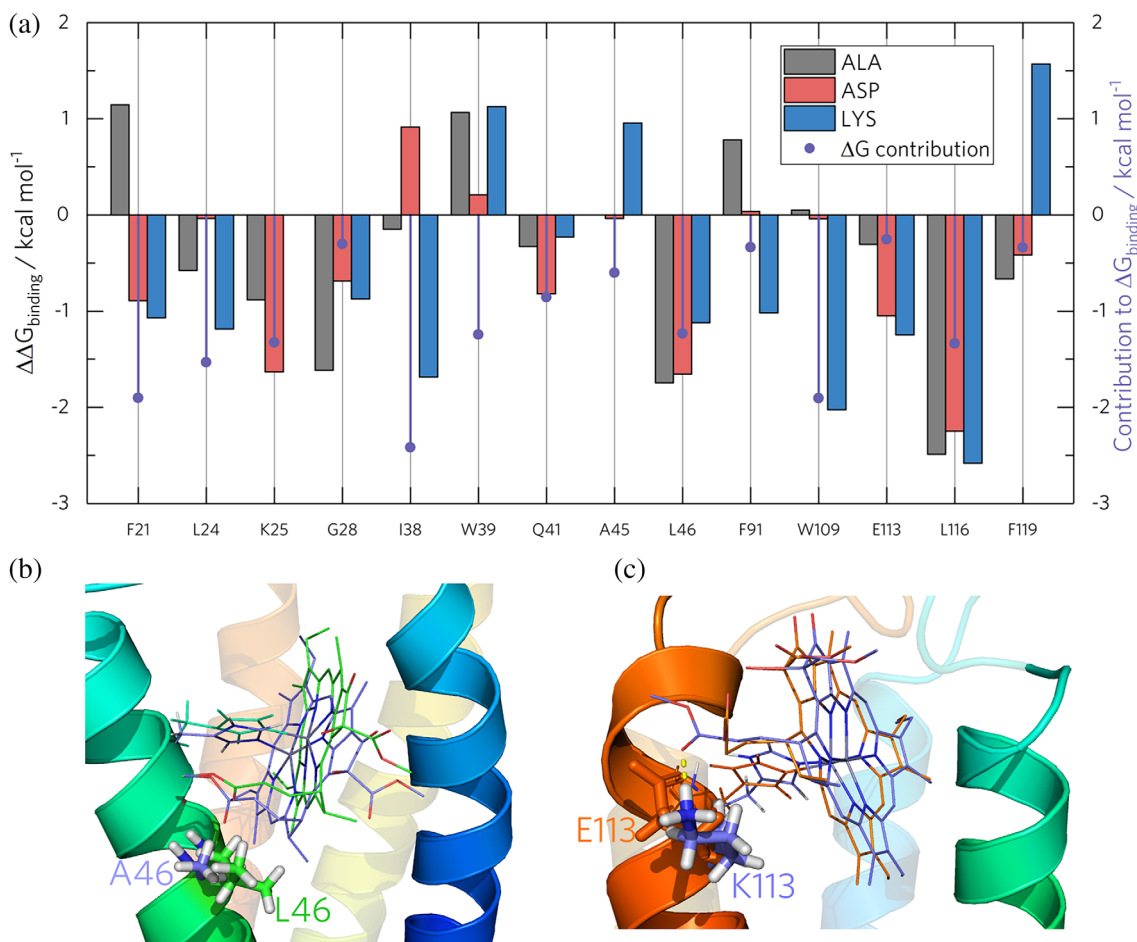


FIGURE 3 (a) Contribution to the binding energy of amino acids close to the chromophore pair in Holo 2 *cis*, together with the difference in binding energy after point mutation of each residue by Ala, Asp, or Lys, as calculated from MD simulations. Binding energies were calculated with QM(PM6)-MMGBSA 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 colored 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.

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 \rightarrow 4A for short).

As a second candidate, we analyzed 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 unfavorable interaction with the chromophore, while providing a favorable electrostatic interaction with the positively charged Lys. These interactions are depicted in Figure 3c (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 \rightarrow 4K for short) for experimental corroboration and analysis, together with the above-described 4L \rightarrow 4A (their sequences are shown in Table 1).

2.4 | Binding to optimized proteins and spectroscopic properties

To analyze chromophore binding, we performed titrations of the two newly proposed variants (4L \rightarrow 4A and 4E \rightarrow 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 \rightarrow 2A: pK values increase from 6.8 for these two, up to 7.4 and 8.2 for 4E \rightarrow 4K and 4L \rightarrow 4A, respectively. Moreover, although the evolution of the spectral shapes is superficially similar, SVD analysis reveals an important

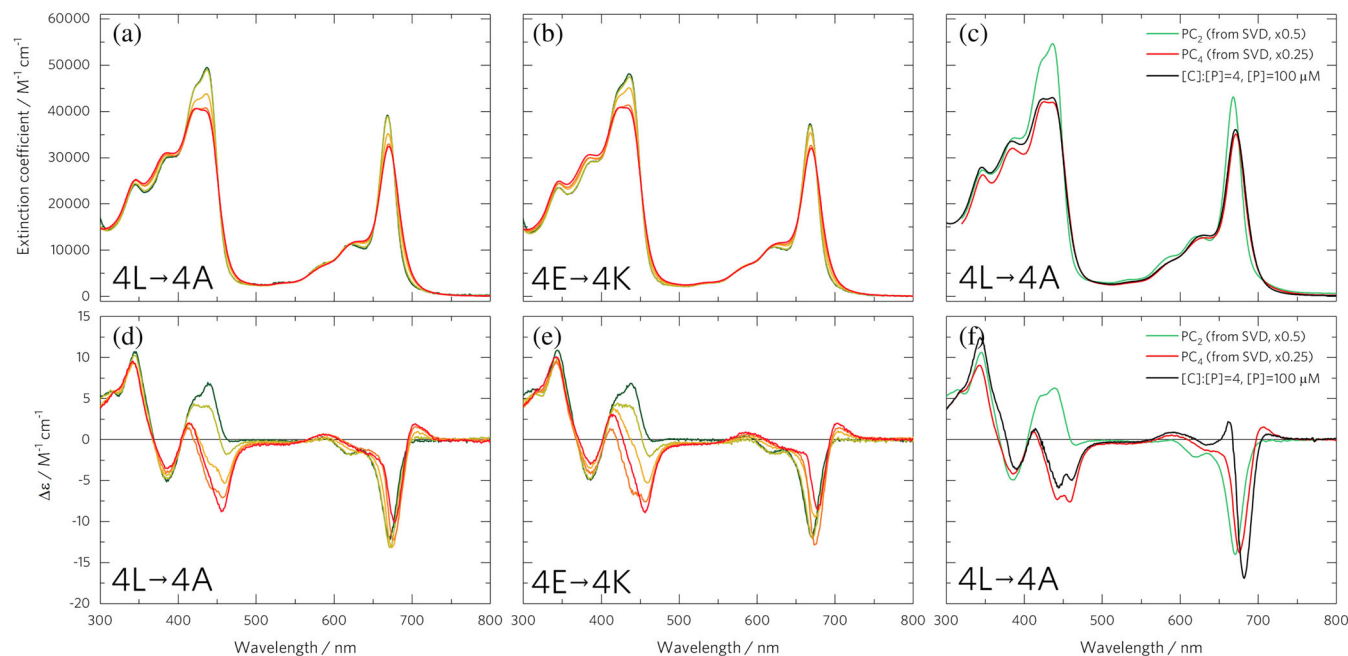


FIGURE 4 Chromophore to protein binding titration experiments monitored by absorption (a and b) and circular dichroism (d and e) spectroscopy, for the 4L → 4A (a and d) and 4E → 4K (b and e) variants. For absorption spectra the protein concentration is 5 μM , whereas for circular dichroism experiments 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.

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 (Figures 4 and 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 behavior than BT6 for [C]:[P] \leq 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 \sim 700 nm and a redshift and decrease of the negative band initially centered 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 (Figures 4 and 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.

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 with those for PC_4 extracted from SVD, and indeed the main features coincide (Figure 4c, f). An important difference is the appearance of a positive feature at \sim 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 analyze the properties of the excitonic excited states we performed steady-state fluorescence. From the SVD

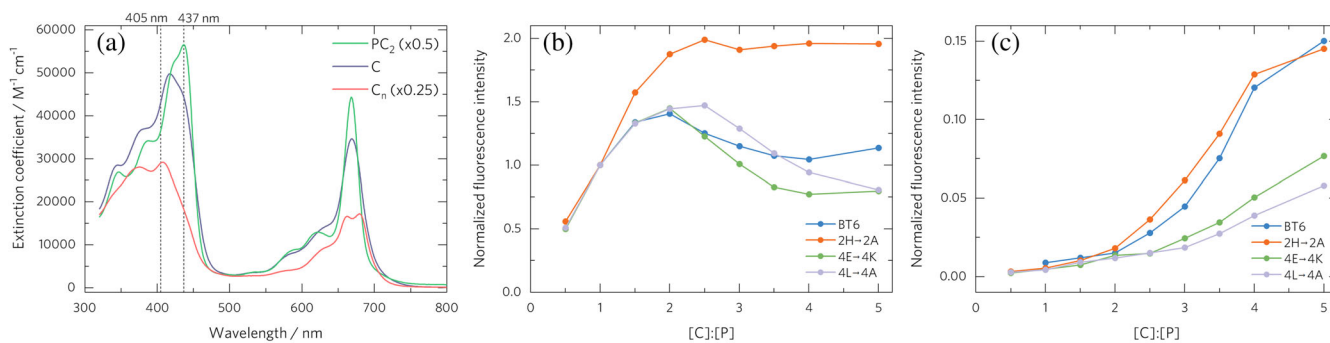


FIGURE 5 (a) Absorption spectra obtained from singular value decomposition of the titrations (4L → 4A variant). The vertical lines highlight the two excitation wavelengths, 405 and 437 nm. (b) Fluorescence emission intensity at 673 nm, with an excitation wavelength of 437 nm. (c) Fluorescence emission 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.

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 analyze 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 behavior: the emission intensity increases for [C]:[P] from 0 to 2, and then it promptly decreases for $2 \leq [C]:[P] \leq 4$. For the 4E → 4K and 4L → 4A variants, this behavior 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) (Hestand & Spano, 2018), 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 $2 \leq [C]:[P] \leq 4$. Although the rest of the spectroscopic evidence points towards 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 5c). 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] ≥ 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), for example, 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 Figures 5c and S14). While employing larger concentrations could minimize this issue, it has the undesired effect of increasing the optical densities of the solutions, thus exacerbating

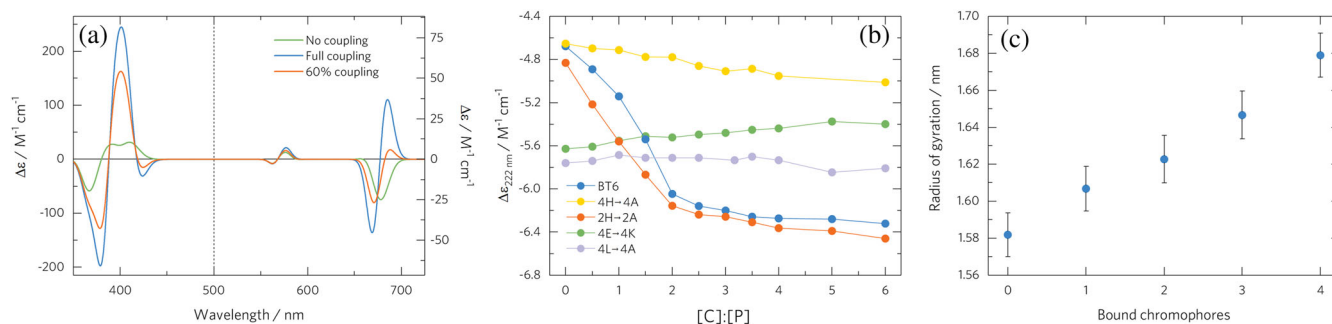


FIGURE 6 (a) Simulated circular dichroism spectra of Holo 2 *cis*, using no (green), scaled (orange), or full (blue) coupling between cofactors. Spectra are the average of 20 calculations performed along 1.0 μs trajectories at the CAM-B3LYP/6-31G*/PCM level, and were shifted -0.25 eV to reproduce the experimental position of the Q_y band. To facilitate visualization, different y-scales are used for the ranges 350–500 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.

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 $\sim 5\%$ 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 LH1 and LH2 complexes of purple bacteria display fluorescence quantum yields of 8%–10% (Monshouwer et al., 1997) (for monomeric bacteriochlorophyll in polar solvents, the value ranges from 11 to 20% (Becker et al., 1991)).

2.5 | Modeling of the electronic properties

We resort again to computational modeling, 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 (Hart et al., 2022; Kasha et al., 1965). In 4E → 4K, for instance, the 0–0 / 0–1 intensity ratio (at ~ 669 and ~ 625 nm, respectively) decreases from 3.8 to 3.0 when comparing PC_2 to PC_4 (Figure S13), indicating relatively weak coupling (Hart et al., 2022).

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 (Curutchet et al., 2011). The heterogeneous polarizable environment of the protein leads to a $\sim 31\%$ attenuation of the coupling, compared with the continuum assumption (Curutchet et al., 2011). Similar strong screening effects have also been observed for the central dimers of the PC630 and PC645 complexes (Corbella et al., 2019). 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

modeling, 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 (Wahadoszamen et al., 2014). In addition, the maquette complex showed a decrease in fluorescence intensity similar to the present case (Cohen-Ofri et al., 2011), while the B820 subunit displays a fluorescence quantum yield of 3% (Monshouwer et al., 1997), 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.

2.6 | 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*) (Koepke et al., 1996), it is informative to analyze 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 (Christensen & Pain, 1991; Farid et al., 2013; Huang et al., 2003; Munson et al., 1996). At the same time, cofactor binding induces a strong structuring (Farid et al., 2013; Huang et al., 2003). Indeed, both BT6 and the 2H → 2A variant display a marked increase in the signal at 222 nm for $[C]:[P] \leq 2$, 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 behavior: 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 (Lau et al., 1984; López-García et al., 2019; Zhou et al., 1992). 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 $69 \pm 3\%$ for the apo structure, that is, 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 (Nagy et al., 2019) yields alpha helix fractions close to 70% for both new variants.

Chromophore uptake decreases the alpha-helix fraction: in the calculated Holo 4 structure, $61 \pm 3\%$ of the residues adopt a helical secondary structure (vs. 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 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 with Holo 2 *trans* (Figure 2), explaining the decrease in the 222 nm to 208 nm ratio at $[C]:[P] \geq 2$ for 4L \rightarrow 4A and 4E \rightarrow 4K.

3 | 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 (Romero et al., 2017). Implementation of these concepts in artificial systems is, however, much harder, and the first fruitful attempts have been attained only recently (Ennist et al., 2022; Farid et al., 2013). Our work illustrates one possible pathway towards that goal, employing MD simulations to effectively optimize 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 (Björn et al., 2009). Instead of designing a new protein in an ad hoc fashion, we instead employ BT6, a versatile four- α -helix structure that has already shown strong cofactor binding (up to two per protein) and photoinduced energy transfer (Farid et al., 2013).

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 \rightarrow 2A) and computational modeling indicate that BT6 can easily accommodate two chromophores at non-neighboring 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) (Cohen-Ofri et al., 2011).

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 neighboring 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-neighboring His residues, and later at neighboring 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 $2 \leq [C]:[P] \leq 4$. The relatively small excitonic coupling, together with prior Stark spectroscopy results on related systems (Wahadoszamen et al., 2014) 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 (Romero et al., 2017). 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 (Cupellini et al., 2018).

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 (Barros et al., 2019). 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 (Selvan et al., 2019), 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 (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021) methods, have recently achieved remarkable accuracy, thus offering an attractive alternative to experimental determination (Pearce & Zhang, 2021). 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 (Garab & van Amerongen, 2009). 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 (Ennist, 2017). 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 (Banal et al., 2017; Hart et al., 2022). However, exact control on the aggregates geometries proved to be difficult, as exemplified by inter-chromophore electrostatic interactions disrupting the DNA duplexes (Hart et al., 2022), or by excitonic interactions that spanned

only 2.5 chromophores despite the scaffolds providing a much larger number of binding sites (Banal et al., 2017). Artificial proteins, on the other hand, can provide a more robust and tailorable scaffold for controlled chromophore binding.

All in all, through the computational optimization 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.

4 | 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 realization of artificial protein-based photosynthetic systems.

AUTHOR CONTRIBUTIONS

Mariano Curti: Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); software (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

Valentin Maffei: Formal analysis (equal); investigation (equal); writing – original draft (equal); writing – review and editing (equal). **Luís Gustavo Teixeira Alves Duarte:** Formal analysis (equal); investigation (equal); writing – review and editing (equal). **Saeed Shareef:** Investigation (equal). **Luisa Xiomara Hallado:**

Investigation (equal). **Carles Curutchet**: Methodology (equal); supervision (equal); writing – review and editing (equal). **Elisabet Romero**: Conceptualization (equal); funding acquisition (equal); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

Mariano Curti acknowledges the funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 801474. Mariano Curti and Elisabet Romero acknowledge funding from the State Research Agency/Spanish Ministry of Science and Innovation (AEI/MICINN/10.13039/501100011033) through the Severo Ochoa Excellence Accreditation CEX2019-000925-S. Luís Gustavo Teixeira Alves Duarte, Saeed Shareef, Luisa Xiomara Hallado, and Elisabet Romero 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. Carles Curutchet 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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 (Álvarez-Moreno et al., 2015), and is accessible through the doi: [10.19061/iochem-bd-1-240](https://doi.org/10.19061/iochem-bd-1-240).

ORCID

Mariano Curti  <https://orcid.org/0000-0002-3721-4358>
 Valentin Maffei  <https://orcid.org/0000-0003-4672-8286>
 Luís Gustavo Teixeira Alves Duarte  <https://orcid.org/0000-0001-5146-2166>
 Saeed Shareef  <https://orcid.org/0000-0002-8412-1939>
 Luisa Xiomara Hallado  <https://orcid.org/0000-0003-3052-1504>
 Carles Curutchet  <https://orcid.org/0000-0002-0070-1208>
 Elisabet Romero  <https://orcid.org/0000-0003-3630-1617>

REFERENCES

- Álvarez-Moreno M, de Graaf C, López N, Maseras F, Poblet JM, Bo C. Managing the computational chemistry big data problem: the ioChem-BD platform. *J Chem Inf Model*. 2015;55:95–103.
- Amaro RE. A reflection on Klaus Schulten. *J. Chem. Theory Comput*. 2017;13:1–2. <https://doi.org/10.1021/acs.jctc.6b01111>
- Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (80–). 2021; 373:871–6. <https://doi.org/10.1126/science.abj8754>
- Banal JL, Kondo T, Veneziano R, Bathe M, Schlau-Cohen GS. Photophysics of J-aggregate-mediated energy transfer on DNA. *J Phys Chem Lett*. 2017;8:5827–33. <https://doi.org/10.1021/acs.jpcclett.7b01898>
- Barros EP, Schiffer JM, Vorobieva A, Dou J, Baker D, Amaro RE. Improving the efficiency of ligand-binding protein design with molecular dynamics simulations. *J. Chem. Theory Comput*. 2019;15:5703–15. <https://doi.org/10.1021/acs.jctc.9b00483>
- Becker M, Nagarajan V, Parson WW. Properties of the excited singlet states of bacteriochlorophyll a and bacteriopheophytin in a polar solvents. *J Am Chem Soc*. 1991;113:6840–8. <https://doi.org/10.1021/ja00018a020>
- Beddard GS, Porter G. Concentration quenching in chlorophyll. *Nature*. 1976;260:366–7.
- Bernardi RC, Melo MCR, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim Biophys Acta: Gen Subj*. 2015;1850:872–7. <https://doi.org/10.1016/j.bbagen.2014.10.019>
- Björn LO, Papageorgiou GC, Blankenship RE, Govindjee. A viewpoint: why chlorophyll a? *Photosynth Res*. 2009;99:85–98. <https://doi.org/10.1007/s11120-008-9395-x>
- Blankenship RE. *Molecular mechanisms of photosynthesis*. United Kingdom: Wiley; 2002. <https://doi.org/10.1002/9780470758472>
- Chalkley MJ, Mann SI, DeGrado WF. De novo metalloprotein design. *Nat Rev Chem*. 2022;6:31–50.
- Christensen H, Pain RH. Molten globule intermediates and protein folding. *Eur Biophys J*. 1991;19:221–9.
- Cohen-Ofri I, van Gestel M, Grzyb J, Brandis A, Pinkas I, Lubitz W, et al. 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*. 2011;133:9526–35.
- Corbella M, Cupellini L, Lipparini F, Scholes GD, Curutchet C. Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α -polypeptide chains. *ChemPhotoChem*. 2019;3:945–56.
- Cupellini L, Caprasecca S, Guido CA, Müh F, Renger T, Mennucci B. Coupling to charge transfer states is the key to modulate the optical bands for efficient light harvesting in purple bacteria. *J Phys Chem Lett*. 2018;9:6892–9.
- Curutchet C, Kongsted J, Muñoz-Losa A, Hossein-Nejad H, Scholes GD, Mennucci B. Photosynthetic light-harvesting is tuned by the heterogeneous polarizable environment of the protein. *J Am Chem Soc*. 2011;133:3078–84.
- Damjanović A, Kosztin I, Kleinekathöfer U, Schulten K. Excitons in a photosynthetic light-harvesting system: a combined molecular dynamics, quantum chemistry, and polaron model study. *Phys. Rev. E*. 2002;65.
- Damjanović A, Ritz T, Schulten K. Energy transfer between carotenoids and bacteriochlorophylls in light-harvesting complex II of purple bacteria. *Phys Rev E*. 1999;59:3293–311.

- Discher BM, Koder RL, Moser CC, Dutton PL. Hydrophilic to amphiphilic design in redox protein maquettes. *Curr Opin Chem Biol.* 2003;7:741–8.
- Du Z, Su H, Wang W, Ye L, Wei H, Peng Z, et al. The trRosetta server for fast and accurate protein structure prediction. *Nat Protoc.* 2021;16:5634–51.
- Eichwurz I, Stiel H, Röder B. Photophysical studies of the pheophorbide a dimer. *J Photochem Photobiol B Biol.* 2000;54:194–200.
- Ennist NM. Design, Structure, and Action of an Artificial Photosynthetic Reaction Centre. United States: University of Pennsylvania; 2017.
- Ennist NM, Zhao Z, Stayrook SE, Discher BM, Dutton PL, Moser CC. De novo protein design of photochemical reaction centers. *Nat Commun.* 2022;13:4937.
- Farid TA, Kodali G, Solomon LA, Lichtenstein BR, Sheehan MM, Fry BA, et al. Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat Chem Biol.* 2013;9:826–33.
- Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure.* 2006;14:437–49.
- Frese RN, Germano M, de Weerd FL, van Stokkum IHM, Shkuropatov AY, Shuvalov VA, et al. Electric field effects on the chlorophylls, pheophytins, and β -carotenes in the reaction center of photosystem II. *Biochem Int.* 2003;42:9205–13. <https://doi.org/10.1021/bi0273516>
- Garab G, van Amerongen H. Linear dichroism and circular dichroism in photosynthesis research. *Photosynth Res.* 2009;101:135–46.
- Goh BC, Hadden JA, Bernardi RC, Singharoy A, McGreevy R, Rudack T, et al. Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annu Rev Biophys.* 2016;45:253–78. <https://doi.org/10.1146/annurev-biophys-062215-011113>
- Hart SM, Banal JL, Castellanos MA, Markova L, Vyborna Y, Gorman J, et al. Activating charge-transfer state formation in strongly-coupled dimers using DNA scaffolds. *Chem Sci.* 2022;13:13020–31.
- Hecht MH, Richardson JS, Richardson DC, Ogden RC. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science (80-).* 1990;249:884–91. <https://doi.org/10.1126/science.2392678>
- Hestand NJ, Spano FC. Expanded theory of H- and J-molecular aggregates: the effects of vibronic coupling and intermolecular charge transfer. *Chem Rev.* 2018;118:7069–163. <https://doi.org/10.1021/acs.chemrev.7b00581>
- Hou T, Wang J, Li Y, Wang W. 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.* 2011;51:69–82.
- Hu X, Damjanović A, Ritz T, Schulten K. Architecture and mechanism of the light-harvesting apparatus of purple bacteria. *Proc Natl Acad Sci.* 1998;95:5935–41. <https://doi.org/10.1073/pnas.95.11.5935>
- Hu X, Ritz T, Damjanović A, Schulten K. Pigment organization and transfer of electronic excitation in the photosynthetic unit of purple bacteria. *J Phys Chem B.* 1997;101:3854–71. <https://doi.org/10.1021/jp963777g>
- Hu X, Schulten K. Model for the light-harvesting complex I (B875) of *Rhodobacter sphaeroides*. *Biophys J.* 1998;75:683–94. [https://doi.org/10.1016/S0006-3495\(98\)77558-7](https://doi.org/10.1016/S0006-3495(98)77558-7)
- Huang SS, Gibney BR, Stayrook SE, Leslie Dutton P, Lewis M. X-ray structure of a maquette scaffold. *J Mol Biol.* 2003;326:1219–25.
- Israelowitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol.* 2001;11:224–30.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9.
- Kaplan J, DeGrado WF. De novo design of catalytic proteins. *Proc Natl Acad Sci.* 2004;101:11566–70. <https://doi.org/10.1073/pnas.0404387101>
- Kasha M, Rawls HR, Ashraf El-Bayoumi M. The exciton model in molecular spectroscopy. *Pure Appl Chem.* 1965;11:371–92. <https://doi.org/10.1351/pac196511030371/html>
- Kodali G, Mancini JA, Solomon LA, Episova TV, Roach N, Hobbs CJ, et al. Design and engineering of water-soluble light-harvesting protein maquettes. *Chem Sci.* 2016;8:316–24.
- Koebke KJ, Pecoraro VL. Development of de novo copper nitrite reductases: where we are and where we need to go. *ACS Catal.* 2018;8:8046–57.
- Koepke J, Hu X, Muenke C, Schulten K, Michel H. The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum rubrum*. *Structure.* 1996;4:581–97.
- Lau SYM, Taneja AK, Hodges RS. 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.* 1984;259:13253–61.
- Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Discovery through the computational microscope. *Structure.* 2009;17:1295–306. <https://doi.org/10.1016/j.str.2009.09.001>
- Lichtenstein BR, Farid TA, Kodali G, Solomon LA, Anderson JLR, Sheehan MM, et al. Engineering oxidoreductases: maquette proteins designed from scratch. *Biochem Soc Trans.* 2012;40:561–6.
- Lishchuk A, Kodali G, Mancini JA, Broadbent M, Darroch B, Mass OA, et al. A synthetic biological quantum optical system. *Nanoscale.* 2018;10:13064–73.
- Lockhart DJ, Boxer SG. Stark effect spectroscopy of *Rhodobacter sphaeroides* and *Rhodospseudomonas viridis* reaction centers. *Proc Natl Acad Sci.* 1988;85:107–11. <https://doi.org/10.1073/pnas.85.1.107>
- Lombardi A, Pirro F, Maglio O, Chino M, Degradó WF. De novo design of four-helix bundle metalloproteins: one scaffold, diverse reactivities. *Acc Chem Res.* 2019;52:1148–59.
- López-García P, Goktas M, Bergues-Pupo AE, Kokscha B, Varón Silva D, Blank KG. Structural determinants of coiled coil mechanics. *Phys Chem Chem Phys.* 2019;21:9145–9.
- Mann SI, Nayak A, Gassner GT, Therien MJ, Degradó WF. 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.* 2021;143:252–9.
- Monshouwer R, Abrahamsson M, van Mourik F, van Grondelle R. Superradiance and Exciton delocalization in bacterial photosynthetic light-harvesting systems. *J Phys Chem B.* 1997;101:7241–8. <https://doi.org/10.1021/jp963377t>
- Moser CC, Sheehan MM, Ennist NM, Kodali G, Bialas C, Englander MT, et al. De novo construction of redox active proteins. *Methods in enzymology.* Volume 580. United States: Academic Press Inc; 2016. p. 365–88.

- Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'Brien R, Sturtevant JM, et al. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* 1996;5:1584–93.
- Nagy G, Igaev M, Jones NC, Hoffmann SV, Grubmüller H. SESCA: predicting circular dichroism spectra from protein molecular structures. *J Chem Theory Comput.* 2019;15:5087–102. <https://doi.org/10.1021/acs.jctc.9b00203>
- Naudin EA, Albanese KI, Smith AJ, Mylemans B, Baker EG, Weiner OD, et al. From peptides to proteins: coiled-coil tetramers to single-chain 4-helix bundles. *Chem Sci.* 2022;13:11330–40.
- Pearce R, Zhang Y. Toward the solution of the protein structure prediction problem. *J. Biol. Chem.* 2021;297(1):100870.
- Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, et al. Molecular dynamics simulations of large macromolecular complexes. *Curr Opin Struct Biol.* 2015;31:64–74. <https://doi.org/10.1016/j.sbi.2015.03.007>
- Romero E, Diner BA, Nixon PJ, Coleman WJ, Dekker JP, van Grondelle R. Mixed exciton-charge-transfer states in photosystem II: stark spectroscopy on site-directed mutants. *Biophys J.* 2012;103:185–94.
- Romero E, Novoderezhkin VI, Van Grondelle R. Quantum design of photosynthesis for bio-inspired solar-energy conversion. *Nature.* 2017;543:355–65.
- Scholes GD, Fleming GR, Olaya-Castro A, van Grondelle R. Lessons from nature about solar light harvesting. *Nat Chem.* 2011;3:763–74.
- Scott AJ, Niitsu A, Kratochvil HT, Lang EJM, Sengel JT, Dawson WM, et al. Constructing ion channels from water-soluble α -helical barrels. *Nat Chem.* 2021;13:643–50.
- Selvan D, Prasad P, Farquhar ER, Shi Y, Crane S, Zhang Y, et al. Redesign of a copper storage protein into an artificial hydrogenase. *ACS Catal.* 2019;9:5847–59. <https://doi.org/10.1021/acscatal.9b00360>
- Strümpfer J, Schulten K. Light harvesting complex II B850 excitation dynamics. *J Chem Phys.* 2009;131:225101. <https://doi.org/10.1063/1.3271348>
- Sutherland GA, Grayson KJ, Adams NBP, Mermans DMJ, Jones AS, Robertson AJ, et al. 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.* 2018;293:6672–81.
- Sutherland GA, Polak D, Swainsbury DJK, Wang S, Spano FC, Auman DB, et al. A thermostable protein matrix for spectroscopic analysis of organic semiconductors. *J Am Chem Soc.* 2020;142:13898–907.
- Tsargorodska A, Cartron ML, Vasilev C, Kodali G, Mass OA, Baumberg JJ, et al. Strong coupling of localized surface plasmons to excitons in light-harvesting complexes. *Nano Lett.* 2016;16:6850–6.
- Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr Opin Chem Biol.* 2018;44:1–8. <https://doi.org/10.1016/j.cbpa.2018.05.003>
- Wahadoszamen M, Margalit I, Ara AM, van Grondelle R, Noy D. The role of charge-transfer states in energy transfer and dissipation within natural and artificial bacteriochlorophyll proteins. *Nat. Commun.* 2014;5:5287.
- Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci USA.* 2020;117:1496–503.
- Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature.* 2013;497:643–6.
- Zhou NE, Kay CM, Hodges RS. Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded α -helical coiled-coils. *J Biol Chem.* 1992;267:2664–70.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Curti M, Maffei V, Teixeira Alves Duarte LG, Shareef S, Hallado LX, Curutchet C, et al. Engineering excitonically coupled dimers in an artificial protein for light harvesting via computational modeling. *Protein Science.* 2023;32(3):e4579. <https://doi.org/10.1002/pro.4579>