Carbon Atoms Speaking Out: How the Geometric Sensitivity of ^{13}C Chemical Shifts Leads to Understanding the Colour Tuning of Phycocyanobilin in Cph1 and AnPixJ

Sascha Jähnigen [†] and Daniel Sebastiani *

Institut für Chemie, Naturwissenschaftliche Fakultät II, Martin-Luther-Universität Halle-Wittenberg, von-Danckelmann-Platz 4, 06120 Halle (Saale), Germany; sascha.jahnigen@ens.psl.eu

* Correspondence: daniel.sebastiani@chemie.uni-halle.de

† Current address: PASTEUR, Département de Chimie, École Normale Supérieure, 24 rue Lhomond, 75005 Paris, France.

check for updates

Abstract: We present a combined quantum mechanics/molecular mechanics (QM/MM) molecular dynamics–statistical approach for the interpretation of nuclear magnetic resonance (NMR) chemical shift patterns in phycocyanobilin (PCB). These were originally associated with colour tuning upon photoproduct formation in red/green-absorbing cyanobacteriochrome AnPixJg2 and red/far-red-absorbing phytochrome Cph1Δ2. We pursue an indirect approach without computation of the absorption frequencies since the molecular geometry of cofactor and protein are not accurately known. Instead, we resort to a heuristic determination of the conjugation length in PCB through the experimental NMR chemical shift patterns, supported by quantum chemical calculations. We have found a characteristic correlation pattern of ^{13}C chemical shifts to specific bond orders within the π-conjugated system, which rests on the relative position of carbon atoms with respect to electron-withdrawing groups and the polarisation of covalent bonds. We propose the inversion of this regioselective relationship using multivariate statistics and to apply it to the known experimental NMR chemical shifts in order to predict changes in the bond alternation pattern. Therefrom the extent of electronic conjugation, and eventually the change in absorption frequency, can be derived. In the process, the consultation of explicit mesomeric formulae plays an important role to qualitatively account for possible conjugation scenarios of the chromophore. While we are able to consistently associate the NMR chemical shifts with hypsochromic and bathochromic shifts in the P_g and P_{fr}, our approach represents an alternative method to increase the explanatory power of NMR spectroscopic data in proteins.

Keywords: protein NMR; theoretical spectroscopy; phycocyanobilin; π-conjugation; colour tuning; multivariate statistics; QM/MM; molecular dynamics

1. Introduction

Tetrapyrroles showcase an overwhelming abundance in living species, accompanied by an impressive functional versatility [1]. They can be classified into cyclic and open-chain forms, the former comprising the well-known porphyrins that build the precursors of hemes and chlorophylls, renowned for their important role in oxygen transport, oxidoreductase enzymes, and photosynthesis [2]. The second group, open-chain tetrapyrroles, also referred to as *bilins*, are ubiquitous in living species as well. Organised in biliproteins, they take the role of a chromophore

and serve phototrophic organisms as accessory pigments [3]. An important class of biliproteins is formed by phytochromes—photosensors that typically entail histidine kinase activity to regulate photomorphogenesis and photoperiodism [4–7]. The occurrence of phytochromes was long assumed to be restricted to higher plants, but over the years members of this class have also been found in cyanobacteria, bacteria, and fungi [6–11]. Recently, variants of (bacterio)phytochromes were found, genuine to cyanobacteria, that are therefore called cyanobacteriochromes (CBCRs) [12,13]. They fulfil a widespread range of tasks in the cell connected to the response towards light, such as positive phototaxis [12,14,15]. Although being phylogenetically related to phytochromes, CBCRs bear unique features: while the sensory module of phytochromes needs to encompass at least a period circadian/aryl hydrocarbon receptor nuclear translocator/single-minded (PAS)–cGMP phosphodiesterase/adenylyl cyclase/FhlA (GAF)–phytochrome specific (PHY) tridomain (Figure 1), CBCRs require only a chromophore-binding GAF domain to maintain their biochemical properties [16]. Most intriguing, however, is the ability of CBCRs to tune their absorption over the entire range of the visible spectrum, including near-IR and near-UV [12,17–20]. The treasure of photosensors provided by nature with which organisms adapt to their environment bears high utility for biochemical applications in bioimaging and optogenetics, but also in therapy and catalysis [14,21–34].
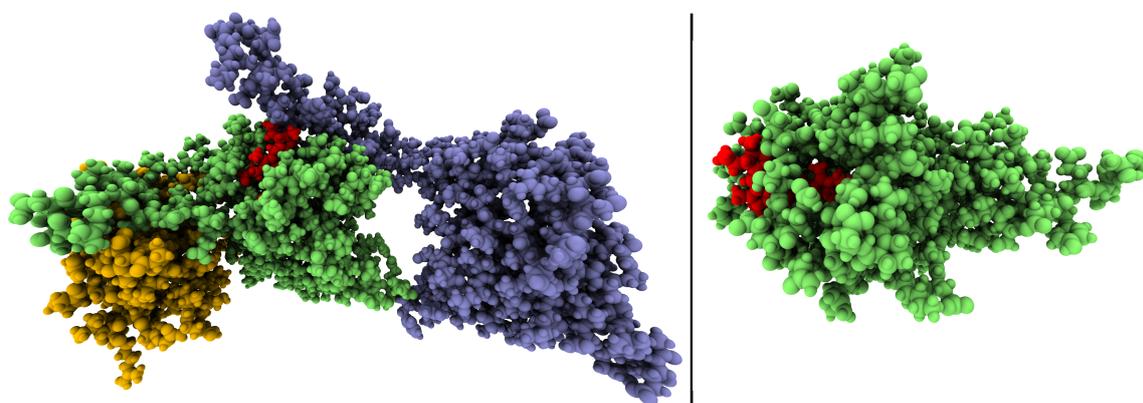

Figure 1. (**Left**) Sensory module of Cph1, a phytochrome of cyanobacterial origin. Phytochromes comprise a PAS (yellow)–GAF (green)–PHY (blue) tridomain, [16] wherein the latter acts as signal transmitter [10,35]. (**Right**) Photoactive GAF domain (green) of AnPixJ, a cyanobacteriochrome (CBCR) [13]. In both cases, the cofactor phycocyanobilin (PCB) is covalently bound to the GAF domain (red). Note: PAS—period circadian/aryl hydrocarbon receptor nuclear translocator/single-minded; GAF—cGMP phosphodiesterase/adenylyl cyclase/FhlA; PHY—phytochrome specific.

Many cyanobacterial phytochromes and CBCRs carry as bilin cofactor, phycocyanobilin (PCB), which operates as a photoreversible switch, toggling between the *ZZZssa* and *ZZEssa* configuration (Scheme 1) [4]. Exciting the chromophore at its absorption maximum, photochemical *Z*/*E* isomerisation of the C_{15}–C_{16} double bond initiates a flip of the D ring, which triggers a biochemical response [6,36–41]. The two states, *Z* and *E*, are denoted as dark state and photoproduct, respectively. The photoproduct is back-transformed into the dark state by photochemical *E*/*Z* isomerisation after excitation at its, typically shifted, absorption maximum. Depending on the colour (*x*) of absorbed light, the states are labelled as "P_x". Phytochromes switch between red-absorbing (P_r) dark states and far-red-absorbing (P_{fr}) photoproducts, [4,8] while the colour diversity in CBCRs is broader. Typical examples of CBCR absorption colour pairs are red/green, green/red, blue/green, red/orange, green/orange, violet/orange, blue/yellow, or blue/teal [12,14,18–20,34,42].

The colour tuning in CBCRs is exceptional, regarding the very limited number of chromophore species they bind. We recently showed that the degree of solvation due to a sealed chromophore binding pocket cannot serve as an explanation model [43]. However, Peng and co-workers experimentally related the dihedral angle of ring D of PCB in allophycocyanin B (a phycobilisome) with the absorption maxima ranging from 590 to 670 nm and thus proved the bathochromic shift's

originating in an increased conjugation length [44]. Recently, Wiebeler and co-workers showed the hypsochromic shift in the green-absorbing photoproduct of Slr1393g3 (a PCB-carrying CBCR) to emerge from a decrease of conjugation, originating from the out-of-plane rotation of ring D ("trapped-twist" model) [17,45]. They underlined the important role of the apoprotein to keep the chromophore in the intended conformational state.
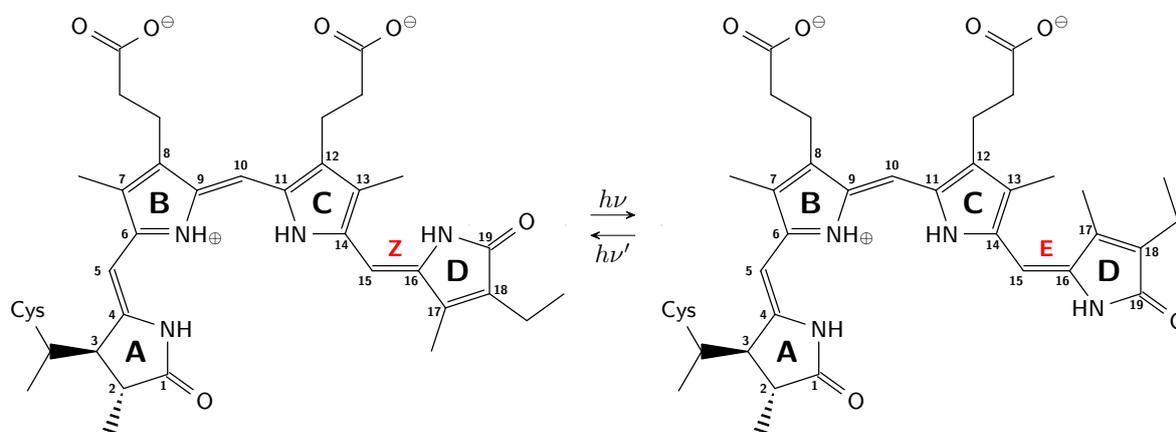
Scheme 1. Structural formula of phycocyanobilin (PCB) in *ZZZssa* (**left**) and *ZZEssa* (**right**) configuration. Reversible photoisomerisation of the C_{15}–C_{16} double bond triggers the photoswitch.

In this work, we present a quantum mechanics/molecular mechanics (QM/MM) study of PCB in a protein environment by means of extensive molecular dynamics (MD) and ab initio ^{13}C-nuclear magnetic resonance (NMR) chemical shift calculations, based on density functional theory (DFT). Comprising the entire protein together with the solvent, our theoretical model is able to account for physical meaningful non-covalent interactions without losing accuracy at describing the chromophore and its instantaneous thermal fluctuations. We focus on two cyanobacterial photosensor proteins (Figure 1): the red/far-red switching Cph1Δ2 (see "Abbreviations"), a well-studied member of the phytochrome superfamily [10,18,35,38–40,46–66] and the red/green switching AnPixJg2 (see "Abbreviations"), a cyanobacteriochrome that has been discovered later [12,13,43,45,67–73]. Both proteins have PCB as a cofactor and exhibit a red-absorbing dark state (P_r). Photoconversion yields a far-red absorbing photoproduct (P_{fr}) in the case of Cph1Δ2, whereas AnPixJg2 renders a green-absorbing state (P_g). Although the spectral shift in opposite directions amounts to a difference of more than 100 nm, photoproduct formation neither implies chemical changes to the chromophore nor a changed protonation state [45,70,72]. However, according to the studies by Peng and Wiebeler it can be concluded that the D-ring twist and thus the effective conjugation length determine the absorption properties [17,44]. To date, no crystal structures of the photoproduct states of Cph1Δ2 and AnPixJg2 exist that could build the basis of theoretical structure models, but we show that relying on the two P_r states, structural fluctuations can profitably be correlated with calculated spectroscopic patterns and increase the explanatory power of NMR spectroscopy towards conjugation effects in the P_g and P_{fr} state, respectively.

2. Results

2.1. How Bond Alternation and Thermal Fluctuations Determine ^{13}C-NMR Chemical Shifts

Extensive MD sampling under the QM/MM regime allows for the inclusion of thermal fluctuations of the molecular structure on the femtosecond and picosecond time scale. Figure 2a shows the computed time evolution of the C_{13}–C_{14} bond length and the ^{13}C chemical shift of atoms C_{13} and C_{14}—exemplary for all similar features in PCB—as they are found in the simulations. The thermal motion of the atoms at 300 K leads to bond length oscillations in the range 1.4–1.5Å, and to corresponding oscillations of the instantaneous carbon NMR chemical shifts within a range

of about 20 ppm. These oscillations are too fast for the NMR experiment; only their averages (dashed lines in Figure 2a, cf. also Table 1) can be compared to corresponding measurements. Figure 2b matches computed ^{13}C chemical shifts of PCB's carbon atom scaffold (C_1 through C_{19}) with reported MAS-NMR data from Matysik and co-workers for the P_r states of AnPixJg2 and Cph1Δ2 at 233 K [68]. The simulations clearly reproduce the characteristic alternating pattern of ^{13}C chemical shifts mirroring the π-conjugated system, which only proves that PCB in the model represents the physical state of the chromophore in the proteins. At the same time, though, the uncertainty in the computed averages impedes distinguishing between the two protein environments of the cofactor. In fact, with chemical shifts differing by less than 5 ppm, AnPixJg2 and Cph1Δ2 are hardly distinguishable experimentally let alone by the QM/MM model given the large fluctuations shown in Figure 2a.
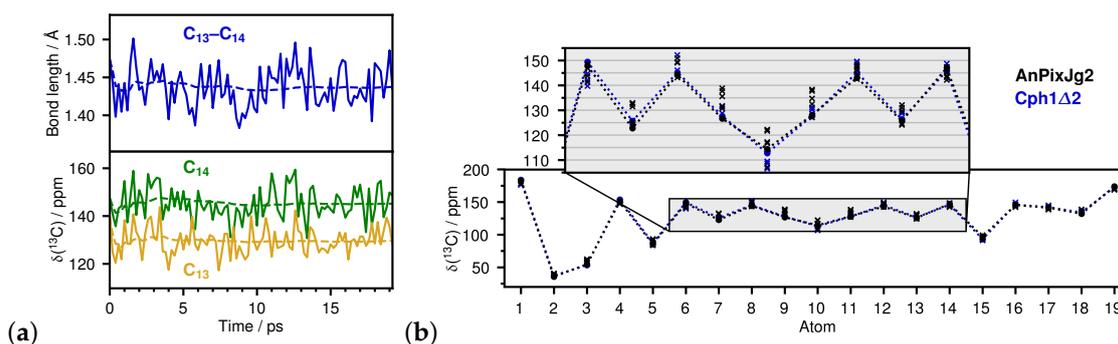

Figure 2. (**a**) Thermal fluctuations exemplified by bond length and ^{13}C chemical shifts of carbon atoms C_{13} and C_{14} in PCB along a quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulation of AnPixJg2 in its P_r form (solid line) and running average (dashed line). For atom numbering see Scheme 1. (**b**) ^{13}C chemical shifts of PCB in AnPixJg2 (black) and Cph1Δ2 (blue). Crosses mark the values computed from QM/MM MD trajectories while full circles indicate the experimental reference (dashed lines to guide the eye) [68].

Although it seems that the uncertainty in the obtained averages of ^{13}C chemical shifts suggests a high sensitivity of the carbon atoms' shielding towards any direct or indirect change in the environment, the further analysis of their dependence on chromophore–apoprotein connections and temperature effects enables a more refined notion of the matter. Computationally, it is straightforward to repeat chemical shift calculations for the same setup, gradually "switching-on" those protein residues that interact with the chromophore, or numerically "cooling-down" the system, which forces the atoms back to their equilibrium positions. Corresponding results can be found in Appendix A (Figure A1). These technical modifications shed light into the susceptibility of ^{13}C chemical shifts in PCB—and into those of other nuclei. It turns out that they are largely insensitive to temperature and non-covalent interactions the chromophore is subjected to due to their mainly quartary nature. This contrasts with the high sensitivity of ^{1}H and ^{15}N nuclei, the chemical shifts of which change strongly with both temperature and supramolecular interactions.

However, Figure 2b undoubtedly indicates that the conjugation pattern and charge distribution in the chromophore strongly determine the extent of ^{13}C chemical shifts. This striking correspondence points to the fact that instantaneous fluctuations like those shown in Figure 2a actually incorporate an intimate connection of bond parameters and nuclear shieldings. How this manifests in PCB will be shown and discussed in the following section; we so far conclude that the chromophore's carbon atoms C_1–C_{19}, though being blind towards supramolecular effects, are an important probe of their immediate covalent situation.

Table 1. Results of the inversion of the geometric sensitivity of ^{13}C chemical shifts shown in Figure 3 based on extensive QM/MM MD simulations and reported NMR measurements by Song et al. [68] for AnPixJg2 (switches from P_r to P_g) and Cph1Δ2 (switches from P_r to P_{fr}). $\Delta\delta^C(P_x–P_y)$ is the experimental difference pattern of ^{13}C chemical shifts going from P_x to P_x; $d(P_r)$ denotes averaged bond lengths calculated from the MD trajectories of both proteins in their respective P_r state; $\Delta d(P_x–P_y)$ is the change in bond length (together with uncertainty) going from P_y to P_x, predicted from principal component analysis and multivariate regression.

| | C$_{13}$–C$_{14}$ | C$_{14}$–C$_{15}$ | C$_{15}$–C$_{16}$ | C$_{16}$–C$_{17}$ | C$_{17}$–C$_{18}$ | C$_{18}$–C$_{19}$ |
	C$_{13}$	C$_{14}$	C$_{15}$	C$_{16}$	C$_{17}$	C$_{18}$
$\Delta\delta^C(P_r–P_r)^{\dagger,*}$	−0.7	+1.2	+3.8	−0.5	−0.9	−2.0
$\Delta\delta^C(P_g–P_r)$ *	−6.1	−7.4	+2.3	−8.4	+4.7	+9.3
$\Delta\delta^C(P_{fr}–P_r)$ *	+4.3	+6.1	−1.6	+5.7	−6.6	+6.5
$\Delta\delta^C(P_g–P_{fr})$ *	−11.1	−12.3	+7.7	−14.6	+10.4	+0.8
$d(P_r)$ **	143	144	137	148	137	149
$\Delta d(P_r–P_r)^{\dagger,**}$	−1.3 ± 0.5	+4.1 ± 0.6	+2.1 ± 0.5	−3.8 ± 0.6	−3.1 ± 0.4	−1.6 ± 0.6
$\Delta d(P_g–P_r)$ **	−16.6 ± 2.0	+4.8 ± 2.3	−14.4 ± 1.7	−4.5 ± 2.3	+19.1 ± 1.3	−6.2 ± 2.1
$\Delta d(P_{fr}–P_r)$ **	+12.8 ± 1.6	−3.2 ± 1.9	+7.8 ± 1.3	−4.6 ± 1.8	−4.1 ± 1.0	+16.0 ± 1.7
$\Delta d(P_g–P_{fr})$ **	−30.7 ± 3.0	+12.0 ± 3.6	−20.1 ± 2.6	−3.7 ± 3.5	+20.1 ± 2.0	−11.4 ± 3.3

† $P_r^{AnPixJg2}$ − $P_r^{Cph1\Delta2}$, * in ppm; from reference [68], ** in pm.

2.2. Correlations of ^{13}C Chemical Shifts and Their Geometric Sensitivity

The interconnection of nuclear shieldings and bond lengths is not particularly surprising, but the remarkable "isolation" of ^{13}C chemical shifts from supramolecular interactions deserves a closer look. Figure 3 presents the linear correlation matrix of ^{13}C shifts with C-C bond lengths in rings C and D of PCB in P_r state. It is based on plotting the instantaneous values obtained from the QM/MM trajectories against each other and extracting statistically significant connections through linear regression. It therefore reveals whether chemical shifts decrease or increase upon increasing the individual bond lengths (see Figure A2 for the explicit correlation diagram). It is important to note that we are entitled to carry out a combined analysis of both proteins alike, given the P_r state of both proteins, AnPixJg2 and Cph1Δ2, having PCB in the same or a very similar configuration (the matrix shown in Figure 3 looks essentially the same for either protein). It is the central result of this report, because it identifies not only pivotal correlations, but furthermore reveals a particular pattern, which in the following we refer to as the geometric sensitivity of ^{13}C chemical shifts. There are numerous representations that describe a sensitivity, yet in this study we are interested in those chromophore parts that are involved in colour tuning, that is, the nuclei describing the conjugation between rings C and D (C$_{13}$–C$_{19}$).

Qualitatively, the geometric sensitivity follows the clear rationale that an atom's chemical shift increases with the length of the originating bonds—and decreases with the next, and increases again with the following bond and so on. In Figure 3 this alternation is marked by red and blue bullets, respectively, shaping an imagined double diagonal; like Figure 2b, it is another representation of the underlying conjugated system. The symmetry of the geometric sensitivity, however, is clearly disturbed leaving certain nuclei with a "preferred" bond they are correlated with. For instance, there is a strong connection of atoms C$_{13}$ and C$_{14}$ to bond C$_{13}$–C$_{14}$, but a rather weak one to bonds C$_{12}$–C$_{13}$ and C$_{14}$–C$_{15}$, respectively. The same observation can be made for C$_{16}$ and C$_{17}$, while the latter shows a strong connection to bond C$_{18}$–C$_{19}$, whereof it is not even a member. Expectedly, the two pyrrole systems mainly exhibit within-ring correlations, but atoms C$_{15}$ and C$_{16}$ are connected to both parts. Amide C$_{19}$, in turn, appears completely independent from the conjugated system.

The characteristic sensitivity in Figure 3 evidently is determined not only by the bond alternation of the conjugated system, but also by bond polarisation due to the relative position of the nuclei to electron-withdrawing groups and substituents. Nevertheless, the direct interpretation of the geometric pattern is difficult and we suggest employing tools from multivariate statistics. As will be shown in the following section, we can use the geometric sensitivity together with experimental evidence to

predict the conjugation length in P_g and P_{fr} states of PCB alike, assuming that the correlation matrix in Figure 3 remains unaltered in either state.
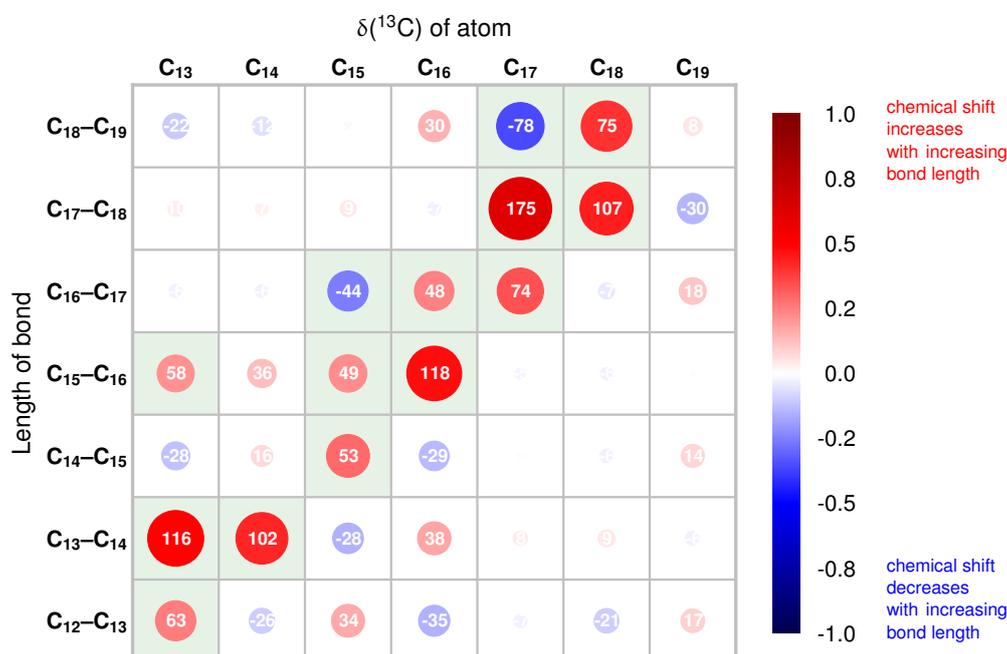

Figure 3. Correlation matrix showcasing the geometric sensitivity of ^{13}C chemical shifts towards changes in selected C-C bond lengths in rings C and D of PCB in AnPixJg2 and Cph1Δ2 (P_r state). Colour and size of the circles correspond to the Pearson correlation coefficient (see colour bar on the right), whereas white numbers denote the slope of linear regression (in ppm Å^{-1}). Mint cells are significant at $p < 0.0001$; $n = 405$ snapshots.

2.3. Inversion of the Geometric Sensitivity Points to P_g and P_{fr} States

In 2015, Song et al. reported a peculiar difference pattern of ^{13}C nuclear shieldings ($\Delta\delta^C$) in PCB upon photoconversion in AnPixJg2 and Cph1Δ2: Nuclei belonging to rings C and D exhibit changes in chemical shifts of up to 10 ppm when transforming from the P_r state into the respective photoproduct [68]. The authors attribute this observation to a "large-scale rearrangement of the critical interactions [that] cannot be explained only by the configurational changes of the chromophore." However, recalling that ^{13}C chemical shifts are blind towards supramolecular perturbations (cf. Figure A1) the changes cannot be explained by the removal of hydrogen bonds or the emersion of a new charged side chain—unless this involves changing the covalent situation of the carbon atoms. Most intriguing, however, is the observation by Song of an opposite trend of the $\Delta\delta^C$ pattern for most atoms upon switching to either P_{fr} or P_g state with P_r lying in the middle. Following the correspondence principle, this trend must be found in the causing process (i.e., photoconversion) as well, which is supported by what is by now commonly understood as the colour tuning mechanism: [17,44] the extent of conjugation in rings B, C and D of the chromophore, where the P_g and P_{fr} states are characterised by a decreasing and increasing conjugation length relative to the P_r state, respectively.

The NMR experimental observations are made amid scarce information on the atomic coordinates of the photoproduct states since no structural data from X-ray diffraction are available. In these terms, the geometric sensitivity may serve to retrieve structural data from the ^{13}C-NMR experiment and to predict the single–double bond conjugation pattern for the P_g and the P_{fr} state. To endow our theoretical observations with quantitative information, we inverted the correlation matrix in Figure 3 and applied the result on the $\Delta\delta^C$ patterns reported by Song et al. [68] We carried out a principal component analysis (PCA) to account for the importance of fluctuations in the feature variables before

using multivariate regression to obtain predicted Δd patterns that correspond to the experimental $\Delta\delta^C$ patterns, as shown in Table 1 (see Appendix A for computational details). We report as a reference the average bond lengths of PCB in the P_r state, calculated from the MD trajectories. They support the inferred change in bond order, for instance to what extent a single or double bond character is retained. The given uncertainties of prediction help to distinguish important from less significant trends upon photoconversion.

Comparing the two P_r states of AnPixJg2 and Cph1Δ2, we do not find big differences between the two proteins, which corresponds to the earlier argument that PCB in both proteins is chemically equivalent. The found numbers are mainly due to the comparably large difference in chemical shift for atom C_{15}, lying at the joint of rings C and D and which is connected to many bonds (cf. Figure 3). Going over to predicted changes in bond length upon photoconversion, that is, differences going from P_r to P_g and P_{fr}, respectively, we are able to distinguish key sites among other, less significant changes. Bond C_{13}–C_{14}, for which we hardly see a difference in the P_r state, undergoes important changes and exhibits a clearly opposite behaviour in the process: while in P_g state it is shortened by about 17 pm, its length increases by about 13 pm in P_{fr}. Bond C_{14}–C_{15} is already longer in P_r of AnPixJg2 and becomes even longer in P_g, but shorter in P_{fr} of Cph1Δ2, amounting to a net difference of 12 pm between the photoproducts. Bond C_{15}–C_{16} again shows opposite behaviour being shorter in the P_g state (–14 pm) and longer in the P_{fr} state (+8 pm). For the bonds forming ring D the prediction tends to assign changes to either one or the other photoproduct, but changes in bond C_{16}–C_{17} are small and not significant. Bond C_{17}–C_{18} is clearly lengthened in P_{fr} (+19 pm) as is bond C_{18}–C_{19} in P_g (+16 pm) with respect to P_r. Consequently, this analysis reveals that it is possible turning the asymmetry of the geometric sensitivity of ^{13}C chemical shifts to precisely account for subtle changes in the covalent structure of PCB, thereby allowing regioselective predictions. In the following discussion we make the attempt to interpret the pattern on a molecular scale.

3. Discussion

The presented results of the inversion of the geometric sensitivity of ^{13}C chemical shifts are based on experimental evidence reported by Song et al. [68] We inferred expected changes in bond length of the PCB chromophore when going from the dark state P_r to the photostates P_g for AnPixJg2 and P_{fr} for Cph1Δ2. We pointed out—in accordance with the common understanding of the colour tuning mechanism in these forms—the antagonistic nature of the changes involved. The results in Table 1 can be used to predict the predominant character of the investigated C-C bonds of PCB in P_g and P_{fr} states. Associating bond lengths with bond orders, we propose the following alternation pattern:

Table 2. Predicted C-C bond character trends in the conjugated system of rings C and D of PCB in P_g and P_{fr} state based on Table 1 and a threshold of 144 pm. There is no reliable prediction possible for bond C_{16}–C_{17}.

	C_{13}–C_{14}	C_{14}–C_{15}	C_{15}–C_{16}	C_{16}–C_{17}	C_{17}–C_{18}	C_{18}–C_{19}
P_g	double	single	double	(single)	single	double
P_{fr}	single	double	single	(single)	double	single

Table 2 suggests that the bond alternation pattern in the two photostates is opposite and we present in the following an explanatory model to translate this into the effective conjugation length. One has to keep in mind that the alternation pattern represents the predominant character of the C-C bonds, but does not compare to pure single or double bonds since pyrrole rings form a π-conjugated system not only between, but also within themselves.

Considering that the shaping of delocalisation of double bonds by means of mesomeric schemes has so far not received due attention, we argue that working out what is actually possible and favourable in terms of bond alternation is a key to fully understand the colour tuning phenomenon in PCB.

3.1. P_g: Short Conjugation

Schemes 2 and 3 illustrate the π-resonance of PCB in P_g state where conjugation has been found to be of short length. We consider the conjugation between rings B and C, which is maintained in all states of the chromophore, as *core conjugation*. Assuming that in P_g state ring D is twisted out of plane, core conjugation alone carries the absorption of green light, which is in line with what has been reported by Wiebeler et al. [17] There exist several mesomeric formulae describing this conjugation, the most important of which are shown in Scheme 2. Therein, the positive charge is delocalised along with the change of bond alternation, but the high electronegativity of the nitrogen atoms shifts the distribution towards the formation of stabilised tertiary carbocations centered at C_8 and C_{12}, thereby rendering bond C_{13}–C_{14} with an increased bond order. A stabilisation of the positive charge to some extent at the pyrrole nitrogen atoms seems reasonable since both proteins present residues at this location that may serve as hydrogen bond acceptors, but have to compete with water molecules—also in AnPixJg2 [43]. In any case this would not swivel the predicted bond character at C_{13}–C_{14}.
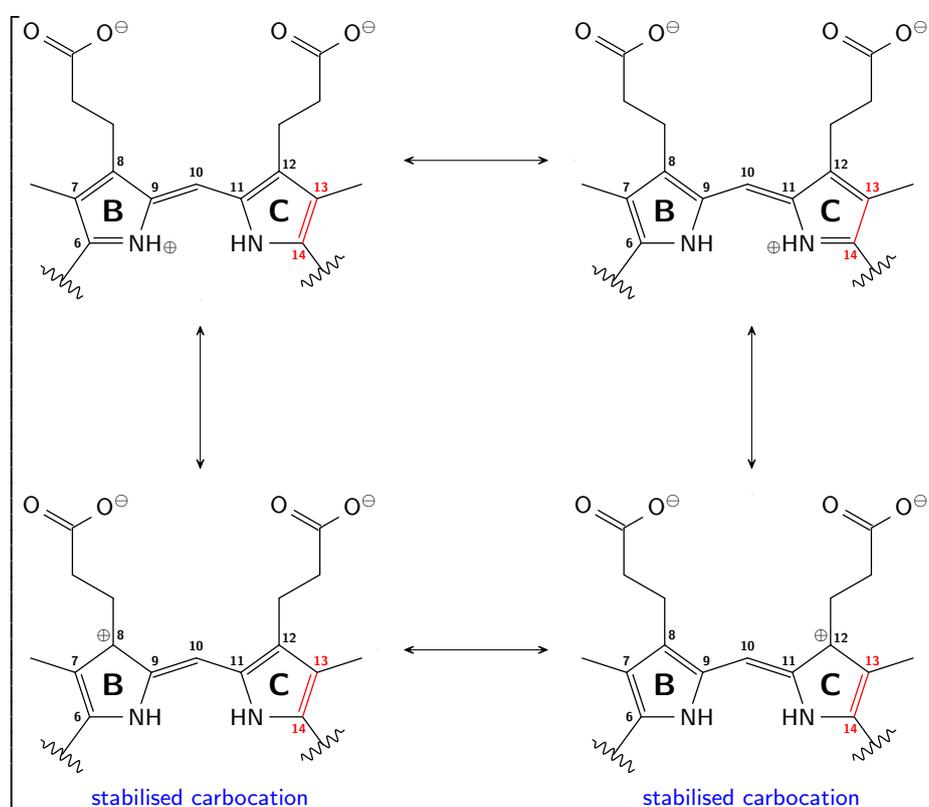
Scheme 2. Mesomeric formulae showcasing the delocalisation of positive charge in PCB between rings B and C. It represents the core conjugation, which is prevalent in the P_g state of the chromophore. The investigated bond C_{13}–C_{14}, marked in red, inherits a double bond character due to the importance of mesomeric forms with a (stabilised) tertiary carbocation.
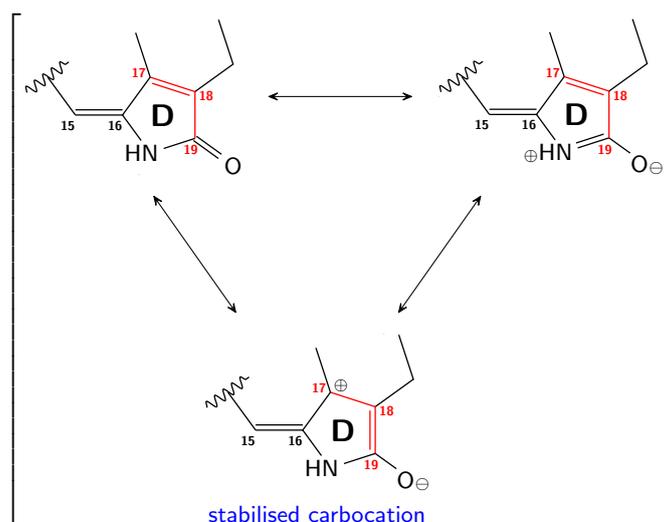
Scheme 3. Negative mesomeric effect ($-M$) exerted by the carbonyl group in ring D of PCB. It corresponds with amide resonance, which is prevalent in the P_g state of the chromophore. The formation of a (stabilised) tertiary carbocation renders bonds C_{17}–C_{18} and C_{18}–C_{19} with a single–double alternation.

Simultaneously, ring D, being fully detached from the conjugated core, is subjected to *amide resonance* pulling electron density from the pyrrole ring and generating a negative partial charge at the carbonyl oxygen atom. However, in pyrrole, resonance does not only include the NH group of the amide bond, but the entire aromatic system. As a consequence another mesomeric form can be found, which is again stabilised through formation of a tertiary carbocation far from the electronegative centres. This has the important consequence that bond C_{17}–C_{18} no longer retains a double bond character, which is shifted to C_{18}–C_{19}, but that of a single bond with atom C_{17} carrying the positive (partial) charge.

We want to note that the methylene bridges are not favourable sites to put a positive partial charge, due to their secondary type, but they still take part in conjugation. Furthermore, it has been argued that a certain single bond character of C_{15}–C_{16} is the underlying cause of "dark conversion" [6].

We conclude that a plausible prediction of the geometric bond length conjugation pattern of the PCB in its P_g state is possible through measured and computed NMR chemical shifts (and their geometric sensitivities). The resulting conjugation pattern, in turn, provides a consistent explanation of the change in absorption wavelength after photoconversion, which is an explanation for the unusual colour change of the bilin chromophore in this protein.

3.2. P_{fr}: Long Conjugation

π-Resonance of PCB in P_{fr} state follows different principles than in the case for the P_g state. As was pointed out by Peng et al., the increase in absorption wavelength is due to the extension of conjugation, which now includes rings B, C, and D since the latter rotates in the plane with the core [44]. We shall call this *full conjugation* because it resembles the longest possible resonance in PCB. Including ring D in the exchange of π-electrons has not only bonds C_{14}–C_{15} and C_{15}–C_{16} toggle their single–double bond pattern, but also withdraws charge density from the ring as the positive charge is now shared as well.

Scheme 4 presents the most important mesomeric formulae in the P_{fr} state and an immediate question might be why there is no favourable carbocation formation in ring D as has been adduced for rings B and C. The reason behind this is that in the case of P_{fr}, an explicit interaction with the apoprotein makes a difference. We have argued before that the carbon atoms do not directly "see" supramolecular effects; however, they are sensitive to indirect effects, which in the present case is the stabilisation of the positive charge by the apoprotein: Rohmer et al. argued that in P_{fr} the negatively charged Asp207 is in position to form a salt bridge with the pyrrole nitrogen atom carrying the positive charge [38]. Based on our analysis, we strongly suggest that this interaction is the dominating factor

in the π-resonance of the P_{fr} state. The new stable mesomeric form that can be found in the P_{fr} state corresponds with the predicted bond alternation pattern in Table 2.
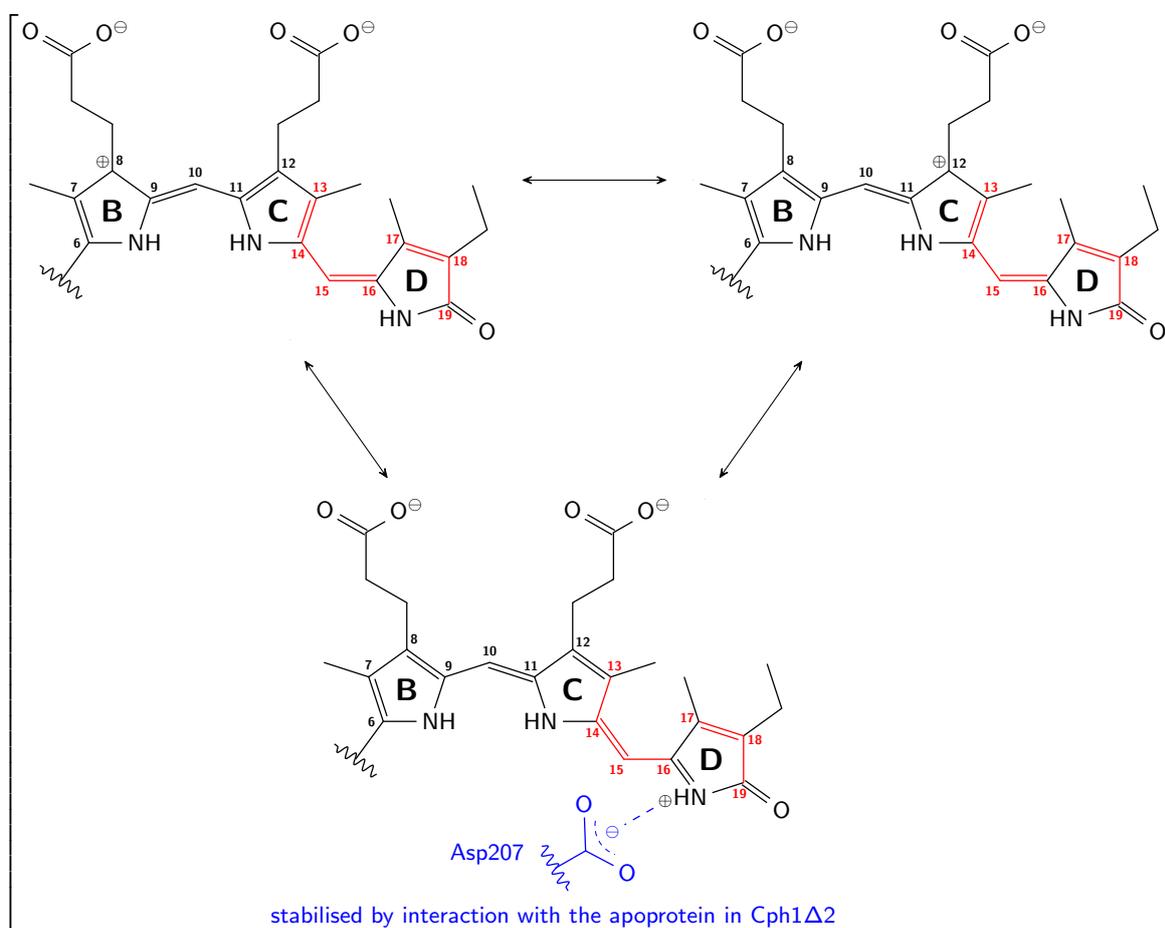
Scheme 4. Mesomeric formulae showcasing the delocalisation of positive charge in PCB between rings B, C, and D. It represents the full conjugation, which is prevalent in the P_{fr} state of the chromophore. Ring D carrying the positive charge is stabilised by a salt bridge to an aspartate residue (Asp207) of the apoprotein, thereby endowing this mesomeric form with high importance [37,38]; bonds C_{13}–C_{14}, C_{14}–C_{15}, and C_{15}–C_{16} are rendered with single–double–single alternation. The presence of positive charge in ring D also suppresses amide conjugation (cf. Scheme 3), which has bonds C_{17}–C_{18} and C_{18}–C_{19} retain their double–single alternation.

In summary, also for PCB in its P_{fr} state a plausible prediction of the bond length conjugation pattern is obtained, based solely on measured and computed NMR chemical shifts (and their geometric sensitivities). Again, the resulting conjugation pattern serves as explanation of the change in absorption wavelength through photoconversion.

4. Conclusions

We have addressed the colour tuning mechanism in the photosensor proteins AnPixJg2 and Cph1Δ2, which are based on the phycocyanobilin (PCB) chromophore, via a quantum chemical method using an indirect approach. While the direct approach (i.e., the calculation of absorption energies in the different configurations) is possible, but difficult due to the absence of reliable structural data, we have chosen to use reported experimental ^{13}C-NMR chemical shifts of the chromophore backbone. These data are available for both P_{fr} and P_{g} state, together with the chemical shift changes with respect to the dark state conformation, P_{r}, of known structure. We infer structural information and derive the respective bond length alternation pattern, which points towards the effective conjugation in

the unknown P_{fr} and P_g states. This can eventually be linked to the changes in absorption wavelength via simple confinement considerations.

Our complementary approach represents an uncommon computational route that is interesting for situations in which clear and reliable structural information (e.g., from X-ray diffraction) is difficult to obtain due to experimental constraints (here, the lifetime of the photoproducts P_{fr} and P_g). The approach requires the quantitative determination and subsequent inversion of a particular structure–property relationship; namely, the dependence of the NMR chemical shifts on the bond length pattern. This can be realised in the form of a matrix representing the transformation of the bond length vector $d(C_i, C_{i+1})$ into the NMR chemical shift vector $\delta_i{}^C$ at linear order. For the special case of AnPixJg2 and Cph1Δ2, we used extensive QM/MM MD simulations and ab initio computations of ^{13}C chemical shifts to determine the geometric sensitivity of the latter. We could show that with the help of multivariate regression this approach leads to a semi-quantitative understanding of the colour tuning mechanism, without the need to have the explicit molecular conformations of the chromophore.

A series of molecular conjugation scenarios using mesomeric formulae that explicitly describe the possibilities of bond alternation in PCB helps to connect the statistical predictions with the actual conjugation patterns found in P_g and P_{fr} state. An important role is played by the stabilisation of the positive charge in tertiary carbocations or by ionic interaction with the protein.

Showcasing the examples of cyanobacteriochrome AnPixJg2 and phytochrome Cph1Δ2, our approach represents an alternative method to interpret NMR data that can readily be transferred to related tetrapyrrole chromophores in various protein environments.

5. Materials and Methods

5.1. Preparations

Starting geometries for P_r states of AnPixJg2 and Cph1Δ2 were obtained as X-ray structures from the Protein Data Bank (PDB; AnPixJg2: 3W2Z, [13] Cph1Δ2: 2VEA [35]). The simulation cell was set up using VMD, [74] solvating the protein in a box of TIP3P water [75]. Possible protonation states of protein residues were set to correspond to pH 7.0. The protonation state of histidine was chosen as follows: AnPixJg2: His293-E, His322-P, rest: D; Cph1Δ2: His260-E, His290-D, rest: D. The PCB chromophore was chosen with fully protonated pyrrole rings according to the experimental evidence [68].

5.2. Force Field Molecular Dynamics Simulations

FFMD simulations were carried out with the NAMD package [76] under periodic boundary conditions (PBC) using the CHARMM22 force field [77,78] in the isothermal-isobaric (NPT) ensemble and the combined Nosé–Hoover Langevin piston method with a period of 200 fs [79,80]. For PCB, force field parameters obtained by Mroginski and co-workers were used [58]. For van der Waals interactions a cutoff of 10 Å was employed. The initial simulation cell dimensions were chosen to include a water layer of 30 Å thickness. Bond lengths between heavy atoms and hydrogen atoms were held fixed using the SHAKE algorithm [81] with a time step of 2 fs. The system was equilibrated, first, by optimisation of the water cell, followed by heat up and equilibration keeping the protein positions fixed. Then, the protein structure was optimised with fixed water positions. Final equilibration for 200 ps was carried out after heating up the entire simulation cell to the desired temperature within 200 ps, followed by the production run (see Table 3 for more details).

Table 3. Computational details of force field and ab initio molecular dynamics calculations for the protein study.

Parameter	FFMD	AIMD	NMR
Ensemble	NPT	NVT	NVT
No. of atoms	37,651/104,960 *	$\approx 100_{QM}$	$\approx 280_{QM}$
Temperature	300 K	300 K	–
Pressure	1.01325 bar	–	–
Barostat	Langevin	–	–
Thermostat	Nosé–Hoover	Nosé–Hoover	–
ABC_{QM} (Å)	– **	30.0, 30.0, 30.0	35.0, 35.0, 35.0
Force Field/Functional	CHARMM22	BLYP-D3	BLYP-D3
Basis Set	–	TZVP-GTH	pcS2/3
Density Cutoff	–	320 Ry	400 Ry
No. of Snapshots	–	–	405
t_{NVP}/t_{NVT}	≈ 1 μs	>20 ps	–
Other	SHAKE [81]		

* AnPixJg2/Cph1Δ2 + solvent box (water); ** solvation layer thickness: 30 Å

5.3. QM/MM Molecular Dynamics Simulations

Snapshots from FFMD simulations were transferred into a QM/MM setup. AIMD simulations were performed as the Born-Oppenheimer type in the NVT ensemble using CP2K 5.1 [82]. The partitioning of the protein into QM and MM part, as well as employed capping atoms, is listed in Table 4. QM/MM bond interfaces were handled using an optimised capping potential introduced between for the bond between α and β carbon atoms [83,84]. The MM part was described by the CHARMM22 force field (see previous section). The size of the QM box was set to 30.0, 30.0, 30.0 Å. The BLYP functional [85,86] was employed together with Grimme's dispersion correction (D3), ref. [87] using the GPW scheme [88], GTH pseudo-potentials, [89–91] a density cutoff of 320 Ry, and the TZVP-GTH basis set. The QM/MM interaction term was evaluated within the GEEP scheme presented by Laino and co-workers [92,93]. Equilibration runs of length 5–10 ps were carried out under massive thermostatting using Nosé–Hoover chains of length 5 at 330 K and a coupling constant of 10 fs [80,94]. For subsequent production runs, QM and MM subsystems were coupled to separate thermostat units with a coupling constant of 500 fs.

Table 4. Information on protein residues included into the QM part, and corresponding capping atoms for response calculations.

	AnPixJg2				Cph1Δ2		
Residue	AIMD	NMR	Capping Atom	Residue	AIMD	NMR	Capping Atom
PCB	x	x	–	PCB	x	x	–
Trp289	x	x	CA	Asp207	x	x	CB/CA **
Asp291	x	x	CA	Ile208	x	x	C/C,CB **
His293	–	x	CA	Arg222	–	x	CB
Arg301	–	x	CA	Arg254	–	x	CB
Tyr302	–	x	CA	Ala256	–	x	CA
Phe308	–	x	CA	Tyr257	–	x	–
His318	–	x	CA	His258	x	x	CA/CB **
Phe319	–	x	CA	Cys259	x	x	–
Ser320	x	x	CA	His260	x	x	C
Cys321	x	x	–	Tyr263	–	x	CA
His322	x	x	C	His290	–	x	CA
Tyr334	–	x	CA	Water*	x	x	–
Water *	–	x	–				

* Water molecules within a distance of 2.6 Å to PCB. ** AIMD /NMR.

5.4. Calculation of Nuclear Shieldings

NMR calculations were performed in the same setup as AIMD simulations, but the size of the QM/MM part was chosen adaptively such that RMSE of chemical shifts were below 1% (cf. Figure A1).

Water molecules within a range of 2.6 Å from the chromophore were included in the QM part as well. Snapshots were extracted from the AIMD trajectory every 200 fs. All atoms in the QM region were treated at an all-electron level using the Gaussian-augmented plane waves (GAPW) method of CP2K 5.1 [95] with density cutoff of 400 Ry and pcS-3 basis set (pcS-2 on oxygen and sulphur atoms) [96]. The size of the QM box was set to 35.0, 35.0, 35.0 Å. Gauge origin was treated using the IGAIM method of CP2K, [95] which is an adaption of the CSGT method based on Atoms in Molecules [97].

5.5. Statistical Analysis

Principal component analysis (PCA) and multivariate regression was carried out with R using the packages tidyverse, ade4, and GGally [98]. For PCA the centred, unscaled ^{13}C chemical shifts of PCB atoms C_{13}–C_{18} were used as variables to account for the importance of fluctuations in these features. PCA was not used for dimensionality reduction; though, consequently, the number of components equalled the number of input variables (6; cf. Figure 4).
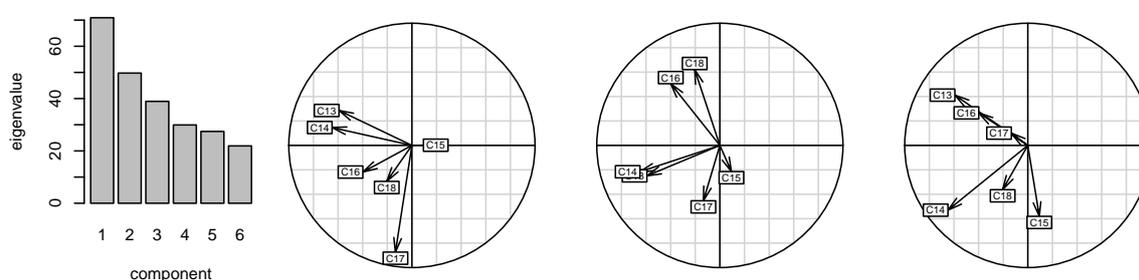

Figure 4. Principal component analysis (PCA) of ^{13}C chemical shift data obtained from the QM/MM MD simulations. From left to right: eigenvalues, components with PC 1 as abscissa and PC 2, 3, 4 as ordinate, respectively. PC 1–4 contain a cumulative projected inertia of 79%.

Multivariate regression and prediction was carried out, inverting the correlation matrix of ^{13}C shifts with C-C bond lengths in rings C and D, shown in Figure 3, based on the relationship

$$\Delta\delta_i^{\mathrm{C}} \propto M_{ij} \cdot \Delta d_j \,, \tag{1}$$

$\Delta\delta_i^{\mathrm{C}}$ being the change in ^{13}C chemical shift (features), Δd_j the change in bond length (outcome/prediction), and M_{ij} denoting elements of the linear correlation matrix (Figure 3); using the previously obtained principal components.

6. Other

Python-based MDAnalysis 1.0.0 [99,100] was used for handling protein topologies. Geometric analysis, QM/MM setups, as well as general post-processing of data was carried out using our own implementations in Python 3.8 [101–104] and numerical libraries NumPy 1.19.1 and SciPy 1.5.2 [105,106]. Plots were generated with Matplotlib 3.3.0 [107]. Molecular visualisations were created with VMD [74] using the Tachyon Ray Tracer [108] All structural formulae were created with LaTeX using chemfig 1.56 [109].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIMD	Ab initio molecular dynamics
AnPixJg2	Second GAF domain (out of four) of a phototaxis regulator found in heterocyst-forming cyanobacteria *Anabaena* (Nostoc) sp. PCC 7120 (all1069) [13,67]. *Pix*, originally *pis* (PhototaxIS), is a class of regulatory genes required for positive phototactic movement (PositIve phototaXis), whereas *J* is a serial letter [15]
BLYP	Becke/Lee-Yang-Parr (exchange/correlation)
CBCR	Cyanobacteriochrome
CHARMM	Chemistry at Harvard macromolecular mechanics
Cph1Δ2	"Cyanobacterial phytochrome 1", i.e., the first phytochrome found in cyanobacteria, more precisely *Synechocystis* sp. PCC 6803 (slr0473) [10,46,47]. It functions as light-regulated histidine protein kinase [10,35]. Δ2 denotes the sensory module of the protein comprising 514 N-terminal residues; *Δ1* comprises 492 N-terminal residues [48]
CSGT	Continuous set of gauge transformations
DFT	Density functional theory
FFMD	Force field molecular dynamics
GAF	cGMP phosphodiesterase/adenylyl cyclase/FhlA
GAPW	Gaussian-augmented plane waves
GEEP	Gaussian expansion of the electrostatic potential
GPW	Gaussian and plane waves
GTH	Goedecker-Teter-Hutter (pseudopotential)
IGAIM	Individual Gauges for Atoms in Molecules
IR	Infrared
MAS	Magic angle spinning
MD	Molecular dynamics
MM	Molecular mechanics
NMR	Nuclear magnetic resonance
NVT	Canonical (ensemble)
PAS	period circadian/aryl hydrocarbon receptor nuclear translocator/single-minded
PBC	Periodic boundary conditions
PCA	Principal component analysis
PCB	Phycocyanobilin
PHY	Phytochrome specific
PP	Pseudopotential
QM	Quantum mechanics
QM/MM	(Hybrid) Quantum mechanics/molecular mechanics
QMP	QM part
RMSE	Root mean square error
UV	Ultraviolett

Appendix A

Appendix A.1. Size of the QM Part and Convergence

Figure A1 shows the relative (normed) root mean square error (RMSE) of calculated NMR shieldings of chromophore atoms. In the left graph, the RMSE plotted against QM part size points to specific residues of importance. Naturally, all (charged) residues involved in hydrogen binding to the chromophore need to be considered in the QM part, including water molecules. Apart from this, no meaningful influence on NMR shieldings can be found. Notably, ^{13}C shieldings are barely affected by the environment, which is certainly related to their mainly tertiary nature. The effect of annealing has already been mentioned. The right graph in Figure A1 underlines the importance of

underlying molecular structures determined by either the level of theory (MM/QM) or the simulated temperature; the relative error of calculated shieldings drops from over 15% to less than 2% after direct annealing (i.e., without further equilibration) of the FFMD snapshot. Interestingly, thermal fluctuations let the error grow again above 5%. The atom types fluctuate differently with ^{1}H and ^{15}N being most susceptible. This corresponds to their chemical exposure to the protein environment.
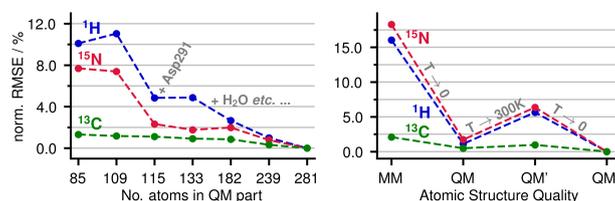

Figure A1. Normalised RMSE of calculated shieldings with increasing size of the QM part (**left**) and after subsequent annealing and equilibration of FFMD snapshots (**right**) at the example of AnPixJg2.

Appendix A.2. Statistical Data

Figure A2 shows the explicit correlation plot of all features and outcomes of this analysis. It corresponds to the correlation matrix shown in Figure 3.
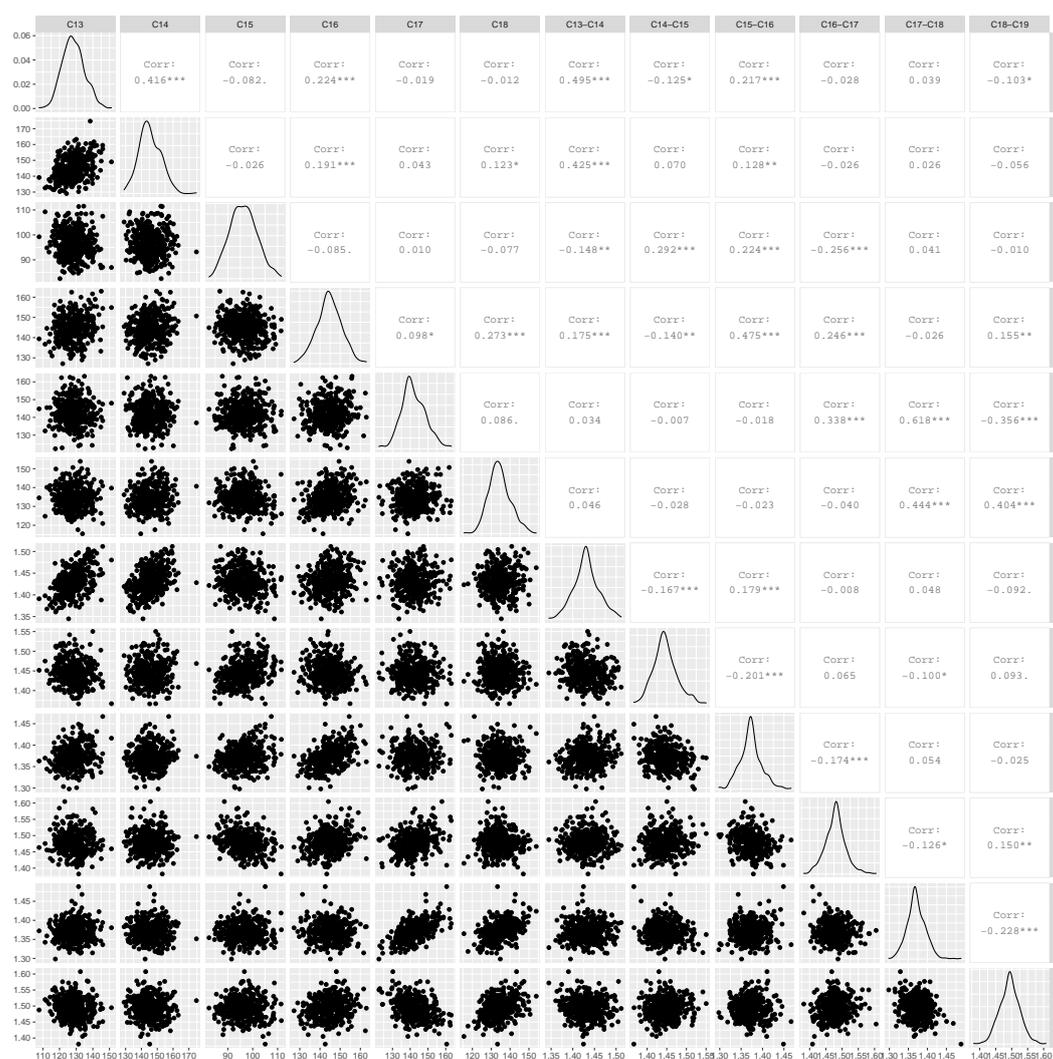

Figure A2. Explicit correlation plot of C-C bond lengths and ^{13}C chemical shifts in PCB.

References

1. Senge, M.O.; MacGowan, S.A.; O'Brien, J.M. Conformational Control of Cofactors in Nature—The Influence of Protein-Induced Macrocycle Distortion on the Biological Function of Tetrapyrroles. *Chem. Commun.* **2015**, *51*, 17031–17063. [CrossRef] [PubMed]

2. Kadish, K.; Smith, K.; Guilard, R. (Eds.) *The Porphyrin Handbook*; Academic: San Diego, CA, USA, 2000.

3. Grossman, A.R.; Schaefer, M.R.; Chiang, G.G.; Collier, J.L. The Phycobilisome, a Light-Harvesting Complex Responsive to Environmental Conditions. *Microbiol. Rev.* **1993**, *57*, 725–749. [CrossRef] [PubMed]

4. Butler, W.L.; Norris, K.H.; Siegelman, H.W.; Hendricks, S.B. Detection, Assay, and Preliminary Purification of the Pigment Controlling Photoresponsive Development of Plants. *Proc. Natl. Acad. Sci. USA* **1959**, *45*, 1703–1708. [CrossRef] [PubMed]

5. Sage, L.C. *Pigment of the Imagination: A History of Phytochrome Research*; Academic: San Diego, CA, USA, 1992.

6. Rockwell, N.C.; Su, Y.S.; Lagarias, J.C. Phytochrome Structure and Signaling Mechanisms. *Annu. Rev. Plant Biol.* **2006**, *57*, 837–858. [CrossRef]

7. Tu, S.L.; Lagarias, J.C. The Phytochromes. In *Handbook of Photosensory Receptors*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2005; pp. 121–149. [CrossRef]

8. Huq, E.; Quail, P.H. Phytochrome Signaling. In *Handbook of Photosensory Receptors*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2005; pp. 151–170. [CrossRef]

9. Vierstra, R.D.; Karniol, B. Phytochromes in Microorganisms. In *Handbook of Photosensory Receptors*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2005; pp. 171–195. [CrossRef]

10. Yeh, K.C.; Wu, S.H.; Murphy, J.T.; Lagarias, J.C. A Cyanobacterial Phytochrome Two-Component Light Sensory System. *Science* **1997**, *277*, 1505–1508. [CrossRef]

11. Schmidt, A.; Sauthof, L.; Szczepek, M.; Lopez, M.F.; Escobar, F.V.; Qureshi, B.M.; Michael, N.; Buhrke, D.; Stevens, T.; Kwiatkowski, D.; et al. Structural Snapshot of a Bacterial Phytochrome in Its Functional Intermediate State. *Nat. Commun.* **2018**, *9*, 4912. [CrossRef]

12. Ikeuchi, M.; Ishizuka, T. Cyanobacteriochromes: A New Superfamily of Tetrapyrrole-Binding Photoreceptors in Cyanobacteria. *Photochem. Photobiol. Sci.* **2008**, *7*, 1159–1167. [CrossRef]

13. Narikawa, R.; Fukushima, Y.; Ishizuka, T.; Itoh, S.; Ikeuchi, M. A Novel Photoactive GAF Domain of Cyanobacteriochrome AnPixJ That Shows Reversible Green/Red Photoconversion. *J. Mol. Biol.* **2008**, *380*, 844–855. [CrossRef]

14. Rockwell, N.C.; Martin, S.S.; Lagarias, J.C. Red/Green Cyanobacteriochromes: Sensors of Color and Power. *Biochemistry* **2012**, *51*, 9667–9677. [CrossRef]

15. Suzuki, F.; Fujita, H.; Ikeuchi, M.; Yoshihara, S.; Geng, X.X. Novel Putative Photoreceptor and Regulatory Genes Required for the Positive Phototactic Movement of the Unicellular Motile Cyanobacterium Synechocystis Sp. PCC 6803. *Plant Cell Physiol.* **2000**, *41*, 1299–1304. [CrossRef]

16. Karniol, B.; Wagner, J.R.; Walker, J.M.; Vierstra, R.D. Phylogenetic Analysis of the Phytochrome Superfamily Reveals Distinct Microbial Subfamilies of Photoreceptors. *Biochem. J.* **2005**, *392*, 103–116. [CrossRef] [PubMed]

17. Wiebeler, C.; Rao, A.G.; Gärtner, W.; Schapiro, I. The Effective Conjugation Length Is Responsible for the Red/Green Spectral Tuning in the Cyanobacteriochrome Slr1393g3. *Angew. Chem. Int. Ed.* **2019**, *58*, 1934–1938. [CrossRef] [PubMed]

18. Rockwell, N.C.; Martin, S.S.; Feoktistova, K.; Lagarias, J.C. Diverse Two-Cysteine Photocycles in Phytochromes and Cyanobacteriochromes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11854–11859. [CrossRef] [PubMed]

19. Enomoto, G.; Hirose, Y.; Narikawa, R.; Ikeuchi, M. Thiol-Based Photocycle of the Blue and Teal Light-Sensing Cyanobacteriochrome Tlr1999. *Biochemistry* **2012**, *51*, 3050–3058. [CrossRef]

20. Rockwell, N.C.; Martin, S.S.; Gulevich, A.G.; Lagarias, J.C. Phycoviolobilin Formation and Spectral Tuning in the DXCF Cyanobacteriochrome Subfamily. *Biochemistry* **2012**, *51*, 1449–1463. [CrossRef]

21. Schäfer, E.; Nagy, F. Light-Activated Intracellular Movement of Phytochrome. In *Handbook of Photosensory Receptors*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2005; pp. 197–210. [CrossRef]

22. Velazquez Escobar, F.; Hildebrandt, T.; Utesch, T.; Schmitt, F.J.; Seuffert, I.; Michael, N.; Schulz, C.; Mroginski, M.A.; Friedrich, T.; Hildebrandt, P. Structural Parameters Controlling the Fluorescence Properties of Phytochromes. *Biochemistry* **2014**, *53*, 20–29. [CrossRef]

23. Rockwell, N.C.; Ohlendorf, R.; Möglich, A. Cyanobacteriochromes in Full Color and Three Dimensions. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 806–807. [CrossRef]

24. Rumyantsev, K.A.; Shcherbakova, D.M.; Zakharova, N.I.; Verkhusha, V.V.; Turoverov, K.K. Design of Near-Infrared Single-Domain Fluorescent Protein GAF-FP Based on Bacterial Phytochrome. *Cell Tissue Biol.* **2017**, *11*, 16–26. [CrossRef]

25. Shu, X.; Royant, A.; Lin, M.Z.; Aguilera, T.A.; Lev-Ram, V.; Steinbach, P.A.; Tsien, R.Y. Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome. *Science* **2009**, *324*, 804–807. [CrossRef]

26. Zhang, J.; Wu, X.J.; Wang, Z.B.; Chen, Y.; Wang, X.; Zhou, M.; Scheer, H.; Zhao, K.H. Fused-Gene Approach to Photoswitchable and Fluorescent Biliproteins. *Angew. Chem. Int. Ed.* **2010**, *49*, 5456–5458. [CrossRef]

27. Simon, J.; Losi, A.; Zhao, K.H.; Gärtner, W. FRET in a Synthetic Flavin- and Bilin-Binding Protein. *Photochem. Photobiol.* **2017**, *93*, 1057–1062. [CrossRef] [PubMed]

28. Piatkevich, K.D.; Subach, F.V.; Verkhusha, V.V. Engineering of Bacterial Phytochromes for Near-Infrared Imaging, Sensing, and Light-Control in Mammals. *Chem. Soc. Rev.* **2013**, *42*, 3441–3452. [CrossRef] [PubMed]

29. Hou, Y.N.; Ding, W.L.; Jiang, S.P.; Miao, D.; Tan, Z.Z.; Hu, J.L.; Scheer, H.; Zhao, K.H. Bright Near-Infrared Fluorescence Bio-Labeling with a Biliprotein Triad. *Biochim. Biophys. Acta Bioenerg. Mol. Cell Res.* **2019**, *1866*, 277–284. [CrossRef] [PubMed]

30. Scheer, H.; Yang, X.; Zhao, K.H. Biliproteins and Their Applications in Bioimaging. *Procedia Chem.* **2015**, *14*, 176–185. [CrossRef]

31. Filonov, G.S.; Piatkevich, K.D.; Ting, L.M.; Zhang, J.; Kim, K.; Verkhusha, V.V. Bright and Stable Near-Infrared Fluorescent Protein for in Vivo Imaging. *Nat. Biotechnol.* **2011**, *29*, 757. [CrossRef] [PubMed]

32. Ryu, M.H.; Gomelsky, M. Near-Infrared Light Responsive Synthetic c-Di-GMP Module for Optogenetic Applications. *ACS Synth. Biol.* **2014**, *3*, 802–810. [CrossRef]

33. Shimizu-Sato, S.; Huq, E.; Tepperman, J.M.; Quail, P.H. A Light-Switchable Gene Promoter System. *Nat. Biotechnol.* **2002**, *20*, 1041. [CrossRef]

34. Wu, X.J.; Yang, H.; Sheng, Y.; Zhu, Y.L.; Li, P.P. Fluorescence Properties of a Novel Cyanobacteriochrome GAF Domain from Spirulina That Exhibits Moderate Dark Reversion. *Int. J. Mol. Sci.* **2018**, *19*, 2253. [CrossRef]

35. Essen, L.O.; Mailliet, J.; Hughes, J. The Structure of a Complete Phytochrome Sensory Module in the Pr Ground State. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14709–14714. [CrossRef]

36. Rüdiger, W.; Thümmler, F.; Cmiel, E.; Schneider, S. Chromophore Structure of the Physiologically Active Form (P(Fr)) of Phytochrome. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6244–6248. [CrossRef]

37. Song, C.; Rohmer, T.; Tiersch, M.; Zaanen, J.; Hughes, J.; Matysik, J. Solid-State NMR Spectroscopy to Probe Photoactivation in Canonical Phytochromes. *Photochem. Photobiol.* **2013**, *89*, 259–273. [CrossRef] [PubMed]

38. Rohmer, T.; Lang, C.; Bongards, C.; Gupta, K.B.S.S.; Neugebauer, J.; Hughes, J.; Gärtner, W.; Matysik, J. Phytochrome as Molecular Machine: Revealing Chromophore Action during the Pfr—Pr Photoconversion by Magic-Angle Spinning NMR Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132*, 4431–4437. [CrossRef] [PubMed]

39. Heyes, D.J.; Khara, B.; Sakuma, M.; Hardman, S.J.O.; O'Cualain, R.; Rigby, S.E.J.; Scrutton, N.S. Ultrafast Red Light Activation of Synechocystis Phytochrome Cph1 Triggers Major Structural Change to Form the Pfr Signalling-Competent State. *PLoS ONE* **2012**, *7*, e52418. [CrossRef] [PubMed]

40. Song, C.; Psakis, G.; Lang, C.; Mailliet, J.; Gärtner, W.; Hughes, J.; Matysik, J. Two Ground State Isoforms and a Chromophore D-Ring Photoflip Triggering Extensive Intramolecular Changes in a Canonical Phytochrome. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3842–3847. [CrossRef]

41. Song, C.; Psakis, G.; Kopycki, J.; Lang, C.; Matysik, J.; Hughes, J. The D-Ring, Not the a-Ring, Rotates in Synechococcus OS-B Phytochrome. *J. Biol. Chem.* **2014**, *289*, 2552–2562. [CrossRef]

42. Hirose, Y.; Rockwell, N.C.; Nishiyama, K.; Narikawa, R.; Ukaji, Y.; Inomata, K.; Lagarias, J.C.; Ikeuchi, M. Green/Red Cyanobacteriochromes Regulate Complementary Chromatic Acclimation via a Protochromic Photocycle. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4974–4979. [CrossRef]

43. Scarbath-Evers, L.K.; Jähnigen, S.; Elgabarty, H.; Song, C.; Narikawa, R.; Matysik, J.; Sebastiani, D. Structural Heterogeneity in a Parent Ground-State Structure of AnPixJg2 Revealed by Theory and Spectroscopy. *Phys. Chem. Chem. Phys.* **2017**, *19*, 13882–13894. [CrossRef]

44. Peng, P.P.; Dong, L.L.; Sun, Y.F.; Zeng, X.L.; Ding, W.L.; Scheer, H.; Yang, X.; Zhao, K.H. The Structure of Allophycocyanin B from Synechocystis PCC 6803 Reveals the Structural Basis for the Extreme Redshift of the Terminal Emitter in Phycobilisomes. *Acta Crystallogr. Sect. D* **2014**, *70*, 2558–2569. [CrossRef]

45. Rockwell, N.C.; Martin, S.S.; Gulevich, A.G.; Lagarias, J.C. Conserved Phenylalanine Residues Are Required for Blue-Shifting of Cyanobacteriochrome Photoproducts. *Biochemistry* **2014**, *53*, 3118–3130. [CrossRef]

46. Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; et al. Sequence Analysis of the Genome of the Unicellular Cyanobacterium Synechocystis Sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-Coding Regions. *DNA Res.* **1996**, *3*, 109–136. [CrossRef]

47. Hughes, J.; Lamparter, T.; Mittmann, F.; Hartmann, E.; Gärtner, W.; Wilde, A.; Börner, T. A Prokaryotic Phytochrome. *Nature* **1997**, *386*, 663–663. [CrossRef] [PubMed]

48. Hahn, J.; Strauss, H.M.; Landgraf, F.T.; Gimenèz, H.F.; Lochnit, G.; Schmieder, P.; Hughes, J. Probing Protein–Chromophore Interactions in Cph1 Phytochrome by Mutagenesis. *FEBS J.* **2006**, *273*, 1415–1429. [CrossRef] [PubMed]

49. Takeda, K.; Terazima, M. Photoinduced Orientation Change of the Dimer Structure of the Pr-i State of Cph1Δ2. *Biochemistry* **2018**, *57*, 5058–5071. [CrossRef] [PubMed]

50. Velazquez Escobar, F.; Lang, C.; Takiden, A.; Schneider, C.; Balke, J.; Hughes, J.; Alexiev, U.; Hildebrandt, P.; Mroginski, M.A. Protonation-Dependent Structural Heterogeneity in the Chromophore Binding Site of Cyanobacterial Phytochrome Cph1. *J. Phys. Chem. B* **2017**, *121*, 47–57. [CrossRef]

51. Stöppler, D.; Song, C.; van Rossum, B.J.; Geiger, M.A.; Lang, C.; Mroginski, M.A.; Jagtap, A.P.; Sigurdsson, S.T.; Matysik, J.; Hughes, J.; et al. Dynamic Nuclear Polarization Provides New Insights into Chromophore Structure in Phytochrome Photoreceptors. *Angew. Chem. Int. Ed.* **2016**, *55*, 16017–16020. [CrossRef]

52. Yang, Y.; Heyne, K.; Mathies, R.A.; Dasgupta, J. Non-Bonded Interactions Drive the Sub-Picosecond Bilin Photoisomerization in the Pfr State of Phytochrome Cph1. *ChemPhysChem* **2016**, *17*, 369–374. [CrossRef]

53. Kim, P.W.; Rockwell, N.C.; Martin, S.S.; Lagarias, J.C.; Larsen, D.S. Heterogeneous Photodynamics of the Pfr State in the Cyanobacterial Phytochrome Cph1. *Biochemistry* **2014**, *53*, 4601–4611. [CrossRef]

54. Kim, P.W.; Pan, J.; Rockwell, N.C.; Chang, C.W.; Taylor, K.C.; Clark Lagarias, J.; Larsen, D.S. Ultrafast E to Z Photoisomerization Dynamics of the Cph1 Phytochrome. *Chem. Phys. Lett.* **2012**, *549*, 86–92. [CrossRef]

55. Spillane, K.M.; Dasgupta, J.; Mathies, R.A. Conformational Homogeneity and Excited-State Isomerization Dynamics of the Bilin Chromophore in Phytochrome Cph1 from Resonance Raman Intensities. *Biophys. J.* **2012**, *102*, 709–717. [CrossRef]

56. Mailliet, J.; Psakis, G.; Feilke, K.; Sineshchekov, V.; Essen, L.O.; Hughes, J. Spectroscopy and a High-Resolution Crystal Structure of Tyr263 Mutants of Cyanobacterial Phytochrome Cph1. *J. Mol. Biol.* **2011**, *413*, 115–127. [CrossRef]

57. Song, C.; Psakis, G.; Lang, C.; Mailliet, J.; Zaanen, J.; Gärtner, W.; Hughes, J.; Matysik, J. On the Collective Nature of Phytochrome Photoactivation. *Biochemistry* **2011**, *50*, 10987–10989, [CrossRef] [PubMed]

58. Kaminski, S.; Mroginski, M.A. Molecular Dynamics of Phycocyanobilin Binding Bacteriophytochromes: A Detailed Study of Structural and Dynamic Properties. *J. Phys. Chem. B* **2010**, *114*, 16677–16686. [CrossRef] [PubMed]

59. Mroginski, M.A.; von Stetten, D.; Escobar, F.V.; Strauss, H.M.; Kaminski, S.; Scheerer, P.; Günther, M.; Murgida, D.H.; Schmieder, P.; Bongards, C.; et al. Chromophore Structure of Cyanobacterial Phytochrome Cph1 in the Pr State: Reconciling Structural and Spectroscopic Data by QM/MM Calculations. *Biophys. J.* **2009**, *96*, 4153–4163. [CrossRef] [PubMed]

60. Rockwell, N.C.; Shang, L.; Martin, S.S.; Lagarias, J.C. Distinct Classes of Red/Far-Red Photochemistry within the Phytochrome Superfamily. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6123–6127. [CrossRef]

61. Hahn, J.; Strauss, H.M.; Schmieder, P. Heteronuclear NMR Investigation on the Structure and Dynamics of the Chromophore Binding Pocket of the Cyanobacterial Phytochrome Cph1. *J. Am. Chem. Soc.* **2008**, *130*, 11170–11178. [CrossRef]

62. Matute, R.A.; Contreras, R.; Pérez-Hernández, G.; González, L. The Chromophore Structure of the Cyanobacterial Phytochrome Cph1 as Predicted by Time-Dependent Density Functional Theory. *J. Phys. Chem. B* **2008**, *112*, 16253–16256. [CrossRef]

63. Rohmer, T.; Lang, C.; Hughes, J.; Essen, L.O.; Gärtner, W.; Matysik, J. Light-Induced Chromophore Activity and Signal Transduction in Phytochromes Observed by 13C and 15N Magic-Angle Spinning NMR. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15229–15234. [CrossRef]

64. Murgida, D.H.; von Stetten, D.; Hildebrandt, P.; Schwinté, P.; Siebert, F.; Sharda, S.; Gärtner, W.; Mroginski, M.A. The Chromophore Structures of the Pr States in Plant and Bacterial Phytochromes. *Biophys. J.* **2007**, *93*, 2410–2417. [CrossRef]

65. Rockwell, N.C.; Lagarias, J.C. The Structure of Phytochrome: A Picture Is Worth a Thousand Spectra. *Plant Cell* **2006**, *18*, 4–14. [CrossRef]

66. Rohmer, T.; Strauss, H.; Hughes, J.; de Groot, H.; Gärtner, W.; Schmieder, P.; Matysik, J. 15N MAS NMR Studies of Cph1 Phytochrome: Chromophore Dynamics and Intramolecular Signal Transduction. *J. Phys. Chem. B* **2006**, *110*, 20580–20585. [CrossRef]

67. Narikawa, R.; Ishizuka, T.; Muraki, N.; Shiba, T.; Kurisu, G.; Ikeuchi, M. Structures of Cyanobacteriochromes from Phototaxis Regulators AnPixJ and TePixJ Reveal General and Specific Photoconversion Mechanism. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 918–923. [CrossRef] [PubMed]

68. Song, C.; Narikawa, R.; Ikeuchi, M.; Gärtner, W.; Matysik, J. Color Tuning in Red/Green Cyanobacteriochrome AnPixJ: Photoisomerization at '5 Causes an Excited-State Destabilization. *J. Phys. Chem. B* **2015**, *119*, 9688–9695. [CrossRef] [PubMed]

69. Xu, X.L.; Gutt, A.; Mechelke, J.; Raffelberg, S.; Tang, K.; Miao, D.; Valle, L.; Borsarelli, C.D.; Zhao, K.H.; Gärtner, W. Combined Mutagenesis and Kinetics Characterization of the Bilin-Binding GAF Domain of the Protein Slr1393 from the Cyanobacterium Synechocystis PCc6803. *ChemBioChem* **2014**, *15*, 1190–1199. [CrossRef] [PubMed]

70. Velazquez Escobar, F.; Utesch, T.; Narikawa, R.; Ikeuchi, M.; Mroginski, M.A.; Gärtner, W.; Hildebrandt, P. Photoconversion Mechanism of the Second GAF Domain of Cyanobacteriochrome AnPixJ and the Cofactor Structure of Its Green-Absorbing State. *Biochemistry* **2013**, *52*, 4871–4880. [CrossRef]

71. Fukushima, Y.; Iwaki, M.; Narikawa, R.; Ikeuchi, M.; Tomita, Y.; Itoh, S. Photoconversion Mechanism of a Green/Red Photosensory Cyanobacteriochrome AnPixJ: Time-Resolved Optical Spectroscopy and FTIR Analysis of the AnPixJ-GAF2 Domain. *Biochemistry* **2011**, *50*, 6328–6339. [CrossRef] [PubMed]

72. Song, C.; Velazquez Escobar, F.; Xu, X.L.; Narikawa, R.; Ikeuchi, M.; Siebert, F.; Gärtner, W.; Matysik, J.; Hildebrandt, P. A Red/Green Cyanobacteriochrome Sustains Its Color despite a Change in the Bilin Chromophores Protonation State. *Biochemistry* **2015**, *54*, 5839–5848. [CrossRef] [PubMed]

73. Rao, A.G.; Wiebeler, C.; Sen, S.; Cerutti, D.S.; Schapiro, I. Histidine Protonation Controls Structural Heterogeneity in the Cyanobacteriochrome AnPixJg2. *bioRxiv* **2020**. [CrossRef]

74. Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

75. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935, [CrossRef]

76. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802, [CrossRef]

77. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616, [CrossRef] [PubMed]

78. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690, [CrossRef] [PubMed]

79. Feller, S.E.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621, [CrossRef]

80. Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519, [CrossRef]

81. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

82. The CP2K Developers Group. 2001–2020. Available online: www.cp2k.org (accessed on 20 November 2020).

83. Schiffmann, C.; Sebastiani, D. Artificial Bee Colony Optimization of Capping Potentials for Hybrid Quantum Mechanical/Molecular Mechanical Calculations. *J. Chem. Theory Comput.* **2011**, *7*, 1307–1315. [CrossRef]

84. Ihrig, A.C.; Schiffmann, C.; Sebastiani, D. Specific Quantum Mechanical/Molecular Mechanical Capping-Potentials for Biomolecular Functional Groups. *J. Chem. Phys.* **2011**, *135*, 214107. [CrossRef]

85. Becke, A.D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100. [CrossRef]

86. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789. [CrossRef]

87. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104. [CrossRef]

88. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128. [CrossRef]

89. Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710. [CrossRef] [PubMed]

90. Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic Separable Dual-Space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662. [CrossRef]

91. Krack, M. Pseudopotentials for H to Kr Optimized for Gradient-Corrected Exchange-Correlation Functionals. *Theor. Chem. Acc.* **2005**, *114*, 145–152. [CrossRef]

92. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184. [CrossRef] [PubMed]

93. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378. [CrossRef]

94. Tuckerman, M.E. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press Inc.: New York, NY, USA, 2010.

95. Weber, V.; Iannuzzi, M.; Giani, S.; Hutter, J.; Declerck, R.; Waroquier, M. Magnetic Linear Response Properties Calculations with the Gaussian and Augmented-Plane-Wave Method. *J. Chem. Phys.* **2009**, *131*, 014106. [CrossRef]

96. Jensen, F. Polarization Consistent Basis Sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125. [CrossRef]

97. Keith, T.A.; Bader, R.F. Calculation of Magnetic Response Properties Using a Continuous Set of Gauge Transformations. *Chem. Phys. Lett.* **1993**, *210*, 223–231. [CrossRef]

98. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.r-project.org/ (accessed on 20 November 2020).

99. Gowers, R.J.; Linke, M.; Barnoud, J.; Reddy, T.J.E.; Melo, M.N.; Seyler, S.L.; Domański, J.; Dotson, D.L.; Buchoux, S.; Kenney, I.M.; et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Proceedings of the 15th Python in Science Conference, Austin, TX, USA, 11–17 July 2016; pp. 98–105. Available online: http://conference.scipy.org/proceedings/scipy2016/pdfs/proceedings.pdf (accessed on 20 November 2020). [CrossRef]

100. Michaud-Agrawal, N.; Denning, E.J.; Woolf, T.B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327. [CrossRef]

101. Oliphant, T.E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [CrossRef]

102. Millman, K.J.; Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **2011**, *13*, 9–12. [CrossRef]

103. Perez, F.; Granger, B.E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21–29. [CrossRef]

104. Python Software Foundation, Version 3.6. Available online: www.python.org (accessed on 20 November 2020).

105. Harris, C.R.; Jarrod Millman, K.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020** *585*, 357–362. [CrossRef]

106. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020** *17*, 261–272. [CrossRef]

107. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

108. Stone, J. Tachyon. An Efficient Library for Parallel Ray Tracing and Animation. Available online: jedi.ks.uiuc.edu/~johns/raytracer (accessed on 20 November 2020).

109. Chemfig—Draw Molecules with Easy Syntax, Version 1.56. 2010–2020 Christian Tellechea. Available online: https://ctan.org/tex-archive/macros/generic/chemfig (accessed on 20 November 2020).

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.