

Understanding proton transfer in phytochromes

vorgelegt von

MSc.

Ronald González Medina

ORCID: 0000-0003-2520-1215 

von der Fakultät II - Mathematik und Naturwissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

-Dr. rer. nat.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Roderich Süßmuth

Gutachter: Prof. Dr. Maria Andrea Mroginski

Gutachter: Prof. Dr. Petra Imhof

Tag der wissenschaftlichen Aussprache: 11. November 2021

Berlin 2022

Zusammenfassung

Phytochrome haben die interessante Fähigkeit, zwischen rot-absorbierenden (Pr) und fern-rot-absorbierenden (Pfr) Zuständen zu photokonvertieren. Dieser Photokonversionsprozess hängt jedoch von Schlüsselinteraktionen zwischen einem Bilin-Chromophor und der Proteinmatrix ab. Daher ist die Identifizierung der chemischen Natur und die Quantifizierung der Chromophor-Protein-Interaktionen ein wesentlicher Schritt zum Verständnis des Photokonversionsprozesses sowie der Protonentransfer-Ereignisse, die während des Photozyklus stattfinden. In dieser Arbeit präsentierten wir einen leistungsstarken und einfachen Ansatz, der auf der Fragment-Molekularorbital-Methode (FMO) basiert, um die Natur und die Stärke der nicht-kovalenten Wechselwirkungen zwischen dem Biliverdin (BV)-Chromophor und dem Protein des *Deinococcus radiodurans* Phytochroms (DrBphP) im Pr-Zustand zu identifizieren und zu quantifizieren.

Eine der interessantesten Eigenschaften der FMO-Methode ist der Ansatz der “pair interaction energy decomposition analysis” (PIEDA). Die Anwendung von PIEDA führte zur Identifizierung des Pyrrolwassers, Asp207 und Glu27 als Schlüsselreste für die Stabilisierung der Pyrrolringe des BV-Chromophors durch die Bildung von sechs H-Bindungen. Darüber hinaus wurden die konservierten Arg254 und His260 als essentielle Reste für die Konformationsstabilität der beiden Propionseitenketten B und C identifiziert. Interessanterweise wurden neue Wechselwirkungen in der Chromophor-Bindungstasche identifiziert, zwei nicht-klassische H-Bindungen (CH/O-Wechselwirkungen) zwischen Asp207 und Tyr263 sowie eine OH/ π -Wechselwirkung zwischen Tyr263 und Ring D des BV-Chromophors, die photochemische Relevanz haben könnten.

Denn die Photokonversionsreaktion von Phytochromen wird zum Teil durch Protonierungszustandsänderungen des Chromophors und von Schlüsselresten der Proteinmatrix erreicht. Die genaue Bestimmung der Protonierungszustände des Chromophors und der titrierbaren Aminosäuren ist ein wesentlicher Schritt, um den Protonentransferprozess zwischen BV und Proteinmatrix zu verstehen. Zu diesem Zweck wurden zwei verschiedene Ansätze auf das Phytochrom Agp2 im Pfr- und Meta-F-Zustand angewendet. Der erste Ansatz basiert auf der Kombination von elektrostatischen Energieberechnungen durch Lösung der Poisson-Boltzmann-Gleichung (PBE) mit klassischen Molekulardynamik (MD)-Simulationen. Der zweite ist die Methode der konstanten pH-Molekulardynamik (CpHMD), die für die Untersuchung mehrerer wichtiger biologischer Prozesse verwendet wurde, die durch Änderungen des Lösungs-pH verursacht werden. Einer der Hauptvorteile der CpHMD-Ansätze ist, dass sich die Protonierungszustände während der Konformationsdynamik ändern können. Somit können

die dynamischen Prozesse, die mit einer Änderung der Protonierungszustände gekoppelt sind, direkt untersucht werden.

Die mit dem ersten Ansatz erhaltenen Ergebnisse führten dazu, dass die Propionssäureseitenkette von Ring C sowohl im Pfr- als auch im Meta-F-Zustand deprotoniert ist. Diese Ergebnisse stehen jedoch im Widerspruch zu den experimentellen Raman-Spektren vom Agp2 Phytochrom. Im Gegensatz dazu führte die Anwendung der CpHMD-Methode auf das Agp2-Phytochrom zu einer sowohl im Pfr- als auch im Meta-F-Zustand protonierten Propionsäureseitenkette des Rings C sowie zur Identifizierung des His278 als potentieller Protonenakzeptor. Die mit der CpHMD-Methode erhaltenen pKa-Werte stehen somit in guter Übereinstimmung mit den experimentellen Ergebnissen. Darüber hinaus identifizierte dieser Ansatz das His278 als potentiellen Protonenakzeptor, da der Protonierungszustand von pscC und His278 hoch korreliert sind. Zusammenfassend lässt sich sagen, dass durch die Verwendung des CpHMD-Ansatzes die dynamischen Prozesse, die mit einer Änderung der Protonierungszustände von Phytochromstrukturen gekoppelt sind, direkt untersucht werden können, während bei der Verwendung klassischer MD-Simulationen die Protonierungszustände während der Konformationsdynamik fixiert sind, was zu einem unrealistischen Sampling führt.

Interessanterweise können die für das Agp2-Phytochrom abgeleiteten Schlussfolgerungen auf alle Bathy-Phytochrome ausgedehnt werden, da die in der Chromophor-Bindungstasche befindlichen Reste hoch konserviert sind. Darüber hinaus zeigen diese Ergebnisse nicht nur die Rolle von His278 im Photozyklus, sondern auch die Wichtigkeit der Durchführung eines besseren Samplings der His-Reste, die in direktem Kontakt mit dem Chromophor-Molekül stehen.

Abstract

Phytochromes have an interesting ability to photoconvert between red-absorbing (Pr) and far-red-absorbing (Pfr) states. However, this photoconversion process depends on key interactions between a bilin chromophore and protein matrix, thus, the identification of the chemical nature and quantification of chromophore-protein interactions constitute an essential step for understanding the photoconversion process, as well as the proton transfer events taking place during the photocycle. In this work, we presented a powerful and straightforward approach based on the fragment molecular orbital (FMO) method to identify the nature and quantify the strength of the noncovalent interactions at a fully quantum mechanical level between the biliverdin (BV) chromophore and protein of the *Deinococcus radiodurans* phytochrome (DrBphP) in the Pr state.

One of the most interesting features of the FMO method is the pair interaction energy decomposition analysis (PIEDA) approach, the application of PIEDA led to the identification of the pyrrole water, Asp207, and Glu27 as key residues for the stabilization of the pyrrole rings of the BV chromophore through the formation of six H-bonds. Furthermore, the conserved Arg254 and His260 were also identified as essential residues in the conformational stability of both propionic side chains B and C. Interestingly, new interactions were identified in the chromophore-binding pocket, two nonclassical H-bonds (CH/O interactions) between Asp207 and Tyr263, and an OH/ π interaction between Tyr263 and ring D of the BV chromophore, which might have photochemical relevance.

Since the photoconversion reaction of phytochromes is achieved in part by protonation state changes of the chromophore and key residues of the protein matrix. The accurate determination of the protonation states of the chromophore and titratable amino acids constitutes an essential step in order to understand the proton transfer process between the BV and protein matrix. To this end, two different approaches were applied to the Agp2 phytochrome in the Pfr and Meta-F states. The first approach is based on the combination of electrostatic energy computations by solving the Poisson Boltzmann equation (PBE) with classical molecular dynamics (MD) simulations. The second one is the constant pH molecular dynamics (CpHMD) method, which has been used for studying several important biological processes caused by changes in solution pH. One of the main advantages of the CpHMD approaches is that the protonation states may change during the conformational dynamics. Thus, the dynamical processes coupled to a change in protonation states can be directly studied.

The results obtained with the first approach led to a propionic side chain of ring C deprotonated in both Pfr and Meta-F states. However, these results are in contradiction with the experimental Raman spectra of Agp2 phytochrome. In contrast, the application of the CpHMD method to the Agp2 phytochrome led to a propionic side chain of ring C protonated in both Pfr and Meta-F states as well as the identification of the His278 as a potential proton site acceptor. Thus, the pKa values obtained with the CpHMD method are in good agreement with the experimental results. Furthermore, this approach identified the His278 as a potential proton site acceptor, since the protonation state of pscC and His278 are highly correlated. In summary, by using the CpHMD approach, the dynamical processes coupled to a change in protonation states of phytochromes structures can be directly studied, whilst, by using classical MD simulations, the protonation states are fixed during the conformational dynamic, which leads to an unrealistic sampling.

Interestingly, the conclusions derived for the Agp2 phytochrome can be extended to all bathy phytochromes, since the residues located in the chromophore binding pocket are highly conserved, furthermore, these results, not only reveal the role of His278 in the photocycle but also the importance of carrying out a better sampling of the His residues that are in direct contact with the chromophore molecule.

Publications

The work presented in this thesis has been published partially in the following articles:

1. Fernandez Lopez, M., Nguyen, A. D., Velazquez Escobar, F., **González, R.**, Michael, N., Nogacz, Z., Piwowarski, P., Bartl, F., Siebert, F., Heise, I., Scheerer, P., Gärtner, W., Mroginski, M. A., and Hildebrandt, P. (2019). Role of the Propionic Side Chains for the Photoconversion of Bacterial Phytochromes. *Biochemistry*, 58(33), 3504-3519.
2. **González, R.**, and Mroginski, M. A. (2019). Fully Quantum Chemical Treatment of Chromophore-Protein Interactions in Phytochromes. *The Journal of Physical Chemistry B*, 123(46), 9819-9830.
3. Battocchio, G., **González, R.**, Rao, Aditya., Schapiro, I., and Mroginski, M. A. (2020). Correlated Motions in *Deinococcus Radiodurans* Bacteriophytochrome Photosensory Domain. *The Journal of Physical Chemistry B*, 124(9), 1740-1750.
4. Mroginski, M. A., ..., **González, R.**, ..., Schapiro, I. (2021) *Frontiers in Multiscale Modeling of Photoreceptor Proteins. Photochemistry and Photobiology*, 2021, vol. 97, n° 2, p. 243-269.

The following publications connected to the results presented in chapter 4 and 5 are in preparation.

5. **González, R.**, and Mroginski, M. A. (2021). Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes
6. Battocchio, G*, **González, R.***, and Mroginski, M. A. (2021). pKa Calculations in Bathy Phytochromes Structures, (*equal contribution)

Table of Contents

Title Page	i
Zusammenfassung	iii
Abstract	v
List of Figures	xi
List of Tables	xvii
1 Introduction	1
1.1 Biological systems	1
1.1.1 Phytochromes	1
1.1.2 Chromophore structure	2
1.2 Electronic structure methods	3
1.2.1 The Self-Consistent Charge Density Functional Tight Binding formalism (SCC-DFTB)	5
1.3 The fragment molecular orbital (FMO) method	5
1.4 Pair interaction energy decomposition analysis (PIEDA)	9
1.5 Computations of Raman spectra using the fragment molecular orbital method .	10
1.6 Molecular dynamic simulations	12
1.6.1 Non-polarizable force fields	12
1.7 Electrostatic energy calculations for computing pKa values	13
1.7.1 Calculating pKa values in proteins	13
1.7.2 Electrostatic energy calculations	15
1.7.3 The Poisson-Boltzmann equation	16
1.8 pKa calculations with constant pH molecular dynamics	16
1.8.1 Titration curve and pKa prediction calculation	17
2 Fully Quantum Chemical Treatment of Chromophore-protein Interactions in Phytochromes	19
2.1 Introduction	19
2.2 Methods	20
2.3 Results and Discussion	21
2.4 Conclusions	34

3	The Fragment Molecular Orbital Method: Application to Phytochromes Structures	37
3.1	Role of the Propionic Side Chains for the Photoconversion of Bacterial Phytochromes	38
3.2	Introduction	38
3.3	Methods	38
3.4	Results and discussion.	40
3.5	Conclusions	48
3.6	Simulation of Raman spectra using the fragment molecular orbital method . .	49
4	Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes	55
4.1	Introduction	55
4.2	Methods	55
4.3	Results and Discussion	58
	4.3.1 Validation of Force Field Parameters.	58
4.4	Conclusions	65
5	pKa Calculations in Bathy Phytochromes Structures	67
5.1	Introduction	67
5.2	Methods	68
5.3	Results and Discussion	69
	5.3.1 pKa caculations based on classical MD simulations and electrostatic energy calculations.	69
	5.3.2 Dowser waters and Karlsberg2+	74
	5.3.3 pKa values using the constant-pH method	76
5.4	Conclusions	77
6	Final Conclusions and Outlook	79
	References	81

List of Figures

1.1	Photosensory core module (PCM) of Agp2 phytochrome structure divided into PAS (color yellow), GAF (color green) and PHY (color blue) domains, (PDB entry 6G1Y, chain A)[9]	2
1.2	left: Photosensory core module of Agp2 phytochrome structure, right: pyrrole water, two highly conserved amino acid residues and BV molecule in the GAF domain.	3
1.3	a) Chemical structure of the tetrapyrrole chromophore and chemical substitutions located at the rings A and D, with R3 equal either a vinyl group (BV and (PΦB)) or an ethyl group (PCB). b) Three different forms of binding between ring A and protein matrix.	4
1.4	Generic phytochrome photocycle with the red light absorbing parent state (Pr) and far-red light absorbing parent state (Pfr).	4
1.5	General procedure for calculating the energy of a molecular system divided into three fragments. In the first step, the monomer densities are built. Then, a self-consistent charge (SCC) procedure is applied to the monomer densities by which the energy of each monomer is solved by means of self-consistent field (SCF) calculations including the electrostatic potential (ESP) generated by the remaining fragments (dash lines). In the third step, fragment pair SCF calculations are performed (i.e., the inter-fragment interaction energy) and, in the last step, the total energy of the system is obtained.	7
1.6	Covalent bond fragmented within the FMO method for ethane molecule C_2H_6 ,	8
1.7	Fragmentation of a dipeptide in the FMO method. (a) The molecule is fragmented at $C\alpha(sp^3)$ carbon atom.(b) The molecule is fragmented at the sp^2 carbon atom.	9
1.8	Phytochrome structure divided into two layers (L1, and L2). The first step in FMO/FDD is the computation of the whole system at the FMO1 level (monomers) for the initial geometry; then, the electronic state of fragments in domain F is frozen for the other geometries during a geometry optimization. Fragments that belong to the polarizable and active region share the same basis set, nevertheless, only fragments that belong to the active region are allowed to move during the optimization.	11

2.1	a) Crystal structure of the <i>Deinococcus radiodurans</i> BphP phytochrome in the Pr state (PDB entry 4q0h)[57] and b) chromophore binding pocket. The structural model contains all crystal water molecules, protein and BV chromophore (6159 atoms). Displayed images were created with UCSF Chimera software[59].	20
2.2	The BV-chromophore divided into four fragments. For preserving the chemical identity of the bilin chromophore and obtain a more detailed description of chromophore-protein, both propionic side chain B (propB) and propionic side chain C (propC) are treated each as a single fragment. Rings B,C and D form only one fragment (rings-BCD), and ring A along with Cys24 form the fragment ring-A-Cys.	22
2.3	a) Phytochrome structure of the DrBphP-hip structural model in the Pr state colored according to displacement after geometry optimization using the FMO-DFTB3-D/PCM method (0.44 and 3.36 Å). This representation shows residues with a low displacement (0.44 Å) in blue, those with a high displacement (3.36 Å) in red, and those in-between as white. The RMSD was also calculated after alignment of backbone heavy atoms of structural model DrBphP-hip with respect to the crystal structure (1.38 Å). b) BV molecule and residues in the vicinity of the BV chromophore of the optimized DrBphP-hip structural model (color cyan) compared to the crystal structure (color gray).	23
2.4	a) H-bond distances between pyrrol water oxygen with His260 (R1), the three pyrrole nitrogens of rings A (R2), B (R3), and C (R4), the carbonyl group of ring A (R5), carbonyl group of Asp207 (R6) and propionic side chain C (R7) in the crystal structure. b) H-bond interactions in the optimized DrBphP-hip structural model. c) H-bond interactions in the optimized DrBphP-hid structural model d) H-bond interactions in the optimized DrBphP-hie structural model.	26
2.5	a) PIEs plot between BV-chromophore and protein for DrBphP-Hip optimized model. b) PIEs plot for DrBphP-Hid optimized model. c) PIEs plot for DrBphP-Hie optimized model.	29
2.6	PIEs plot (left) and PIEDA plot (right) between BV-chromophore and protein of the DrBphP-hip structural model for a) ring-A-Cys b) rings-BCD c) propB and d) propC fragments. In PIEDA plots the energy terms: electrostatics, exchange repulsion, charge-transfer, dispersion and solvent screening are colored coded blue, red, yellow, green and red wine respectively.	32
2.7	Structure of the chromophore binding pocket highlighting key residues and water molecules interacting with the BV tetrapyrrole in the DrBphP-hip structural model.	33
2.8	Zoom into the structure of the chromophore binding pocket highlighting key residues and water molecules interacting with propionic side chains B and C of the BV cofactor in the DrBphP-hip structural model.	34

2.9	Interactions detected by FMO in the chromophore binding pocket: two non-classical H-bonds (CH/O interactions)[81, 82] between the carbonyl oxygens of the Asp207 and the side chain of Tyr263; one OH/ π interaction between the hydroxyl of Tyr263 and ring D of the BV-chromophore.	35
3.1	Chemical structures of the biliverdin monomethylesters with the ester function (blue) at the propionic side chain of ring B (BVM-B, left) and ring C (BVM-C, right). The chromophores are shown in the ZZZssa configuration of the Pr state; in Pfr, the chromophore adopts the ZZEssa configuration generated by rotation around the C(15)=C(16) bond (red). The chromophore is covalently bound to the protein via addition of a Cys side chain to the vinyl function of ring A (yellow).	39
3.2	Chromophore binding pockets of (left) Agp1-BVM-B (cyan) and (right) Agp1-BVM-C (green). The optimized QM/MM structures of the two Agp1 variants are compared with the QM/MM structure of Agp1-BV (gray).	40
3.3	Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the propionic side chain B (top) and propionic side C (bottom) computed for models Agp1-BV (green), Agp1-BVM-B (blue) and Agp1-BVM-C (red).	41
3.4	Structures of the chromophore binding pockets (left) Agp2-BV (cyan) and (right) Agp2-BVM-B (pink) in the Pfr state. The optimized QM/MM structures are compared with the crystal structure of Agp2-BV (gray, PDB entry 6G1Y).	42
3.5	Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the PsB (top) and PsC (bottom) computed for models Agp2-BV (green) and Agp2-BVM-B (blue) in the Pfr state.	43
3.6	Structures of the chromophore binding pockets of the (left) Agp2-BVM-B-keto (orange) and (right) Agp2-BVM-B-enol (green) in the Meta-F(P) state. The optimized QM/MM structures are compared with the optimized QM/MM structure of Agp2-BV in the Meta-F(P) state (gray).	43
3.7	Structures of the chromophore binding pockets of photoproduct states of Agp2-BVM-B-enol with (left) cationic His278 (green) and (right) charge-neutral His278, protonated at N δ (orange). The two structural models for Meta-F(D) are compared with the QM/MM-optimized structure of Agp2-BVM-B (enol) in the Meta-F(P) state carrying a protonated PsC (gray).	44
3.8	Pair Interaction Energies (PIE) in kcal/mol between protein matrix and rings B, C, D and ring A computed for the models Agp2-BV-enol (light green), Agp2-BV-keto (green), Agp2-BVM-B-enol (light blue) and Agp2-BVM-B-keto (blue).	45
3.9	Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the propionic side chains, rings B,C,D and A calculated for the models Agp2-BVM-B-enol in the Meta-F state with His278 doubly protonated and neutral His278d (protonated at delta position), compared to the Meta-f state (enol) (see Figure 3.7)	47

3.10	Cyanobacteriochrome structure divided into two layers (L1, and L2). The electronic state of fragments in domain F (Layer 1) is frozen during a geometry optimization. Fragments that belong to the polarizable and active region (Layer 2) share the same basis set, nevertheless, only fragments that belong to the active region are allowed to move during the optimization.	50
3.11	The optimized FMO/FDD structure of the cyanobacteriochrome structure (Slr1393, PDB entry 5M82[93]) in the Pg state (green) is compared with the crystal structure (gray). The rmsd was calculated after alignment of backbone heavy atoms of the optimized structural model with respect to the crystal structure (0.38 Å).	51
3.12	PCB molecule of Slr1393 in Pg state and residues in the vicinity of the PCB chromophore of the optimized structural model (color cyan) compared to the crystal structure (color gray). The rmsd for the PCB molecule was calculated for the heavy atoms compared to the PCB in the crystal structure (0.34 Å).	51
3.13	Experimental (green lines) and FMO/FDD calculated (purple lines) Raman spectra of Slr1393-g3 in the Pg state. The experimental spectrum was taken from the Ref.[102]	53
3.14	QM/MM (green lines) and FMO/FDD calculated (purple lines) Raman spectra of Slr1393-g3 in the Pg state. The QM/MM spectrum was provided by Giovanni Battocchio[102]	53
4.1	Potential energy scans for model compounds. The torsional potential were scanned in steps of 5° . The red curves represent the scans derived at the MP2/6-31G(d) level of theory, and the blue curves denote result from optimized torsion parameters (see Table).	59
4.2	Potential energy scans for model compounds. The torsional potential were scanned in steps of 5° . The red curves represent the scans derived at the MP2/6-31G(d) level of theory, and the blue curves denote result from optimized torsion parameters (see Table).	60
4.3	Optimized fragments MP2/6-31G(d) (color blue) and empirical MM structures (color green) of the BV chromophore without propionic side chains. The rmsd values were calculated for heavy atoms (no hydrogen atoms) after optimal superposition.	61
4.4	Left: root-mean-square deviation (RMSD) of backbone atoms for Agp2 phytochrome structure compared to the crystal structure (Pfr state) after 50ns. Right: RMSD of heavy atoms for the BV molecule compared to the crystal structure after.	63
4.5	Left: root-mean-square deviation (RMSD) of backbone atoms for Agp2 phytochrome structure compared to the crystal structure (Meta-F state) after 50ns. Right: RMSD of heavy atoms for the BV molecule compared to the crystal structure after.	64

4.6	Left: Chromophore binding pocket (Agp2 in Pfr state). Crystal structure (color gray) superimposed with average structures after 50 ns MD simulation. Right: Chromophore binding pocket (Agp2 in Meta-F state). Crystal structure (color gray) superimposed with average structures after 50 ns MD simulation.	64
5.1	Three MD trajectories of the Agp2 phytochrome with propionic side chain (attached to ring C) deprotonated and protonated.	70
5.2	Root-mean-square deviation (RMSD) of backbone atoms compared to the Agp2 phytochrome crystal structure (Pfr state) after 50ns.	71
5.3	Root-mean-square deviation (RMSD) of heavy atoms of the BV molecule of Agp2 in Pfr state compared to the crystal structure after 50ns.	71
5.4	Root-mean-square deviation (RMSD) of backbone atoms of Agp2 phytochrome compared to the crystal structure (Meta-F state) after 50ns.	72
5.5	Root-mean-square deviation (RMSD) of heavy atoms of the BV molecule of Agp2 (Meta-F state) compared to the crystal structure after 50ns.	72
5.6	Time evolution (50ns) of computed pKa values of pscC for Agp2 in Pfr state. Three MD trajectories with different pscC protonation were used for evaluation with Karlsberg2+ program.	73
5.7	Time evolution (the last 10ns) of computed pKa values of pscC for Agp2 in Meta-F state. Three MD trajectories with different pscC protonation were used for evaluation with Karlsberg2+ program.	73
5.8	Left: Chromophore binding pocket (Agp2 in Pfr state) after 50 ns MD simulation. Right: Chromophore binding pocket (Agp2 in Meta-F state) after 50 ns MD simulation.	74
5.9	Time evolution (the last 20ns) of computed pKa values of pscC for Agp2 in Pfr state using classical MD simulations and Karlsberg2+ program. The “md-pra-h1-cw” trajectory includes crystal waters and the “md-pra-h1-cwd” trajectory includes crystal and Dowser waters.	75
5.10	Left: Chromophore binding pocket and crystal water molecules of Agp2 in the Pfr state, for this MD frame the pscC has a pKa=-10, Right: Chromophore binding pocket including crystal water and Dowser water molecules of Agp2 in the Pfr state, for this MD frame the pscC has a pKa=6.1.	76

List of Tables

2.1	Selected structural parameters of BV molecule in DrBphP-Hip, DrBphP-Hid and DrBphP-Hie models. Bond lengths are given in Å while bond angles, torsional angles and tilt angles between rings in degree (°).	25
2.2	Root Mean Square Deviations (Å) of the position of heavy atoms of selected residues located in CBP relative to crystal structure (PDB entry 4q0h) after alignment of the heavy atoms of the BV moieties.	25
2.3	Hydrogen-bond Network in CBP involving the BV chromophore, pyrrole water, His260 and Asp207. Hydrogen donor - acceptor distances (Å) are defined in Figure 2.4a.	27
3.1	BV chromophore and protein structural properties for Agp2-BV, Agp2-BVM-B, and Agp2-BV-PAiRFP2[77, 89, 90]	40
4.1	Interaction Energies (kcal/mol) and Distances (Å) between water and fragment RDV Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).	57
4.2	Interaction Energies (kcal/mol) and Distances (Å) between water and fragment RCD Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).	57
4.3	Interaction Energies (kcal/mol) and Distances (Å) between water and fragment RBC Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).	57
4.4	Interaction Energies (kcal/mol) and Distances (Å) between water and fragment RAS Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).	58
4.5	Force constants (k), multiplicity (n) and phase (δ) of dihedral parameters for model compounds.	58
4.6	Force constants (k), multiplicity (n) and phase (δ) of dihedral parameters for model compounds.	61
4.7	Atom Types, Labels and Optimized Partial Charges of the BV Molecule	62
5.1	pKa values calculated for pscC and some residues of the chromophore binding pocket with Karlsberg2+ in combination with classical MD simulations for the Agp2 in Pfr and Meta-F state.	73
5.2	Mean pKa values and population distribution after 100 ns for the Agp2 in both Pfr and Meta-F states.	77

1

Introduction

1.1 Biological systems

1.1.1 Phytochromes

Phytochromes are red-light dimeric photoreceptors found in plants, bacteria and fungi[1, 2, 3]. These photoreceptors are responsible for various processes such as seed germination, flowering time, neighbor perception and avoidance, and phototaxis. Phytochromes use a linear methine-bridged tetrapyrrole chromophore for light absorption. The chromophore allows photoconversion between red-absorbing (Pr) and far-red absorbing (Pfr) states and is found partially buried in the interior of the protein structure forming the chromophore-binding pocket (CBP). In the particular case of bacterial phytochromes from *Deinococcus radiodurans* (DrBphP) and *Agrobacterium fabrum* (Agp2), the bilin chromophore consists of a fully protonated biliverdin (BV) molecule which is covalently attached to the protein matrix through a thioether linkage at Cys24 [4]. It is important to mention that the crystal structure of DrBphP was the first phytochrome structure ever solved[5, 6] and has been used as structural model for investigating the behaviour and function of phytochromes deriving from many other organisms[2, 3].

Phytochromes are composed of two functional domains, the N-terminal light-sensing domain and the a C-terminal signaling domain[7, 8]. The C-terminal portion commonly consist of a histidine kinase (HK) domain, whilst the N-terminal domain is responsible for photoperception and is composed of three domains: PAS (Per, Arnt, Sim), GAF (cGMP-phosphodiesterase, adenylate cyclase, FhlA) and PHY (phytochrome specific) (see Figure 1.1). The bilin chromophore binds the PAS domain but is found buried in the GAF domain, the polypeptide chain comprising these two domains is named as the chromophore-binding domain (CBD) (see Figure 1.2).

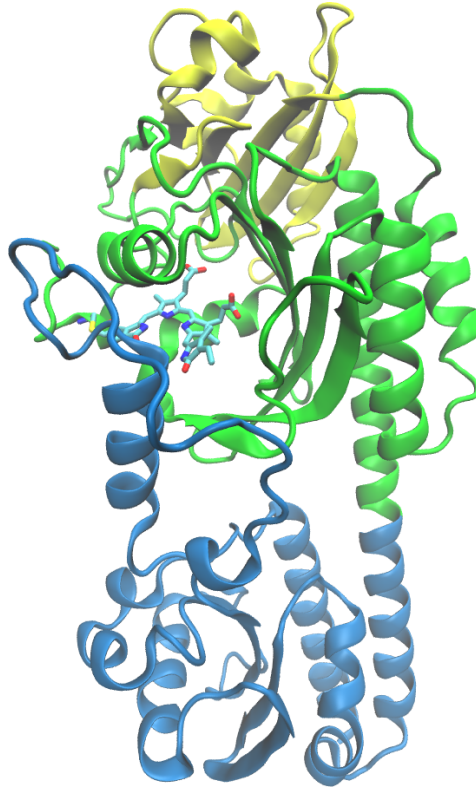


Figure 1.1: Photosensory core module (PCM) of Agp2 phytochrome structure divided into PAS (color yellow), GAF (color green) and PHY (color blue) domains, (PDB entry 6G1Y, chain A)[9]

1.1.2 Chromophore structure

All known phytochromes contains a linear methine-bridged tetrapyrrole chromophore in the N-terminal domain. The chemical structure of the tetrapyrrole chromophore can vary depending the type of phytochrome. In plant phytochromes, a P Φ B is the molecule responsible for light absorption. whilst in phytochromes of bacterial and fungal organisms, the cofactor corresponds to the phycocyanobilin (PCB) and biliverdin (BV)[2, 10]. Figure 1.3 shows the four pyrrole rings of the chromophore labeled from A to D and two propionic side chains attached to the pyrrole rings B and D. Commonly, the chemical substitutions in the rings A and D may include a vinyl group (BV and P Φ B) and an ethyl group (PCB).

In the particular case of BV molecule, the binding to the protein matrix can lead to different chemical structures (BV-like or P Φ B-like) as shown in Figure 1.3. Usually, all four nitrogens of phytochrome cofactors are protonated in the Pr and Pfr state leading to a positive charge (+1) delocalized over the inner pyrrole rings, in contrast, the two propionic side chains have each a negative charge (-1). However, in bathy phytochromes, recently was shown that the propionic side chain of ring C is protonated in Pfr state even above pH 7.8, which is a common structural element of the chromophore-binding pocket in bathy phytochromes[11, 12, 13]. Regarding the conformation/configuration of the protein-bound cofactor, the double bond of the three methine bridges between rings A-B, B-C and C-D are normally found in Z (Zusammen) and E (Entgegen) configurations. Whereas the conformational states of each methine bridge single bond are denoted as s (syn) and a (anti)[14]. For example, the BV molecule of the Agp2 phytochrome shows a ZZE_{ssa} configuration in the Pfr state, whilst the cofactor of the DrBphP phytochrome in the Pr state shows a ZZZ_{ssa} configuration[9].

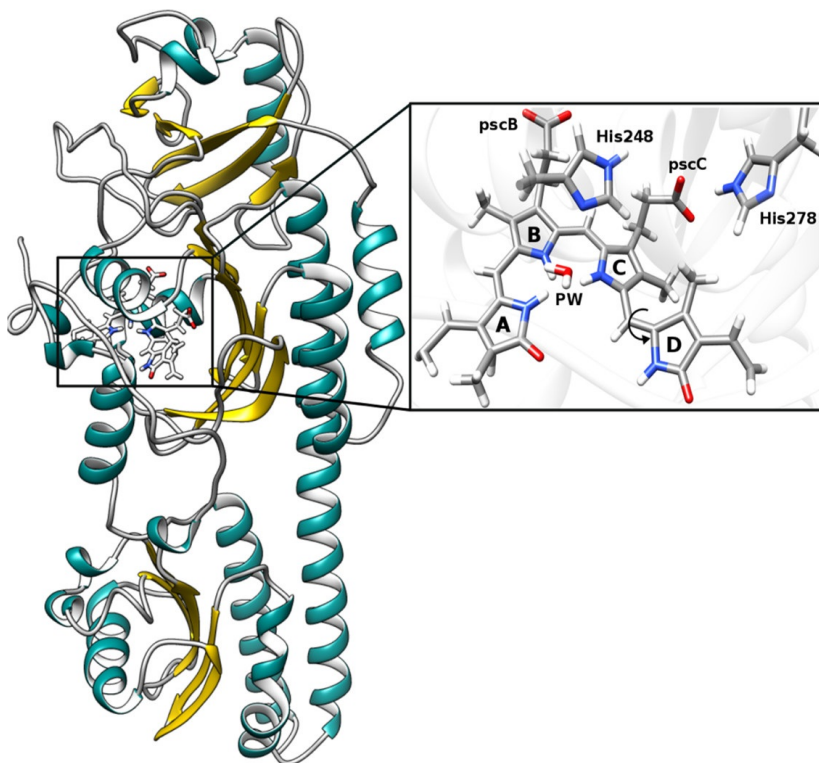


Figure 1.2: left: Photosensory core module of Agp2 phytochrome structure, right: pyrrole water, two highly conserved amino acid residues and BV molecule in the GAF domain.

In general, the photoconversion process[15, 16, 17] (see Figure 1.4) from the red-absorbing form Pr to the far-red-absorbing form Pfr is initiated by a Z to E isomerization around the C-D methine bridge (C15-C16) of the cofactor. This process involves the formation of at least two intermediate states, called Lumi-R and Meta-R, and depending of the species this transformations can take 1 to 1000 ms. The photochemical reaction from Pfr to Pr state proceeds via Lumi-F and Meta-F states. Interestingly, several studies have revealed a secondary structural change of the "tongue" in the PHY domain[11, 18, 19] (a β -sheet-to- α -helix transition) in most of phytochromes.

1.2 Electronic structure methods

The main goal of the modern electronic structure methods is the calculation of different properties of molecules such as their total energy, dipole moments, potential energy surfaces, vibrational frequencies, excitation energies, and many others. However, to achieve this aim, the many-electron Schrödinger equation must be solved within the Born-Oppenheimer approximation[20] in an efficient way. To this end, a great number of approximate methods have been proposed in order to obtain the best balance between accuracy and computational cost. Overall, the electronic structure methods can be divided into ab initio and semiempirical methods. In this work, we will focus on the Hartree-Fock (HF) (ab initio) and SCC-DFTB (semiempirical) methods[21, 22] within the fragment molecular orbital (FMO) scheme[23, 24].

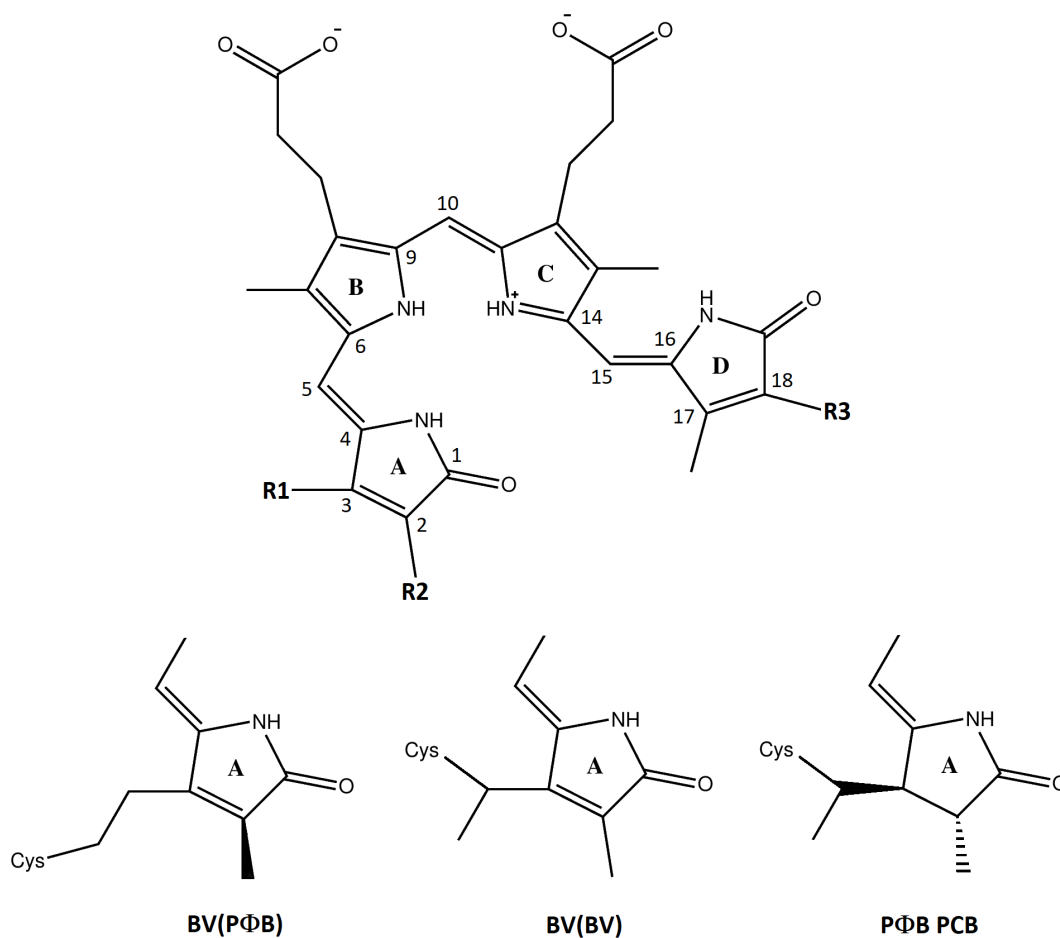


Figure 1.3: a) Chemical structure of the tetrapyrrole chromophore and chemical substitutions located at the rings A and D, with R3 equal either a vinyl group (BV and (PΦB)) or an ethyl group (PCB). b) Three different forms of binding between ring A and protein matrix.

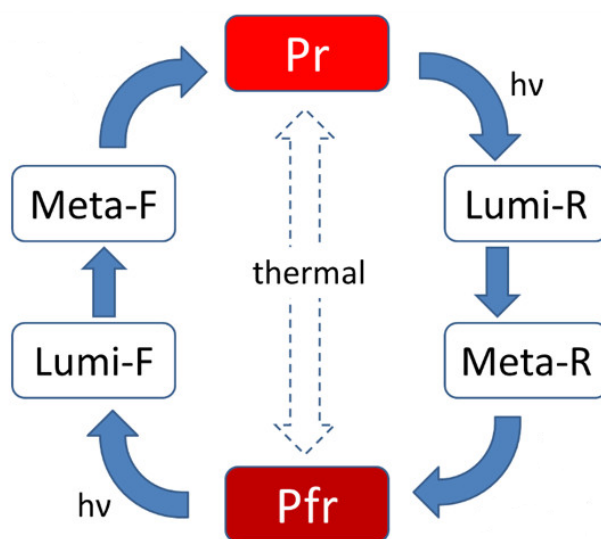


Figure 1.4: Generic phytochrome photocycle with the red light absorbing parent state (Pr) and far-red light absorbing parent state (Pfr).

1.2.1 The Self-Consistent Charge Density Functional Tight Binding formalism (SCC-DFTB)

The self-consistent charge density functional tight-binding (DFTB) theory[22] is based on density functional (DFT) theory and is a useful tool for performing electronic structure calculations of large molecular complex systems. One of the most important advantages of the SCC-DFTB method is the possibility of using parameters to avoid the calculations of one and two-electron integral expressions. Furthermore, the calculations can include only the valence electrons explicitly, as a result, the DFTB calculations can be speedup about three orders of magnitude compared to DFT. Moreover, the combination of DFTB with the FMO method (FMO-DFTB) has demonstrated to be capable of performing geometry optimizations for molecular systems containing up to one million atoms using modest computational resources[24].

The total energy of NCC-DFTB, E^{NCC} is given by:

$$E^{NCC} = \sum_i \sum_{\mu\nu} n_i c_{\mu i}^* c_{\nu i} H_{\mu\nu}^0 + \sum_{A>B} E_{AB}^{rep}, \quad (1.1)$$

where E^{NCC} corresponds to the nonself-consistent-charge (NCC) term (sometimes called ‘‘DFTB1’’). the indexes μ and ν denote atomic orbitals (AO), A and B denote atoms, and n_i is the occupation number of the i -th molecular orbital (MO). $c_{\mu i}$ are the expansion coefficients of the i -th MO in an AO basis. In eq.1.1, the $H_{\mu\nu}^0$ term corresponds to the nonperturbed Hamiltonian in the basis of optimized AOs. E_{AB}^{rep} is defined as the two-body repulsive energy term, which is independent of the electronic structure and can be precomputed and tabulated as a function of interatomic distance R_{AB} using DFT calculations of model systems.

The total energy of SCC-DFTB, E^{SCC} can be defined as:

$$E^{SCC} = E^{NCC} + \frac{1}{2} \sum_{AB} \gamma_{AB} \Delta q_A \Delta q_B, \quad (1.2)$$

where the total energy E^{SCC} is called DFTB2, the term γ_{AB} depends on the distance R_{AB} between two atoms A and B and their chemical hardness. The induced charge on each atom A is denoted as Δq_A . Thus, the charge self-consistent Hamiltonian is given by:

$$H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu}^X \sum_C (\gamma_{AC} + \gamma_{BC}) \Delta q_C, \quad (1.3)$$

The overlap matrix elements $S_{\mu\nu}$ can also be precalculated using the DFT and tabulated.

1.3 The fragment molecular orbital (FMO) method

The FMO method[25, 23] poses as a promising tool for studying large molecular systems by means of efficient parallel computations. In this approach, the system under consideration is divided into fragments (*monomers*). Each fragment calculation involves an *ab initio* calculation in the presence of the electrostatic field due to the other fragments; monomer densities converge self-consistently in an iterative procedure. Exchange interactions and correlation effects between

1. Introduction

fragments can be included by calculations on *dimers*—i.e. pairs of fragments—and, optionally, *trimers*. The total electronic energy within the two-body FMO expansion (FMO2) of a system containing N fragments can be expressed as:

$$E^{FMO2} = \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J), \quad (1.4)$$

where E_I is the energy of monomer I immersed in the external electrostatic potential generated by the remaining monomers; E_{IJ} is the interaction energy of dimer IJ , which is also immersed in the external electrostatic potential of the other fragments (See Figure 1.5). In order to improve the accuracy it is possible to define the three-body FMO expansion, also known as FMO3:

$$E^{FMO3} = E^{FMO2} + \sum_{I>J>K}^N \Delta E_{IJK}, \quad (1.5)$$

The above expression includes the triple corrections that can be calculated from trimer energies E_{IJK} . In the FMO method, monomer energies can be obtained by solving the corresponding *ab initio* equations, where the mono-electronic Hamiltonian $\mathbf{H}$ is modified by the inclusion of two additional terms, the electrostatic potential $\mathbf{V}$ and the projection operator $\mathbf{P}$:

$$\tilde{\mathbf{H}}_{\mu\nu}^X = H_{\mu\nu}^X + V_{\mu\nu}^X + P_{\mu\nu}^X, \quad (1.6)$$

where $H_{\mu\nu}^X$ is the standard one-particle matrix for electrons in monomer (dimer or trimer) X . It includes the kinetic energy of electrons and its interaction with the classical particles in monomer X . The second term in Eq. (1.6) gives the external electrostatic potential due to the remaining monomers, which is defined as:

$$V_{\mu\nu}^X = \sum_{K(K \neq X)}^N \left\{ \sum_{A \in K}^{N_c^K} \langle \phi_\mu | \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} | \phi_\nu \rangle + \sum_{\sigma\lambda \in K} D_{\sigma\lambda}^K \langle \phi_\mu \phi_\sigma | \phi_\nu \phi_\lambda \rangle \right\}. \quad (1.7)$$

The first term in Eq. (1.7) is the attraction between the electron density of monomer X and the nuclei of fragments $K \neq X$ and the second term corresponds to electron density-density repulsion. Index K is summed over all monomers N , except X , indexes μ, ν, σ and λ are summed over all basis functions of electrons in monomer X .

In Eq. (1.7), Z_A and $\mathbf{R}_A$ are classical nuclei charges and coordinates, respectively. $\mathbf{D}$ is the density matrix of monomer K . The third term in Eq. (1.6) is only included when covalent bonds are fragmented, for example, for studying biomolecules, there, B is a positive constant equal to 10^6 au and $P_{\mu\nu}^i = \langle \phi_\mu^\alpha | \phi_i^h \rangle \langle \phi_i^h | \phi_\nu^\alpha \rangle$ is a projection operator which is used to locate electronic molecular orbitals, where ϕ_i^h are hybridized electronic orbitals placed on the atoms involved in the fragmented covalent bonds.

In the FMO method, covalent bonds can be fragmented without using capping approaches at the cutting point[26]. Figure 1.6 shows the breaking of the C-C covalent bond in the ethane molecule C_2H_6 , this molecule can be divided into two CH_3 methyl groups, however, one could apply the simplest scheme fragmentation, that is, assigning the electron pair of the fractioned bond to one fragment, thus, one methyl group gets 10 electrons and the other one 8 electrons

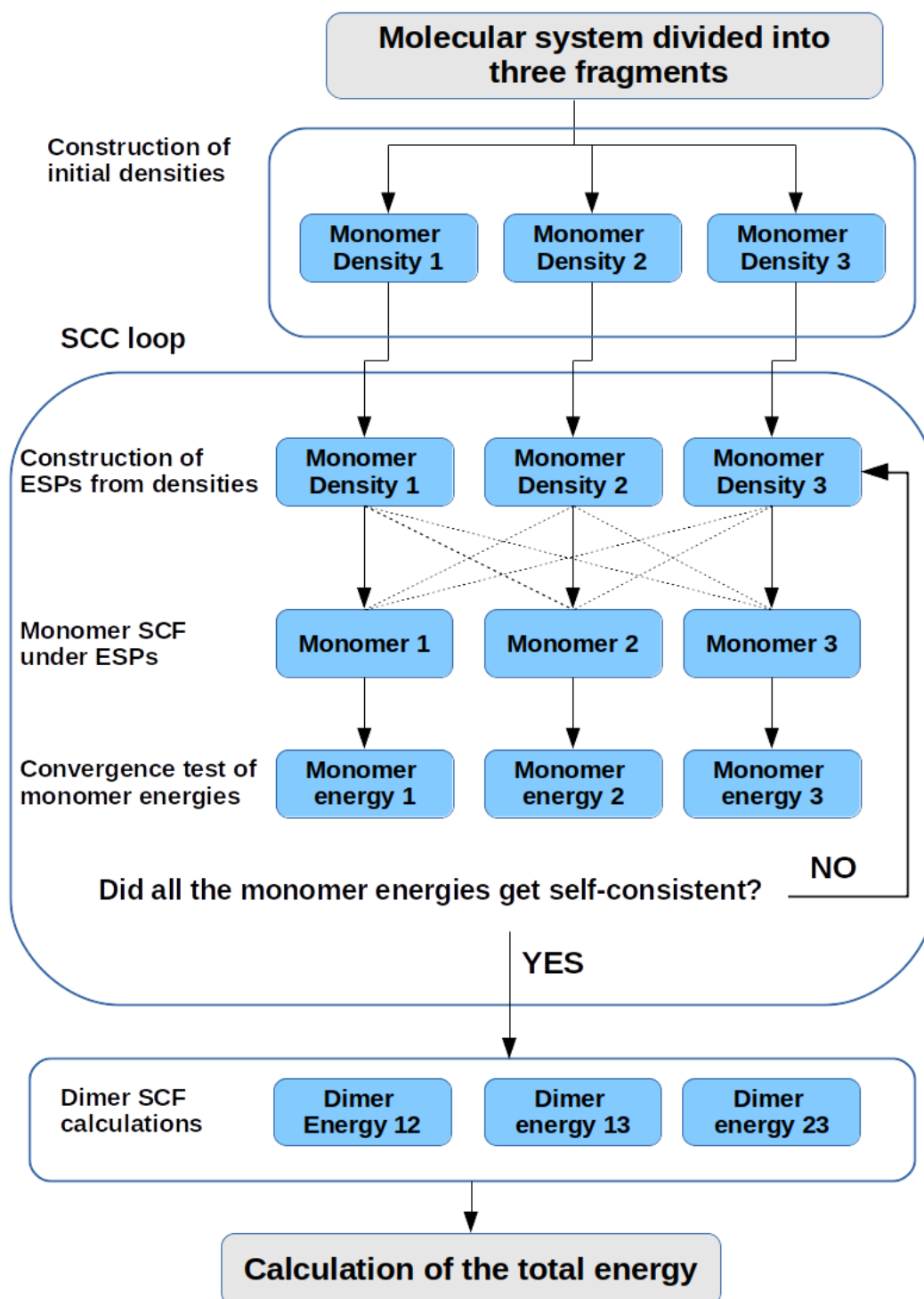


Figure 1.5: General procedure for calculating the energy of a molecular system divided into three fragments. In the first step, the monomer densities are built. Then, a self-consistent charge (SCC) procedure is applied to the monomer densities by which the energy of each monomer is solved by means of self-consistent field (SCF) calculations including the electrostatic potential (ESP) generated by the remaining fragments (dash lines). In the third step, fragment pair SCF calculations are performed (i.e., the inter-fragment interaction energy) and, in the last step, the total energy of the system is obtained.

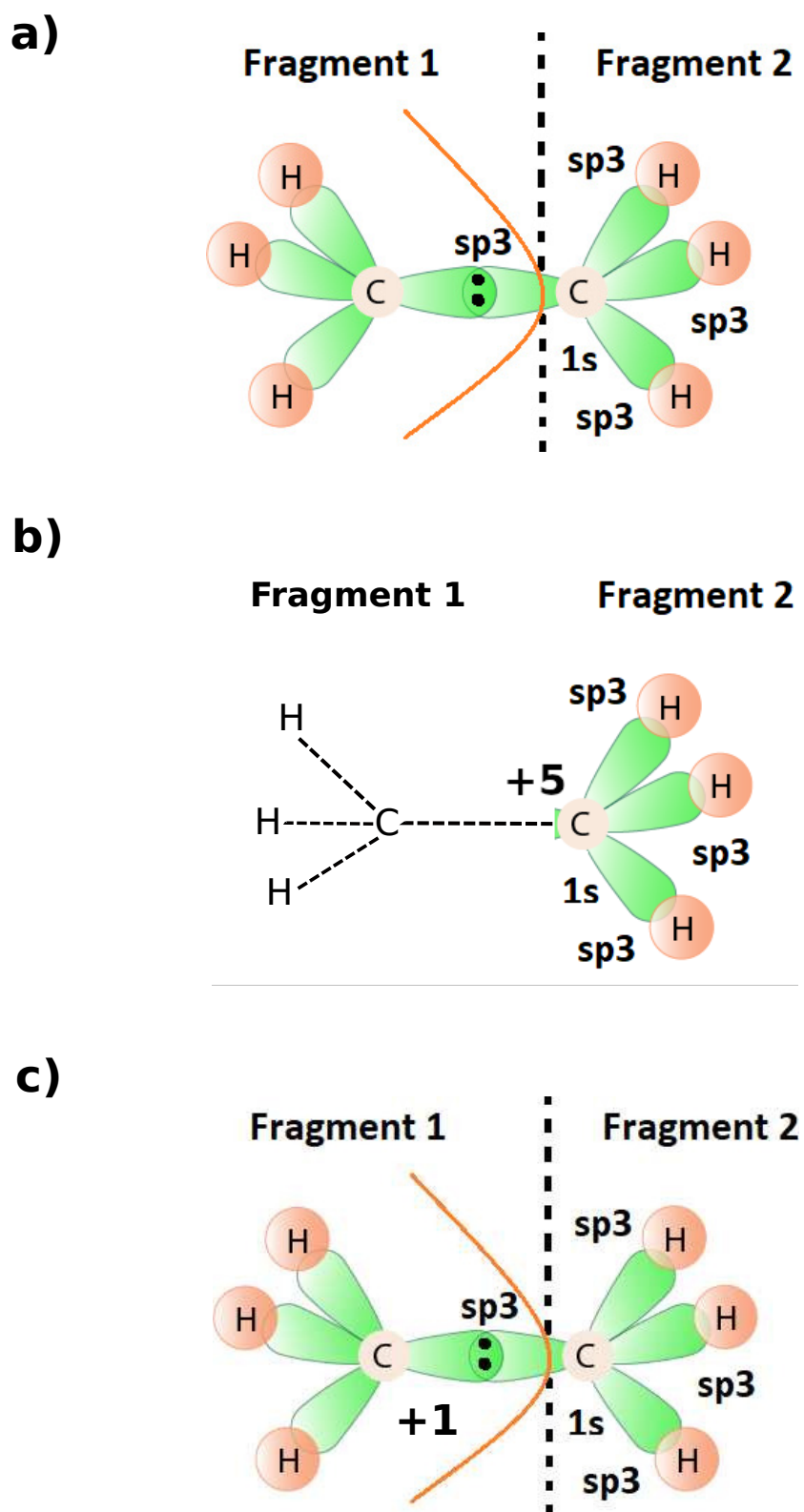


Figure 1.6: Covalent bond fragmented within the FMO method for ethane molecule C_2H_6 ,

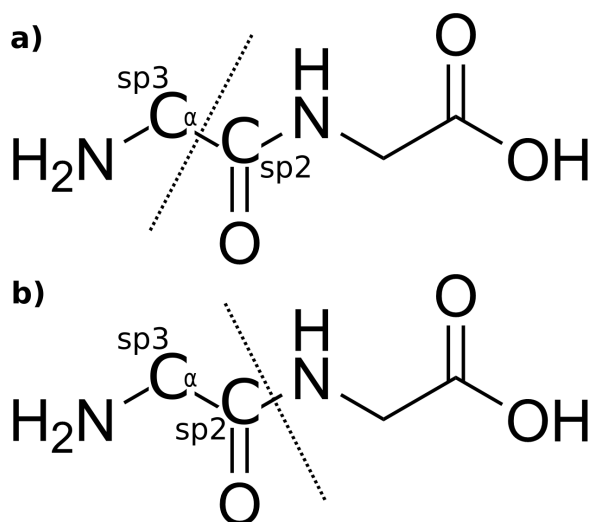


Figure 1.7: Fragmentation of a dipeptide in the FMO method. (a) The molecule is fragmented at $C_{\alpha}(sp^3)$ carbon atom. (b) The molecule is fragmented at the sp^2 carbon atom.

leading to fragments with formal charges of -1 and +1, respectively. Therefore, in order to avoid an attraction of charged electron densities between both fragments, it is necessary to include an additional step. This involves the reassignment of one proton and one electron from the C atom (fragment 2) to the fragment 1, thus, the fragment 1 retains ten electrons and ten protons, whilst the fragment 2 keeps eight electrons and eight protons. Furthermore, a quasiautom C has to be added to fragment 1, this quasiautom C contains carbon basis functions and an atomic charge of +1, while C atom in fragment 2 has a charge of +5 with the carbon basis functions.

In the last step, hybrid orbitals can be used to divide the atomic orbital space. For instance, for carbon atoms it is possible to define sp^3 hybridized orbitals. Thus, the fragment 1 only needs to maintain the latter sp^3 orbital, and the other four orbitals (three sp^3 and core 1s) projected out, in contrast, the fragment 2 retains four orbitals (see Figure 1.6b) and only one orbital projected out. Therefore, the fragmentation scheme in the FMO method leads to the preservation of charges and multiplicities. Recently, Akinaga et al.[27] proposed a fragmentation scheme that allows the fragmentation at sp^2 carbon atoms in FMO method (see Figure 1.7), for proteins, this will be very helpful for doing a natural fragmentation, that is, one amino acid residue per fragment, by doing this, the definition of residues and fragments in the FMO method will be the same, it is important to mention that currently in the FMO method the fragmentation of a protein is performed at C_{α} atoms, this means fragments in FMO are different (by a carbonyl $C=O$ group) regarding conventional residues.

1.4 Pair interaction energy decomposition analysis (PIEDA)

One of the most interesting features of the FMO method is the pair interaction energies (PIEs) approach, which has been applied mostly to the study of protein-ligand, protein-DNA and protein-protein interactions, the PIE between two fragments can be calculated by:

$$\Delta E_{IJ}^{int} = E_{IJ} - E_I - E_J \quad (1.8)$$

The ΔE_{IJ}^{int} can be decomposed into five energy terms: electrostatics (ΔE^{es}), exchange-repulsion (ΔE^{ex}), charge transfer (ΔE^{ct}), dispersion (ΔE^{di}) and solvent screening (ΔE^{solv}):

$$\Delta E_{IJ}^{int} = \Delta E_{IJ}^{es} + \Delta E_{IJ}^{ex} + \Delta E_{IJ}^{ct} + \Delta E_{IJ}^{di} + \Delta E_{IJ}^{solv} \quad (1.9)$$

This method is called PIE decomposition analysis (PIEDA)[28], therefore, by using PIEDA it is possible to obtain valuable information into the chemical nature of non-covalent interactions between proteins and ligands. Non-covalent interactions like salt bridges, hydrogen bonds, or polar interactions are dominated by the electrostatic and charge transfer terms while hydrophobic interactions are dominated by the dispersion term[29]. MP2 (2nd order Møller-Plesset perturbation theory)[30] calculations are additionally performed for evaluating dispersion interactions. Solvent effects are included via the polarizable continuum model (PCM)[31] (solvent screening energy term ΔE_{IJ}^{solv}).

1.5 Computations of Raman spectra using the fragment molecular orbital method

Simulations of IR and Raman spectra of biomolecules can be performed by using the FMO method within the frozen domain with dimers (FDD) approach[32, 33]. FMO/FDD is a multilayer scheme that allows the combination of two different level of theories by using two layers, L_1 and L_2 . In the first step, monomer energies (FMO1) are obtained for the entire system S, this step is necessary for getting the fragment densities and is carried out at the level L_1 . Dimer energy computations (FMO2) of the polarizable domain B are done at the level of theory corresponding to L_2 layer including the contributions of the fragments in F and B regions. However, fragments belong to the F region are frozen whilst fragments that belong to the B region are updated for each structure during the geometry optimization. In FMO/FDD, only the fragments that belong to the active region A are allowed to move, this means, only these coordinates will change during the optimization. In the FMO/FDD approach, the total energy of a molecular system is defined as:

$$E^{FMO/FDD} = \sum_{I \in B} E_I'^{L_2} + \sum_{\substack{I > J \\ I \in A, J \in B}} \Delta E_{IJ}^{L_2} + \sum_{I \in A, J \in F} \Delta E_{IJ}'^{L_2, L_1} \quad (1.10)$$

where $E_X'^{L_2}$ ($X = I, J$, or IJ) corresponds to the internal fragment energy of X, that is, the RHF energy of X without the contribution of the embedding energy. $\Delta E_{IJ}^{L_2}$ is the RHF dimer contribution for two fragments (I,J) that belong to the L_2 . The last term ($\Delta E_{IJ}'^{L_2, L_1}$) corresponds to the electrostatic interaction between fragments I and J that are far each other. Applying the second derivative to the equation 1.10 with respect to nuclear coordinates a and b, one can obtain the analytic Hessian in the FMO/FDD method, which is given by,

$$\frac{\partial^2 E^{FMO/FDD}}{\partial a \partial b} = \sum_{I \in B} \frac{\partial^2 E_I'^{L_2}}{\partial a \partial b} + \sum_{\substack{I > J \\ I \in A, J \in B}} \frac{\partial^2 \Delta E_{IJ}^{L_2}}{\partial a \partial b} + \sum_{I \in A, J \in F} \frac{\partial^2 \Delta E_{IJ}'^{L_2, L_1}}{\partial a \partial b} \quad (1.11)$$

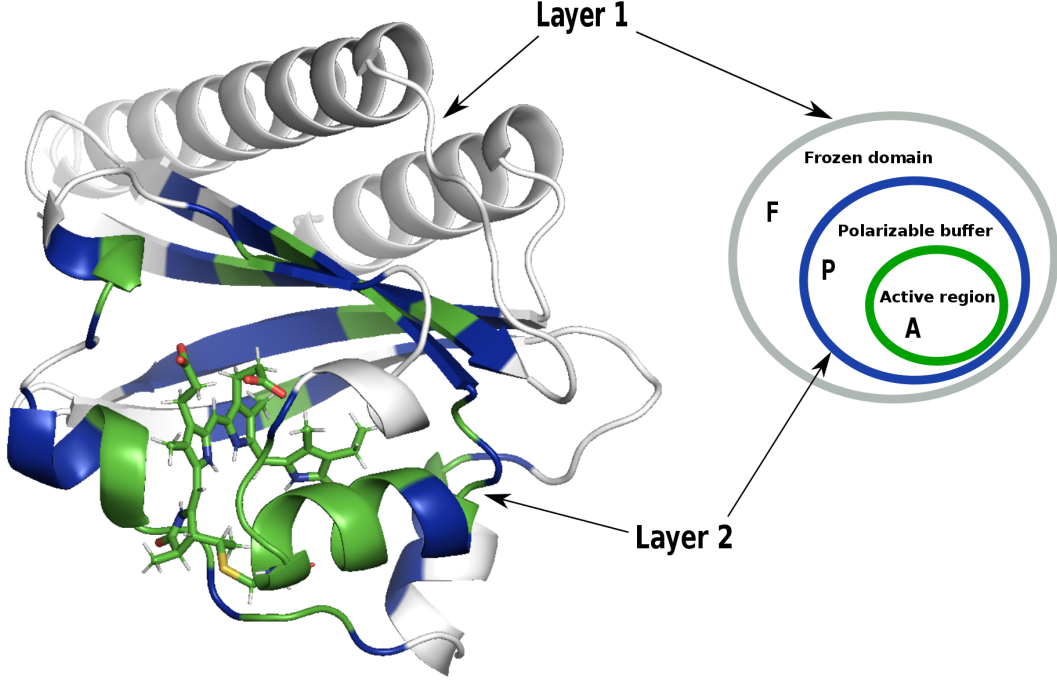


Figure 1.8: Phytochrome structure divided into two layers (L1, and L2). The first step in FMO/FDD is the computation of the whole system at the FMO1 level (monomers) for the initial geometry; then, the electronic state of fragments in domain F is frozen for the other geometries during a geometry optimization. Fragments that belong to the polarizable and active region share the same basis set, nevertheless, only fragments that belong to the active region are allowed to move during the optimization.

The normal Raman activity[34, 35] of a vibrational mode i can be calculated in the FMO/FDD method by means of:

$$J_i = 45\alpha_i'^2 + 7\gamma_i'^2 \quad (1.12)$$

In the equation 1.12 the symmetric contribution α_i' is defined as:

$$\alpha_i' = \frac{1}{3} \left[\left(\frac{\partial \alpha_{xx}}{\partial Q_i} \right) + \left(\frac{\partial \alpha_{yy}}{\partial Q_i} \right) + \left(\frac{\partial \alpha_{zz}}{\partial Q_i} \right) \right] \quad (1.13)$$

Whilst the anisotropic derivative contribution γ_i' is defined as:

$$\gamma_i'^2 = \frac{1}{2} \left[\left(\frac{\partial \alpha_{xx}}{\partial Q_i} - \frac{\partial \alpha_{yy}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{yy}}{\partial Q_i} - \frac{\partial \alpha_{zz}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{zz}}{\partial Q_i} - \frac{\partial \alpha_{xx}}{\partial Q_i} \right)^2 + 6 \left\{ \left(\frac{\partial \alpha_{xy}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{yz}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{zx}}{\partial Q_i} \right)^2 \right\} \right] \quad (1.14)$$

The polarizability tensor α_{xy} corresponds to the second derivative of the energy with respect to the external electric fields $\mathcal{F}_x$ and $\mathcal{F}_y$ and is given by:

$$\alpha_{xy} = \frac{\partial^2 E}{\partial \mathcal{F}_x \partial \mathcal{F}_y} \quad (1.15)$$

On the other hand, $\partial\alpha_{xy}/\partial Q_i$ is the derivative of the polarizability tensor with respect to a normal coordinate Q_i , which is a linear combination of atomic coordinates.

$$Q_i = \sum_{a=1}^{3M} l_{ai} a \quad (1.16)$$

In equation 1.16 the letter a corresponds to a Cartesian coordinate of an atom, where M is the number of atoms and l_{ai} is the eigenvector of the Hessian for the normal mode i . The polarizability tensor derivatives $\partial\alpha_{xy}/\partial Q_i$ can be obtained from the Cartesian derivative $\partial\alpha_{xy}/\partial a$ as

$$\frac{\partial\alpha_{xy}}{\partial Q_i} = \sum_{a=1}^{3M} l_{ai} \frac{\partial\alpha_{xy}}{\partial a} \quad (1.17)$$

Normal coordinates associated with vibrational frequencies can be obtained by calculating the second derivatives of the energy with respect to nuclear coordinates a and b and transform them into mass-weighted Cartesian coordinates using atomic masses m_a :

$$H'_{ab} = \frac{\partial^2 E}{\partial a \partial b} / \sqrt{m_a m_b} \quad (1.18)$$

The mass-weighted Hessian matrix H' can be diagonalized for obtaining the vibrational frequencies and the normal coordinates as its eigenvalues and eigenvectors, respectively.

The normal Raman activity calculation in FMO/FDD can be summarized as follows, in the first step, a geometry optimization needs to be done, then, this optimized geometry can be used for doing a calculation of the Hessian and subsequently, a normal vibrational analysis. It is important to mention that in the FMO/FDD method, the hessian will be calculated only for those atoms that belong to the active region A. Therefore, this approach will be very useful for vibrations highly localized in the active region.

1.6 Molecular dynamic simulations

Molecular dynamics (MD) simulations have become a tool well established for obtaining a deeper insight into the conformational variability of a protein as well as to investigate its dynamic properties in atomistic detail. These calculations can be performed using quantum mechanical (QM) or molecular mechanics (MM) as levels of theory. However, quantum MD simulations of proteins with thousands of atoms remain computationally expensive and as a result, these calculations often are restricted to systems containing a very small number of atoms. In contrast, MD simulations based on molecular mechanics using all-atom empirical force fields have enabled the study of molecular systems with even millions of atoms[36].

1.6.1 Non-polarizable force fields

The interesting ability of MD to investigate biomolecules largely depends on the quality of force fields parameters which must be optimized to describe interactions between particles. In classical MD simulations, these particles are treated as point charges with fixed atomic charges and they obey the laws of Newtonian mechanics, as a consequence, the electrons are

neglected and chemical reactions cannot be studied. Furthermore, most of the force fields do not take into account the polarization of the atoms, for example, due to the presence of the other atoms surrounding. Nevertheless, great efforts have been made to include the effect of the polarization or to study chemical reactions by using polarizable force fields[37, 38] and reactive force fields[39], respectively. In this work, we will focus on the non-polarizable force fields.

A typical force field can be divided into two different types of interactions: bonded and nonbonded interactions. The bonded interactions correspond to bond stretching, angle bending, bond rotation (torsion), and out-of-plane movements (improper torsion) as is shown in Figure X, while the nonbonded interactions contain the charge-charge and van-der-Waals interactions. The van-der-Waals interactions can be described by a Lenard-Jones potential. Thus, an expression of the potential energy function of one of the most widely used force fields (CHARMM)[40, 41] is defined by:

$$V = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(b - b_\theta)^2 + \sum_{dihedrals} k_\phi[1 + \cos(n_\phi - \delta)] + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_{UB}(S - S_0)^2 + \sum_{nonbonded} \epsilon \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_l r_{ij}} \quad (1.19)$$

In equation 1.19 the first four terms correspond to the bonded interactions whereas the last two term describes the nonbonded interactions, the fifth term is the Urey-Bradley expression which accounts for angle bending using 1,3 nonbonded interactions, it is important to mention that by definition, the nonbonded forces are only applied to atom pairs separated by at least three bonds. Usually, the optimization of dihedrals parameters is the most difficult part in the equation1.19, since this step requires the potential energy surfaces (PES) of the dihedrals angles involved in a molecular system by using QM computations.

1.7 Electrostatic energy calculations for computing pKa values

Proteins contain different titratable residues, which are essential for the understanding of inter-residues, protein-solvent, and protein-ligand interactions, furthermore, these residues play a key role in several proprieties of proteins, such as folding, conformational stability, and catalytic activity. Thus, the prediction of accurate pK_a values is the basis for understanding the pH-dependent properties of a protein.

1.7.1 Calculating pKa values in proteins

In general, the equilibrium of the dissociation of a weak acid (HA) can be defined by:



1. Introduction

and the dissociation constant is:

$$K_A = \frac{[H^+][A^-]}{[HA]} \quad (1.21)$$

At this point, it is very useful to be able to relate the pH of a weak acid solution with its dissociation constant. To that end, we can use the Henderson-Hasselbalch equation:

$$pK_a = pH - \log \left(\frac{[A^-]}{[HA]} \right) \quad (1.22)$$

Since, by definition $pH = -\log[H^+]$ and the $pK_a = -\log(K_a)$. Furthermore, the equation 1.22 can be also defined in terms of the protonation probability ρ of the acid HA as follows,

$$pK_a = pH - \log \left(\frac{1 - \rho}{\rho} \right) \quad (1.23)$$

and doing some transformations:

$$\rho = \frac{10^{(pK_a - pH)}}{1 + 10^{(pK_a - pH)}} \quad (1.24)$$

Finally, the Gibbs free energy for the deprotonation of the acid HA is defined by:

$$\Delta G(HA \rightarrow A) = -\ln(10) * RT * (pK_a - pH) \quad (1.25)$$

where R corresponds to the universal gas constant and T the absolute temperature.

It should be noted that by using the above equations we can only compute the pKa value of a titratable molecule in aqueous solution. Therefore, the calculation of the pKa value of a titratable group within a protein requires a different approach. This approach should take account the interactions between the titratable group and other charges in the protein, as well as changes in the dielectric environment of the titratable group that are produced by transfer it from aqueous solution to the protein[42, 43]. Accordingly, the Gibbs free energy can be written as:

$$\Delta G(HA \rightarrow A) = -\ln(10) * RT * (pK_a - pH) + \Delta G_p \quad (1.26)$$

In the equation 1.26 the ΔG_p term is the change in protonation free energy because of the presence of the protein. In general, the protonation states of titratable residues in a protein are pH dependent and as a result, a protonation state change of a residue can affect the protonation states of other residues in protein. Therefore, ΔG_p should be expressed as a function of protonation states of all titratable residues in protein. This function can be defined by using the vector $\mathbf{p}(pH)$. For example, one can use $\mathbf{p}_{ref}$ as a reference protonation vector, while $\mathbf{p}_i$ can be any protonation vector. Thus, the computation of the free energy needed for changing the protonation states of all titratable residues in protein at a certain pH value is expressed by the function $\Delta G_p(\mathbf{p}_{ref} \rightarrow \mathbf{p}_i, pH)$. One of the main advantages of this function is the possibility to compare the energies of different protonation patterns (protonation vectors)

and consequently, to determine the probability of all protonation states at a certain pH value. Therefore, the probability $\rho(\mathbf{p}_i)$ of a protonation pattern $\mathbf{p}_i$ at a certain pH value is:

$$\rho(\mathbf{p}_i, pH) = \frac{e^{-\frac{\Delta G_p(\mathbf{p}_i, pH)}{k_b T}}}{\sum_j N_p e^{-\frac{\Delta G_p(\mathbf{p}_j, pH)}{k_b T}}} \quad (1.27)$$

In equation 1.27 k_b corresponds to the Boltzmann factor, T the absolute temperature and N_p the total number of possible protonation states. However, due to N_p is too large for evaluating the $\rho(\mathbf{p}_i)$ using the equation 1.27, even for very small proteins, a Metropolis-Monte-Carlo algorithm is needed for evaluating probabilities. This algorithm has been implemented in the software Karlsberg[44, 45, 46].

1.7.2 Electrostatic energy calculations

Commonly, the most significant types of interactions that determine pKa shifts in proteins are electrostatic interactions. Thus, an accurate description of the electrostatics in proteins is necessary for obtaining correct pKa shifts. According to the electrostatic theory, the force acting between two point charges q_1 and q_2 in vacuum separated by the distance r can be described with Coulomb's law:

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1 q_2}{|\vec{r}|^2} \frac{\vec{r}}{|\vec{r}|} \quad (1.28)$$

In equation 1.28, ϵ_0 describes the vacuum permittivity and $\vec{r}$ is the vector connecting the position of the two charges. The electrostatic potential of the q_1 is given by the Coulomb's potential:

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1}{|\vec{r}|} \quad (1.29)$$

And the electrostatic interaction energy of q_1 and q_2 is given by:

$$E_{elec} = q_2 \cdot \phi_1(\vec{r}) \quad (1.30)$$

It should be noted that a reasonably description of the electrostatics in a protein must include the Coulomb interactions between the charges that model the protein structure, as well as the interaction of the charges with the solvent. Therefore, a more general description of the electrostatic potential is needed for studying molecular systems like proteins solvated in water. The Poisson-Boltzmann equation (PBE) offers the possibility of including an inhomogeneous dielectric medium by using different dielectric constants for the protein and solvent. Furthermore, the PBE allows the inclusion of the ions represented as salt concentration in the solution.

1.7.3 The Poisson-Boltzmann equation

The electrostatic energy calculations for a protein solvated in pure water can be performed by using the Poisson equation as shown below:

$$\vec{\Delta}\varepsilon(\vec{r})\vec{\Delta}\phi(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon_0} \quad (1.31)$$

In the equation above, $\varepsilon(\vec{r})$ is a scalar function representing the dielectric constant. This function accounts the effect of the solvent around the protein by assigning a dielectric constant of $\varepsilon_w=80$. The choice of the dielectric constant inside the protein remains an open field of discussion, its value can vary in a range of $\varepsilon_p=4$ to $\varepsilon_p=20$. It has even been suggested that the dielectric constant should be inhomogeneous throughout the protein[47], for example, assigning a different dielectric constant for residues located at the interior of a protein ($\varepsilon_p=4$), while a higher value for those residues located on the protein surface ($\varepsilon_p=20$). Nevertheless, Aleksandrova et al.[48] recently showed that the inclusion of the electronic polarization effects by using the polarizable Drude force field led to a more realistic description of proteins for the prediction of pKa values. This means that a polarizable model is relatively insensitive to the choice of the internal dielectric constant in contrast to a non-polarizable force field. These findings highlight the importance of including a polarizable force field in the future computational studies of biomolecules.

In order to take into account the ions as certain salt concentration in the solution, the equation 1.31 can be modified by using the Debye-Hückel theory, resulting in the non-linear Poisson-Boltzmann Equation (PBE):

$$\vec{\Delta}\varepsilon(\vec{r})\vec{\Delta}\phi(\vec{r}) = -4\pi \left[\rho(\vec{r}) + k^2 \frac{kT}{e_c} v(\vec{r}) \sinh \left(\frac{e_c \phi(\vec{r})}{k_B T} \right) \right] \quad (1.32)$$

where k corresponds to the inverse Debye length

$$k = \sqrt{\frac{8\pi N_A e_c^2 I_s}{k_B T}} \quad (1.33)$$

The parameter $I_s = \frac{1}{2} \sum_i c_i z_i^2$ represents the ionic strength. e_c the elementary charge, N_A the Avogadro's number, T absolute temperature and the k_B the Boltzmann constant. The final expression of the PBE corresponds to the following linearized form:

$$\vec{\Delta}\varepsilon(\vec{r})\vec{\Delta}\phi(\vec{r}) + k^2 v(\vec{r})\phi(\vec{r}) = -4\pi\rho(\vec{r}) \quad (1.34)$$

One of the most interesting features of the linearized PBE equation 1.34 is the additivity of the obtained electrostatic potentials and energies. In this work, the software APBS[49] was used for solving the LPBE.

1.8 pKa calculations with constant pH molecular dynamics

Many biological processes, such as protein-ligand binding[50], enzyme catalysis[51], and pH-dependent protein folding[52] have commonly been studied using classical MD simulations, however, one of the most significant weaknesses of this type of approaches is that they require

to maintain fixed the protonation states of titratable sites during the simulation, which can lead to vague results. In order to address this issue, in recent years, the constant-pH molecular dynamics (CpHMD) based-methods have been used for studying many important biological processes caused by changes in solution pH. One of the main advantages of the CpHMD approaches is that the protonation states may change in the course of conformational dynamics. Thus, the dynamical processes coupled to a change in protonation states can be directly studied. In this work we will focus on the constant-pH MD algorithm implemented in the AMBER program[53].

The CpHMD method implemented in the AMBER program utilize a discrete protonation state model[54], this means that the protonation states of a protein are described by a vector $\mathbf{n} = (n_1, n_2, n_3, \dots, n_N)$ where n_i corresponds to an integer number representing the protonation state of titratable site i . In this model, the protonation states are sampled using a Monte Carlo method periodically during an MD simulation, after a certain number of MD steps a new protonation state is chosen. Then, the transition free energy for the protonated and deprotonated forms can be calculated by:

$$\Delta G = k_B T (pH - pK_{a,ref}) \ln(10) + \Delta G_{elec} - \Delta G_{elec,ref} \quad (1.35)$$

In the equation 1.35 k_B is the Boltzmann constant, T is temperature, $pK_{a,ref}$ is the pKa value of the reference compound, ΔG_{elec} is the electrostatic term of the free energy calculated with molecular mechanics for the titratable group, and $\Delta G_{elec,ref}$ is the electrostatic contribution of the transition free energy for the reference compound, this value can be obtained by using thermodynamic integration calculations, which will be treated as a internal parameter in AMBER[55, 56]. Nevertheless, quantum mechanical effects cannot be taken into account for the calculation of the free energy difference between the protonated and deprotonated forms of a titratable residue by using molecular mechanics, since this process involves the formation and breaking of covalent bonds. Therefore, in order to address this issue, a reference compound with known pKa can be used to cancel the nonelectrostatic contribution (QM part) of the transition free energy, assuming that the nonelectrostatic transition free energy value is approximately the same regardless of electrostatic environment. Based on the above, the division of the total transition free energy into the electrostatic and nonelectrostatic parts lead to equation 1.35.

In the last step, the ΔG term is used to decide by applying the Metropolis criterion if the transition will be accepted or rejected. For the case, where the transition is accepted, MD is performed with the titratable group in the new protonation state, if the transition is rejected, the MD is carried out with no modification of the protonation state.

1.8.1 Titration curve and pKa prediction calculation

The analytical form of the titration curve of a titratable residue can be obtained by using a more generalized form of the Henderson-Hasselbalch 1.22 (HH) equation:

$$pK_a = pH - n \log \left(\frac{[A^-]}{[HA]} \right) \quad (1.36)$$

1. Introduction

and by exponentiating both sides of the above equation, the titration curve of an interacting ionizable residue is given by:

$$s = \frac{1}{1 + 10^{n(pK_a - pH)}} \quad (1.37)$$

In the equation 1.37 s corresponds to the fraction of deprotonation and n is the Hill coefficient. The Hill plot is obtained by plotting $[A^-]/[HA]$ as a function of pH.

Fully Quantum Chemical Treatment of Chromophore-protein Interactions in Phytochromes

The project presented in this chapter led to the following publication:

González, R., and Mroginiski, M. A. (2019). Fully Quantum Chemical Treatment of Chromophore-Protein Interactions in Phytochromes. *The Journal of Physical Chemistry B*, 123(46), 9819-9830.

2.1 Introduction

In this work, the fragment molecular orbital (FMO) method was applied for optimizing the 3D structure of the CBD of *Deinococcus radiodurans* bacteriophytochrome (DrBphP) in the Pr state, (PDB entry 4q0h)[57] at a quantum mechanical level. The geometries of three structural models of DrBphP, each of which encompassing over 6000 atoms, were optimized using the FMO-DFTB3-D/PCM method[58]. Pair interaction energies (PIEs) calculations were applied for determining the most stable protonation state of the His260 using the three optimized structural models. Finally, PIE decomposition analysis (PIEDA) was used for identifying the nature and quantifying the strength of the non-covalent interactions between chromophore and protein matrix. To the best of our knowledge, the FMO method was applied for the first time, to refine at quantum mechanical level phytochrome structures. “Reproduced in part with permission from J. Phys. Chem. B 2019, 123, 46, 9819–9830, Copyright © 2019 American Chemical Society”

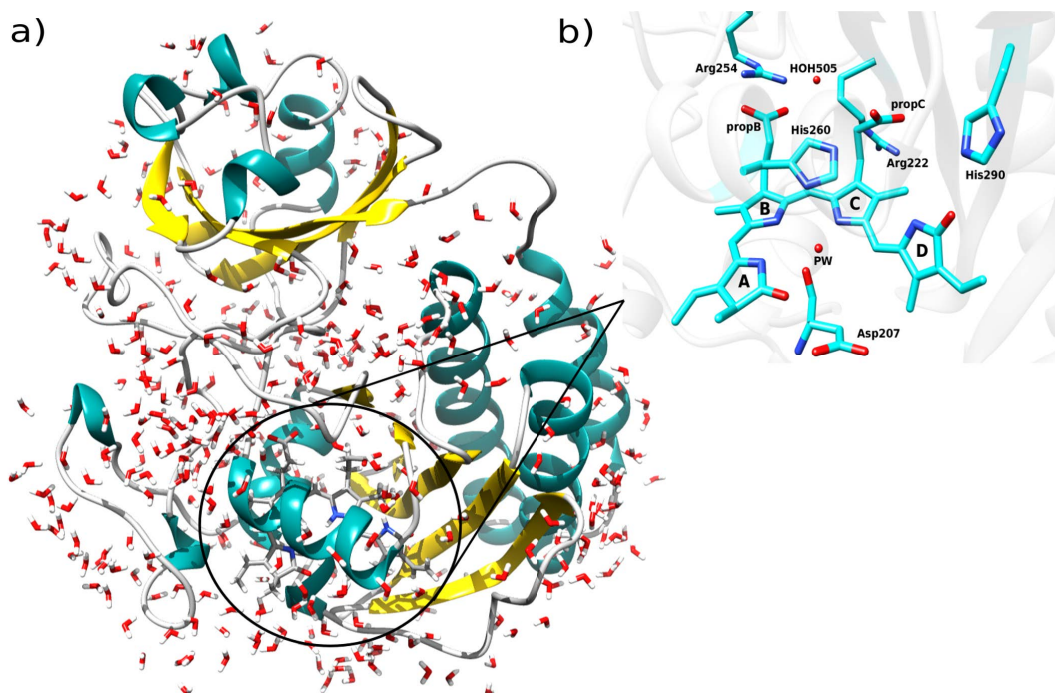


Figure 2.1: a) Crystal structure of the *Deinococcus radiodurans* BphP phytochrome in the Pr state (PDB entry 4q0h)[57] and b) chromophore binding pocket. The structural model contains all crystal water molecules, protein and BV chromophore (6159 atoms). Displayed images were created with UCSF Chimera software[59].

2.2 Methods

Structural models. The structural models addressed in this work were constructed using the crystal structure of the CBD of DrBphP in the Pr state (PDB entry 4q0h) solved at 1.157 Å resolution [57]. All crystal waters were explicitly included in the models (see Figure 2.1a).

The protonation states of all titratable residues at pH 7 were estimated by solving the linearized Poisson-Boltzmann equation of the electrostatic potential[42]. These computations were performed with the Karlsberg2+ software package[60] using the CHARMM36m force field[61] for the calculation of the electrostatic energy. Due to their close proximity to the BV chromophore, the protonation state and tautomeric form of His260 and His290 are evaluated with special care. In the case of the His290, its protonation state was investigated by Velazquez et al.[12] in Cph1 phytochrome and by Takiden et al.[62] in Agp1 phytochrome. In both spectroscopic/computational studies, it was concluded that His290 is protonated at the N ϵ , which is in good agreement with the protonation state predicted by Karlsberg2+ for DrBphB in the present work. This is however not the case for His260, for which contradicting assignments have been suggested. While Karlsberg2+ predicts a neutrally charged histidine holding a proton at the N ϵ in agreement with previous PROPKA calculations [63, 64], earlier spectroscopic studies on DrBphP favor the tautomer protonated at N δ (*vide supra*).[65, 63, 64, 66]. Thus, in order to investigate in greater detail the effect of His260 protonation on the structure of the CBP and the resulting protein-chromophore interactions, three models were constructed differing on the protonation and the tautomeric state of His260. These models are named as DrBphP-Hie, DrBphP-Hid and DrBphP-Hip where Hie allude a neutrally charged His260

with a proton at N ϵ ; Hid, a neutrally charged His260 with a proton at N δ and Hip a cationic histidine with protons at N ϵ and N δ .

Regarding the chromophore, the BV cofactor is considered fully protonated with deprotonated propionic side chains on the basis of resonance Raman spectra of DrBphP [4] and NMR experiments of Cph1 phytochrome [66, 67].

Fragment molecular orbital (FMO) method. In this work, the FMO method was applied to the three models of the CBD of DrBphP in the Pr state described above. Calculations were performed with the FMO code version 5.3[68] as embedded in General Atomic and Molecular Electronic Structure System (GAMESS) which is a general *ab initio* quantum chemistry package GAMESS [69]. For geometry optimizations and PIEs calculations we used the density-functional tight-binding (DFTB) method with the Grimme’s dispersion correction (D3) and the 3OB parameter set[70, 71]. Solvent effects were included via the polarizable continuum model (PCM)[31]. In these calculations, the entire BV molecule was treated as monomer.

In order to better localize and analyze the interactions between chromophore and protein matrix, PIEs calculations were also carried treating the BV chromophore as composed of four fragments (*vide infra*). These calculations were performed using the MP2 method (2nd order Møller-Plesset perturbation theory)[30] with the 6-31G* basis set and the PCM solvent model.

Evaluation of the resulting FMO-optimized structural models was done by computing root-mean-square deviations (rmsd) of the atomic positions of heavy atoms relative to the backbone atoms of crystal structure after alignment of the backbone atoms. These computations were performed using the VMD software[72]. For better identifying and localizing structural differences between optimized and experimental geometries, the rmsd of the positions of the heavy atoms in each residue were computed and mapped onto the protein structure in the form of a color code ranging from blue (rmsd = 0.44 Å) to red (rmsd = 3.36 Å). In this case, the structures were aligned using the ‘measure fit’ tool of VMD.[72]

Fragmentation of the DrBphP model. All three models (DrBphP-Hip, DrBphP-Hie and DrBphP-Hid) addressed in this work were fragmented using Facio program[73]. The BV-chromophore consisting of two propionic side chains and four pyrrole rings with a high electron delocalization was fragmented manually. In order to preserve the chemical identity of the bilin molecule and obtain a more detailed description of chromophore-protein interactions, we chose the fragmentation shown in Figure 2.2. Accordingly, both propionic side chains, B (propB) and C (propC) can be treated each as a single fragment. Adequate description of the high electron delocalization of BV-molecule can only be achieved by joining rings B, C and D in one fragment (rings-BCD) separated from ring A which, along with Cys24, form the fourth fragment (ring-A-Cys). By using this fragmentation scheme it is possible to identify key interactions between protein and specific regions of BV-chromophore.

2.3 Results and Discussion

Geometry optimization using the FMO method. The geometries of all three structural models of the CBD of DrBphP in the Pr state were optimized using the FMO-DFTB3-D/PCM

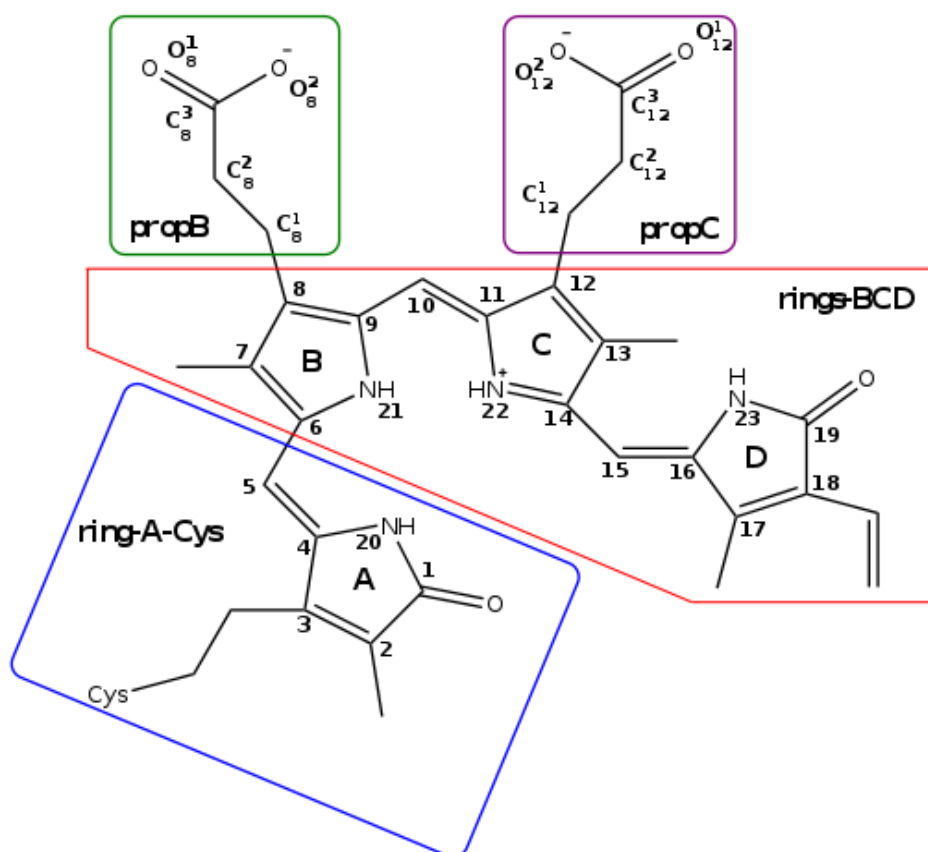


Figure 2.2: The BV-chromophore divided into four fragments. For preserving the chemical identity of the bilin chromophore and obtain a more detailed description of chromophore-protein, both propionic side chain B (propB) and propionic side chain C (propC) are treated each as a single fragment. Rings B,C and D form only one fragment (rings-BCD), and ring A along with Cys24 form the fragment ring-A-Cys.

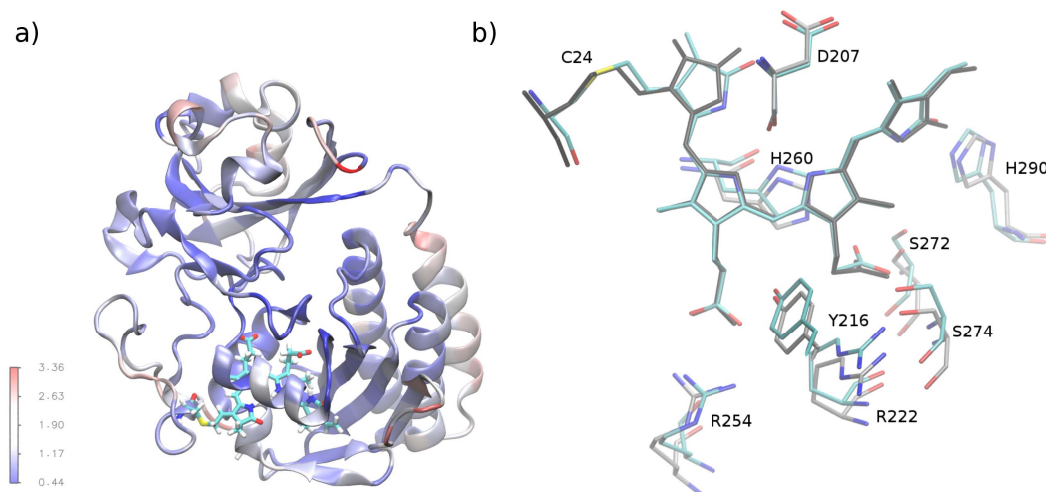


Figure 2.3: a) Phytochrome structure of the DrBphP-hip structural model in the Pr state colored according to displacement after geometry optimization using the FMO-DFTB3-D/PCM method (0.44 and 3.36 Å). This representation shows residues with a low displacement (0.44 Å) in blue, those with a high displacement (3.36 Å) in red, and those in-between as white. The RMSD was also calculated after alignment of backbone heavy atoms of structural model DrBphP-hip with respect to the crystal structure (1.38 Å). b) BV molecule and residues in the vicinity of the BV chromophore of the optimized DrBphP-hip structural model (color cyan) compared to the crystal structure (color gray).

method (~ 6159 atoms). The root-mean-square deviations (rmsd) of the atomic positions of heavy atoms were computed for all models after alignment of backbone atoms of the protein with respect to the crystal structure. Values around 1.40 Å predicted for all three optimized structural models (DrBphP-Hie, DrBphP-Hid and DrBphP-Hip) reflect only minor deviations with respect to the experimental crystal structure (PDB entry 4q0h). The degree of difference between original crystal- and FMO-optimized structures can be visualized in Figures 2.3a, S1 and S2. According to these plots, protein residues that show high displacements upon geometry optimization, above 2.2 Å, (in red color) with respect to the crystal structure are mainly those that are exposed to the solvent and, in turn, are highly flexible. Clear examples are the N-terminal domain (residues 3 to 8), the $\alpha 4$ helix in the GAF domain (residues 136 to 152), the $\alpha 3$ helix (residues 80 to 89) in the GAF domain as well as β -turns and loops regions of both domains. In the case of the $\alpha 4$ -helix, the relatively large rmsd values, above 1.5 Å predicted for each residue results from the absence of the sister molecule, present in the crystalline state of the DrBphP dimer. In contrast, residues embedded in the protein matrix display relatively low rmsd values, below 1.5 Å, (in blue and white colors). This is particularly true for the β sheets in the GAF and PAS domains as well as for most residues in the CBP (Table 4.6 and Table S1).

Structure of the BV-chromophore. Optimization of the protein geometry with the FMO-DFTB approach leads to subtle though relevant changes of the structure of the BV moiety as reflected by the relatively low rmsd values of about 0.4 Å computed for the three models (see table 2.1). As discussed later, these rmsd values are mainly the result of a slight twisting of the tetrapyrrole backbone and minor reorientation of the propionic side chain of ring C (see Figure S3).

Independent of the protonation state of His260, the DFTB method predicts significant changes of the methine bridge bond lengths such as the elongation of the $C4 = C5$ and $C10 - C11$ bonds and shortening of the $C5 - C6$, $C9 - C10$ and $C14 - C15$ bonds which can be interpreted as an increase of π electron conjugation at the methine bridges (see table 2.1). These changes of bond lengths are accompanied by variations of the corresponding bond angles and torsional angles. Noteworthy is the increase in more than 10° of the tilt angles $\tau(A,B)$ and $\tau(B,C)$ relative to the crystal geometry suggesting a more twisted tetrapyrrole conformation rather than a planar arrangement of rings B and C. These variations of the tilt angles are particularly high for the DrBphP-Hie model for which values of 22.2° , 19.0° and 47.6° are predicted for $\tau(A,B)$, $\tau(B,C)$ and $\tau(C,D)$, respectively. Similar elevated tilt angles were also reported for the phycocyanobilin chromophore of Cph1[12] and, in less degree, for BV in Agp1 phytochrome [62] using in both cases density functional theory.

Regarding the propionic side chain on ring C (propC), slight changes (less than 15°) of the torsional angles $C12 - C_{12}^1 - C_{12}^2 - C_{12}^3$, $C_{12}^1 - C_{12}^2 - C_{12}^3 - O_{12}^1$ and $C_{12}^1 - C_{12}^2 - C_{12}^3 - O_{12}^2$ are predicted (table 2.1). In the case of the DrBphP-Hid model, the carboxyl group rotates 14° and moves away from His260 in order to reduce the $O \cdots N\epsilon$ repulsion. In contrast, in the DrBphP-Hie and DrBphP-Hip structures, the carboxyl group rotates 14° toward His260 strengthening the hydrogen bond interaction. The $O_{12}^1 \cdots N\epsilon$ distance reduces to 2.90 and 2.71 Å compared to 3.28 Å predicted for the DrBphP-Hid model. Thus, despite the practically identical rmsd values predicted for the three structural models, the protonation state and tautomeric state of the His260 does significantly influence the structure and, in turn, the electronic properties of the BV molecule. The same observation was reported previously for other phytochromes. [12, 62]

Structure of the Chromophore Binding Pocket. The rmsd values of the atomic position of heavy atoms of selected residues in the CBP (See Figure 2.3b) relative to the crystal structure after alignment of BV molecules are listed in Table 4.6 for the three structural models. Accordingly, Ser274, Arg254, Arg222 and the Ser272 show the highest rmsd values independent of the structural model. These high values result from slight shifts of the BV molecule with respect to the central β -sheet of the GAF domain as consequence of electrostatic interactions involving primarily the anionic propionic side chains. For instance, in case of the Ser274, the side chain slightly moves towards propC by reducing the hydrogen-bond distance between the hydroxyl- function and the carboxylic group from 2.68 Å (crystal structure) to 2.66, 2.59 and 2.64 Å for the optimized DrBphP-Hip, DrBphP-Hid and DrBphP-Hie structural models, respectively. The same effect is predicted for the Ser272 which is also hydrogen bonded to the propC. For this residue, the hydrogen-bond distance reduces from 2.71 Å (crystal structure) to ~ 2.60 Å in the optimized models.

The rearrangement of the Ser274 in the CBP triggers the movement of the cationic side chain of Arg222 towards the BV cofactor. This process is accompanied by the strengthening of the hydrogen-bond between the side chain of Arg222 and the hydroxyl group of Ser274 as reflected by the shortening of the $N \cdots O$ distance from 2.94 Å (crystal structure) to ~ 2.88 Å (optimized structures).

Furthermore, the conserved Arg254 residue forms two salt bridges with the propB. Upon geometry optimization, slight reduction of $O \cdots N$ hydrogen bond distances from 2.83 and 2.88

Table 2.1: Selected structural parameters of BV molecule in DrBphP-Hip, DrBphP-Hid and DrBphP-Hie models. Bond lengths are given in Å while bond angles, torsional angles and tilt angles between rings in degree (°).

	4q0h	DrBphP-Hip	DrBphP-Hid	DrBphP-Hie
$C4 = C5$	1.377	1.385	1.380	1.382
$C5 - C6$	1.479	1.415	1.419	1.416
$C4 - C5 - C6$	130.1	128.5	131.0	128.9
$N - C4 - C5 - C6$	9.9	14.3	13.5	15.8
$C4 - C5 - C6 - N$	1.6	-3.0	-3.0	-5.4
$\tau(A,B)$	8.5	17.1	17.4	22.2
$C9 - C10$	1.483	1.400	1.400	1.392
$C10 - C11$	1.398	1.405	1.398	1.404
$C9 - C10 - C11$	130.8	131.0	131.0	130.4
$N - C9 - C10 - C11$	3.2	3.2	6.3	5.6
$C9 - C10 - C11 - N$	0.6	11.2	10.1	13.3
$\tau(B,C)$	2.2	13.5	15.6	19.0
$C14 - C15$	1.489	1.433	1.434	1.433
$C15 = C16$	1.379	1.375	1.374	1.372
$C14 - C15 - C16$	128.7	125.4	126.0	130.4
$N - C14 - C15 - C16$	-150.2	-161.8	-158.3	-157.8
$C14 - C15 - C16 - N$	17.7	22.5	20.1	18.2
$\tau(C,D)$	49.4	48.5	48.6	47.6
$C12 - C_{12}^1 - C_{12}^2 - C_{12}^3$	76.0	70.0	82.3	75.0
$C_{12}^1 - C_{12}^2 - C_{12}^3 - O_{12}^1$	-130.4	-116.5	-144.4	-115.8
$C_{12}^1 - C_{12}^2 - C_{12}^3 - O_{12}^2$	51.9	64.9	38.4	65.5

Table 2.2: Root Mean Square Deviations (Å) of the position of heavy atoms of selected residues located in CBP relative to crystal structure (PDB entry 4q0h) after alignment of the heavy atoms of the BV moieties.

	DrBphP-Hip ^a	DrBphP-Hid ^b	DrBphP-Hie ^c
BV	0.38	0.36	0.42
Ser274	1.12	1.21	1.15
Arg254	1.07	1.06	1.03
Arg222	1.00	1.04	0.98
Ser272	0.95	1.03	1.00
His260	0.69	0.69	0.87
Tyr216	0.62	0.61	0.58
Asp207	0.59	0.56	0.58
His290	0.54	0.60	0.62

^a His260 doubly protonated; ^b His260 protonated at the δ position; ^c His260 protonated at the ϵ position.

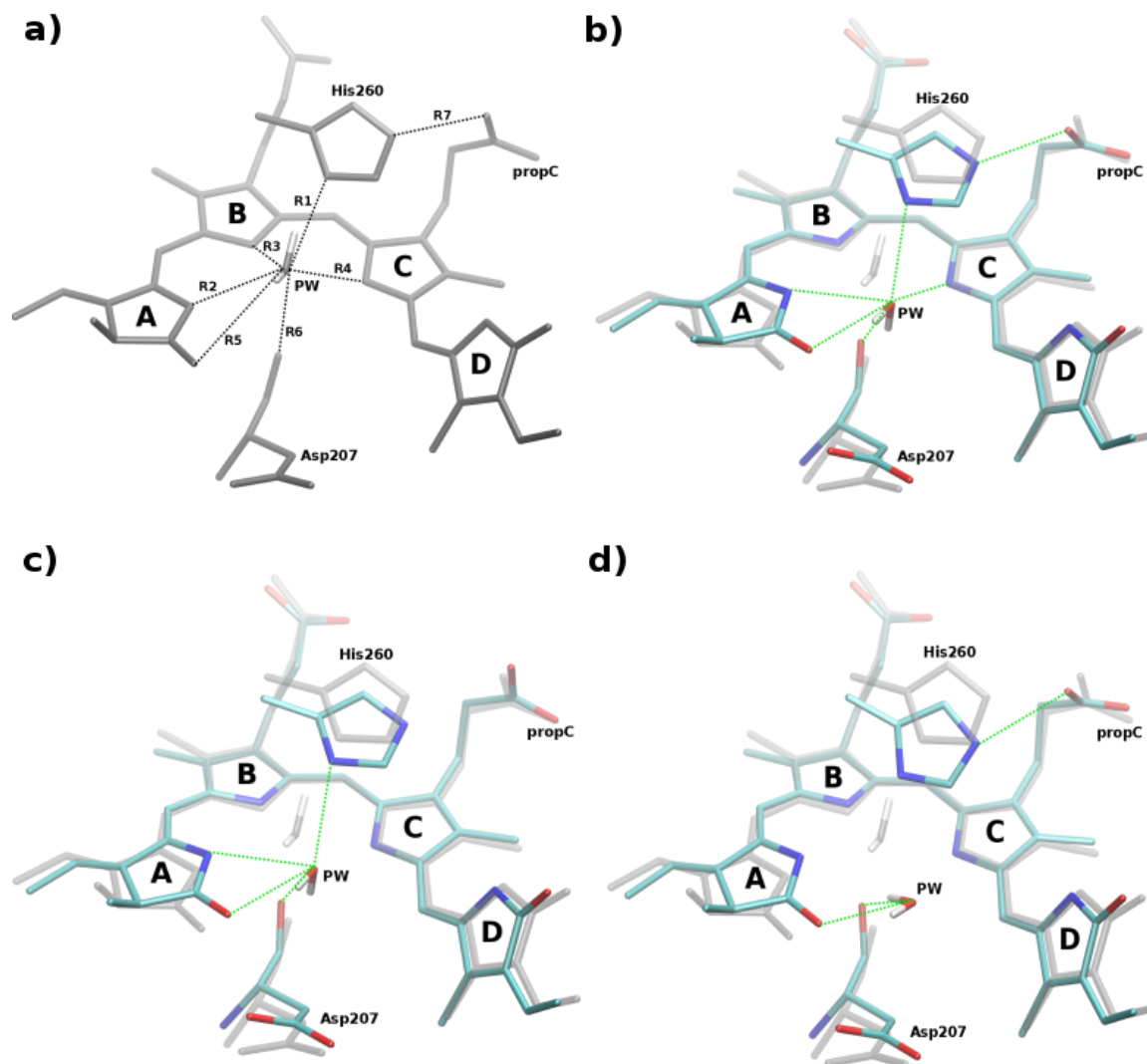


Figure 2.4: a) H-bond distances between pyrrole water oxygen with His260 (R1), the three pyrrole nitrogens of rings A (R2), B (R3), and C (R4), the carbonyl group of ring A (R5), carbonyl group of Asp207 (R6) and propionic side chain C (R7) in the crystal structure. b) H-bond interactions in the optimized DrBphP-hip structural model. c) H-bond interactions in the optimized DrBphP-hid structural model d) H-bond interactions in the optimized DrBphP-hie structural model.

Å (in the crystal structure) to ~ 2.77 and ~ 2.73 Å (in the optimized structures) are predicted which lead to the strengthening of the corresponding salt bridges and thus contributes to the stabilization of the BV cofactor in the binding pocket (*vide infra*). Notwithstanding, the relatively large rmsd value computed for this residue (~ 1.1 Å) is mainly produced by the displacement of the backbone atoms on a flexible turn.

As can be seen in Table 4.6 the change in the protonation state of the His260 does not have a significant influence on the rmsd values of the residues Tyr216, Asp207 and His290 that are in the vicinity of the BV-chromophore. This is not the case for the His260 itself. Table 4.6 shows that the rmsd values computed for the heavy atoms of the His260 are the same for the DrBphP-Hip and DrBphP-Hid structural models (0.69 Å) while a much larger value is predicted for the DrBphP-Hie structure (0.87 Å). Such low rmsd values of the His260 in DrBphP-Hip and DrBphP-Hid models favor a tautomeric form of His260 carrying a proton on N δ .

Table 2.3: Hydrogen-bond Network in CBP involving the BV chromophore, pyrrole water, His260 and Asp207. Hydrogen donor - acceptor distances (Å) are defined in Figure 2.4a.

Structural model	R1	R2	R3	R4	R5	R6	R7
Crystal structure	2.76	2.78	2.95	3.08	3.73	3.06	2.80
DrBphP-Hip	2.77	2.84	3.69	3.00	2.75	2.59	2.71
DrBphP-Hid	2.86	2.84	3.69	3.02	2.80	2.61	3.28
DrBphP-Hie	3.25	3.30	4.26	3.12	2.82	2.70	2.90

An essential element of the CBP of phytochromes is the so called 'pyrrole' water (PW). This highly conserved water molecule does not only confer stability to the chromophore but is also involved in functional relevant hydrogen transfer processes[74, 75, 12, 76] together with the His260. According to the crystal structure, the pyrrole water acts as a bridge between the pyrrole rings of the BV cofactor, the Asp207 and the His260 through a network of hydrogen bonds (see Figure 2.4a). Thus, minor variations in the position and orientation of the pyrrole water will translate into the strength of these interactions. Table 2.3 shows the distances (R1 to R7) between hydrogen acceptor- and hydrogen donor groups in the CBP involving the His260, BV cofactor, PW and Asp207 for all three optimized models. The R1 distance between PW and N δ of His260 (2.76 Å in the crystal structure) significantly increases if this position remains deprotonated as in the DrBphP-Hie model (3.25 Å in the optimized structure), indicating the rupture of the corresponding hydrogen bond. In this case, the PW is oriented in such a way that the interactions with the carbonyl group of ring A in BV and with the carbonyl backbone of Asp207 are strongly enhanced, as reflected by the shortening of the R5 and R6 distances. This last effect has been also predicted for the other two structural models. In addition, the presence of a proton at N ϵ of His260, as in models DrBphP-Hip and DrBphP-Hie, promotes the formation of a hydrogen bond with the propionic side chain C (*vide supra*). Protonation of His260 at the position delta, however, has minor effect on the hydrogen bond distance R1 with the pyrrole water oxygen (2.77 Å and 2.86 Å in the DrBphP-Hip and DrBphP-Hid optimized structures, respectively, compared to 2.76 Å measured in the crystal structure). Furthermore, R2 distances lower than 3.0 Å reflect the formation of hydrogen bonds with rings A and C of the BV moiety, as it is the case in DrBphP-Hip model. Interestingly, based on geometrical criteria, only a cationic histidine is able to close the His260-mediated hydrogen bond network connecting the propC with the inner rings of the BV molecule (Figure 2.4b). The relevance of such a intramolecular hydrogen bonded network has been recently discussed for bacterial bathy phytochromes[77].

PIEs calculations. The His260 is a highly conserved residue that plays a crucial role in the conformational stability of the CBP in phytochromes and is most likely involved in transient proton transfer events[74, 75, 12, 76]. Recently, Rumfeldt et al.[78] determined the importance of His260 for BV protonation by means of pH-dependent UV-Vis spectroscopy and spectral decomposition modeling. As mentioned above, the His260 is able to interact by means of hydrogen-bonds with the propC of the BV-chromophore and the PW. However, depending on its tautomeric and protonation states, the strength of these interactions and , in turn, the stability of the CBP, may change significantly.

In order to identify the protonation and tautomeric state of the His260 which most stabilizes the BV chromophore in the CBP, we computed the corresponding pair interaction energies

(PIEs) using the optimized structural models DrBphP-Hip, DrBphP-Hid and DrBphP-Hie. It should be noted that for these computations, the BV-chromophore was treated as a single fragment. Figures 2.5a, 2.5b and 2.5c show the PIEs between BV-chromophore and protein, deduced for each model.

The FMO analysis of the DrBphP-Hip model led to the identification of twenty major interactions between the BV molecule and protein matrix. As shown in Figure 2.5a the BV chromophore is stabilized mainly by three residues; Arg254 (-116.27 kcal/mol), His260 (-84.59 kcal/mol) and Arg222 (-74.17 kcal/mol) in the CBP. Among them, the His260 is involved in a hydrogen bond network with the propC and five water molecules, including the PW as discussed above. It is also important to note that there is a large electrostatic component between the cationic His260 and both propionic side chains B and C (negatively charged) that contributes significantly to the stabilization of the CBP (see below).

This strong electrostatic component clearly vanishes in the DrBphP-Hid model containing a neutrally charged His260 residue. As depicted in Figure 2.5b the FMO method predicts that the largest contributions to BV-protein interaction energies derive from the Arg254 (-117.79 kcal/mol), Arg222 (-76.15 kcal/mol) and Arg302 (-36.21 kcal/mol). In this model, no direct hydrogen bond interactions are predicted between the BV molecule and the His260.

A slightly different picture is obtained for the DrBphP-Hie model harboring also a neutrally charged His260 (Figure 2.5c). Here, due to the formation of a strong hydrogen bond with the propC, the His260 contributes with -42.51 kcal/mol together with the Arg254 (-117.71 kcal/mol) and Arg222 (-74.57 kcal/mol) to the stabilization of BV in the CBP.

The fact that the PIEs computed for the Arg254 and Arg222 residues are essentially the same for the three models though significantly different for the His260, clearly highlights the relevance of this titratable residue in regulating the energetic of the BV chromophore in its environment. While single protonation of the His260 at either N ϵ or N δ positions has a destabilizing effect on the chromophore pocket, a cationic His260, protonated at both nitrogen sites, has the opposite effect.

In all structural models the BV chromophore (negatively charged) is destabilized by the same anionic residues (Glu125, Glu127, Glu306, Asp 226, Asp95, Asp171, Glu27, Glu189, Asp300 and Glu166) as reflected by their positive contribution to the PIEs (Figures 2.5a, 2.5b and 2.5c). These interactions are of pure electrostatic nature (Figure S5). The long-range character of these interactions is evident when evaluating the magnitude of the PIE associated to a certain residue and its distance to the propionic side chains of BV. For instance, Glu125 which is 11 Å away contributes shows a PIE of 33 kcal/mol while Glu166 located 21 Å away from BV has PIE of 18 kcal/mol. Figures 2.5a, 2.5b and 2.5c also show that most of these repulsive electrostatic interactions are counteract by attractive counterparts (Arg222, Arg302, Arg172, Arg26, Arg218 and Arg100).

PIEDA calculations of the DrBphP-Hip structural model. In this work, PIEDA was used for identifying the nature and quantifying the strength of non-covalent interactions between BV-chromophore and protein using the DrBphP-Hip structural model. Alike the FMO calculations described in the former sections, the BV molecule was divided into four fragments (Figure 2.2) in order to achieve a better understanding of the chromophore-protein interactions. Furthermore, these PIEDA calculations were performed using the MP2 method

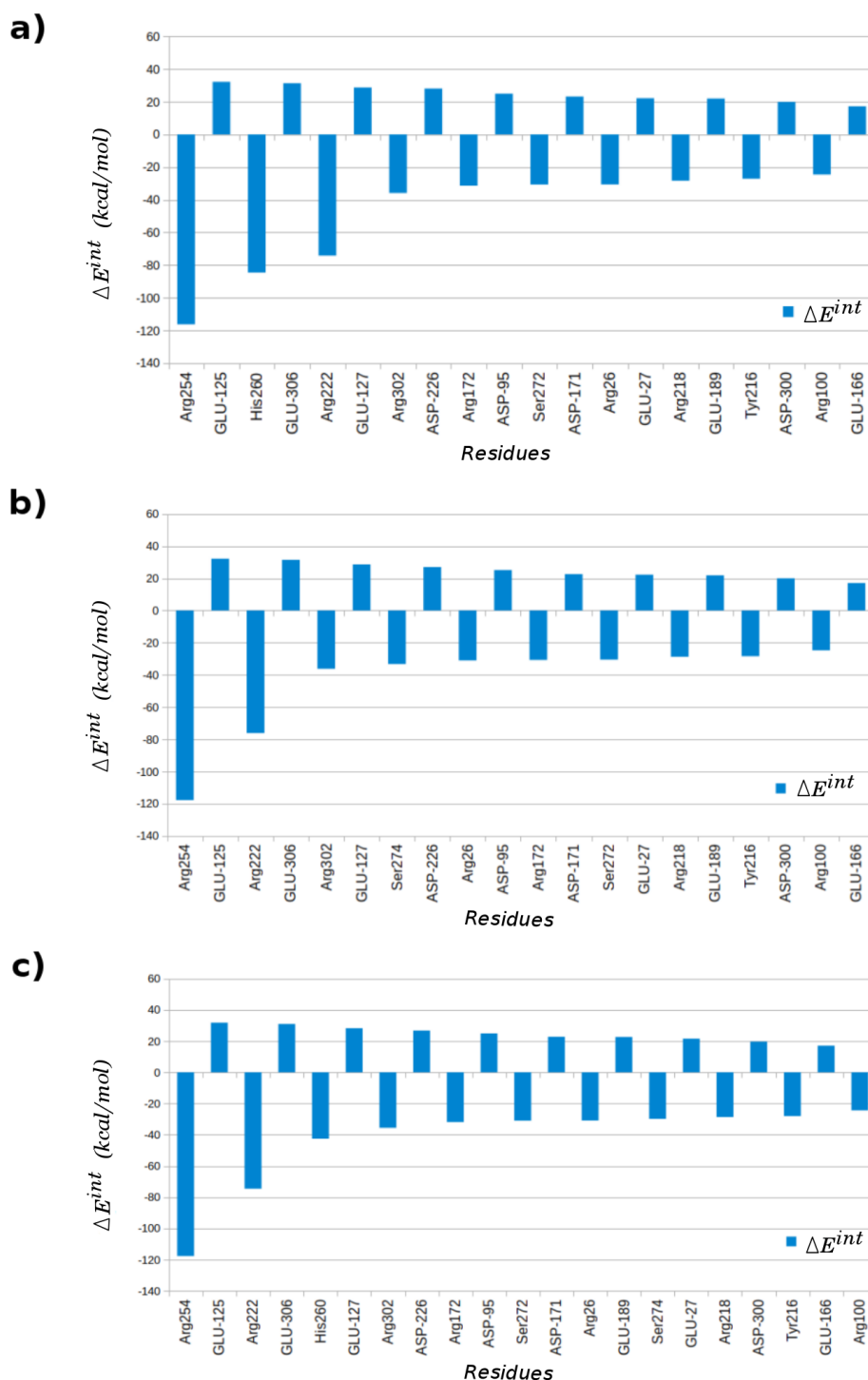


Figure 2.5: a) PIEs plot between BV-chromophore and protein for DrBphP-Hip optimized model. b) PIEs plot for DrBphP-Hid optimized model. c) PIEs plot for DrBphP-Hie optimized model.

with the 6-31G* basis set and the PCM solvent model. The division of the BV molecule in smaller fragments not only allows reducing the computational cost but also provides a more detailed analysis of the PIEs by specifying these interactions with respect to parts of the BV molecule. These fragments will be named as *ring-A-Cys*, *rings-BCD*, *propB* and *propC* (Figure 2.2). More information on the fragmentation scheme can be found in the methodology section.

PIEDA involving fragment ring-A-Cys. As illustrated in Figure 2.6a, FMO calculations identified eleven significant interactions between fragment *ring-A-Cys* and DrBphP phytochrome structure. The fragment *ring-A-Cys* is stabilized mainly by the Arg26 (-20.84 kcal/mol). This interaction is dominated by the electrostatic term as reflected by the large blue area of the corresponding PIEDA column. The two H-bonds formed between the PW (HOH504) and the C=O and NH groups of ring A (Figure 2.7) contribute with -14.22 kcal/mol to the total PIE involving this fragment. Another H-bond interaction is detected by FMO between the carbonyl backbone of the Thr21 and the sulfur atom of Cys24 (-9.77 kcal/mol). This interaction shows large exchange repulsion character (red area in the corresponding PIEDA column) compensated by attractive electrostatic interaction as well as dispersion and charge transfer components, characteristic of a weak hydrogen bond interaction.[79] Additional hydrogen-bonds contributing to the stabilization of ring A are those involving HOH608 (-9.60 kcal/mol), HOH759 (-9.59 kcal/mol), HOH575 (-9.22 kcal/mol) and Glu25 (-5.045 kcal/mol).

The *fragment ring-A-Cys* is destabilized by HOH566, Asp207, HOH605 and Glu22. These interactions are dominated mainly by the repulsion term, which is reflected in the large red area of the corresponding PIEDA columns (Figure 2.6a). An exception is the interaction with HOH566 where the electrostatic component is the dominant term. In all cases, the dispersion term contribute to the stabilization of *fragment ring-A-Cys*. However, this attractive interaction is compensated either by the repulsion component or by the electrostatic energy term.

PIEDA involving fragment rings-BCD. FMO analysis led to the identification of sixty-two major interactions ($|PIE| \geq 5$ kcal/mol) between fragment *rings-BCD* and protein (Supplementary Fig. S6 and S7). However, in order to reduce the complexity of the analysis we selected just the twenty largest interaction energies (Figure 2.6b). The backbone of Asp207 formed three H-bonds with rings A, B, and C of BV-chromophore (-56.80 kcal/mol) (Fig. 2.7). Furthermore, the side chain of the Asp207 (negatively charged) has a large electrostatic interaction with the fragment rings-BCD (positively charged) (-48.27 kcal/mol). One H-bond between the side chain of Glu27 and a methyl group of the fragment rings-BCD (-50.05 kcal/mol) was detected by FMO. Another significant interaction between charged Glu25 and charged fragment rings-BCD (-46.55 kcal/mol) was found by FMO (a purely electrostatic interaction, see Figure 2.6b). The PIEs with GLU-18 (-23.94 kcal/mol), ASP-181 (-23.12 kcal/mol), GLU-189 (-21.71 kcal/mol), GLU-306 (-21.33 kcal/mol), GLU-125 (-18.89 kcal/mol) and ASP-95 (-18.52 kcal/mol) detected by FMO have a purely electrostatic nature (Figure 2.6b).

The *fragment rings-BCD* (positively charged) is mainly destabilized by His260 (positively charged). This interaction is driven by the electrostatic and exchange-repulsion terms as is reflected by the large blue and red areas of the corresponding PIEDA column. The dispersion and charge transfer terms contribute to the stabilization of *fragment rings-BCD*, nevertheless,

these terms are overcome by the electrostatic and repulsion energy components. Other three significant interactions were detected by FMO as destabilizing interactions (Arg222, Arg254 and Arg213). In all cases, the electrostatic component is the responsible for this destabilization. The remaining six interactions (Arg265, Arg218, Arg202, Lys177, Arg172 and Arg26) also are dominated by the electrostatic energy term.

Although FMO identified Asp207 and Glu27 as key residues for the stabilization of the pyrrole rings B, C and D by forming four H-bonds, the electrostatic component also plays an essential stabilizing role as reflected by the blue columns in the PIEDA plot (Figure 2.6b). The main interactions detected by FMO are consistent with previous studies [57, 6, 2, 5], however, some hydrophobic interactions reported in these studies based on the crystal structure, were not identified by FMO as stabilizing interactions. For example, FMO detected a repulsive interaction between Pro209 and rings-BCD (13.03 kcal/mol) which is in contradiction with these previous works. As illustrated in Figure S7, the dispersion term contributes to the stabilization of the rings-BCD (-6.10 kcal/mol), while the electrostatic (6.34 kcal/mol) and repulsion (9.00 kcal/mol) energy terms destabilize the rings-BCD and they overcome the dispersion energy term.

PIEDA involving fragment propB. FMO calculations identified seventy significant interactions between propionic side chain B (propB) and protein (Supplementary Fig. S8 and S9). However, as mentioned earlier, we will focus on the twenty largest interaction energies. Figure 2.6c shows PIEs between propB and protein. It is worth noting that one pair has significantly larger interaction energy. This interaction detected by FMO corresponds to a salt bridge between the conserved Arg254 and propB (-128.19 kcal/mol) (Fig. 2.8) and contributes mainly to the conformational stability of the BV-chromophore. Another significant interaction found by FMO is that between charged Arg222 and charged propB (-49.16 kcal/mol). This interaction is dominated by the electrostatic term as illustrated by the large blue area in the PIEDA column. Two other interactions dominated by the electrostatic term were found between the *fragment propB* with the positively charged His260 (-49.09 kcal/mol) and the Arg218 (-31.74 kcal/mol). One hydrogen bond was formed between Tyr216 side chain and propB, although this interaction is mainly driven by the electrostatic term, dispersion and charge transfer terms also contribute significantly to it. Three electrostatic interactions were found between propionic side chain B with the Arg265 (-27.42 kcal/mol), Arg302 (-24.68 kcal/mol) and Arg213 (-21.94 kcal/mol). Two additional hydrogen bonds were formed between propB and Ser257 backbone (-26.43 kcal/mol) and water molecule HOH505 (-21.89 kcal/mol).

Two significant destabilizing interactions were identified by FMO between *fragment propB* and protein, namely, Glu27 and propionic side chain C. These interactions are of pure electrostatic nature as can be seen in the whole blue area of the corresponding PIEDA columns. The remaining eight interactions (Glu125, Glu25, Glu306, Asp95, Glu18, Glu127, Asp207 and Asp226) also are dominated by the electrostatic energy term.

PIEDA involving fragment propC. FMO analysis detected seventy-three strong interactions between propionic side chain C and protein (Supplementary Fig. S10 and S11). Figure 2.6d shows the twenty largest pair interaction energies between propC and protein. Among them, two pairs have considerably larger interaction energy. The first pair corresponds to a H-bond formed between prop C and the side chain of His260 (-100.34 kcal/mol) (Fig. 2.8). This interaction

2. Fully Quantum Chemical Treatment of Chromophore-protein Interactions in Phytochromes

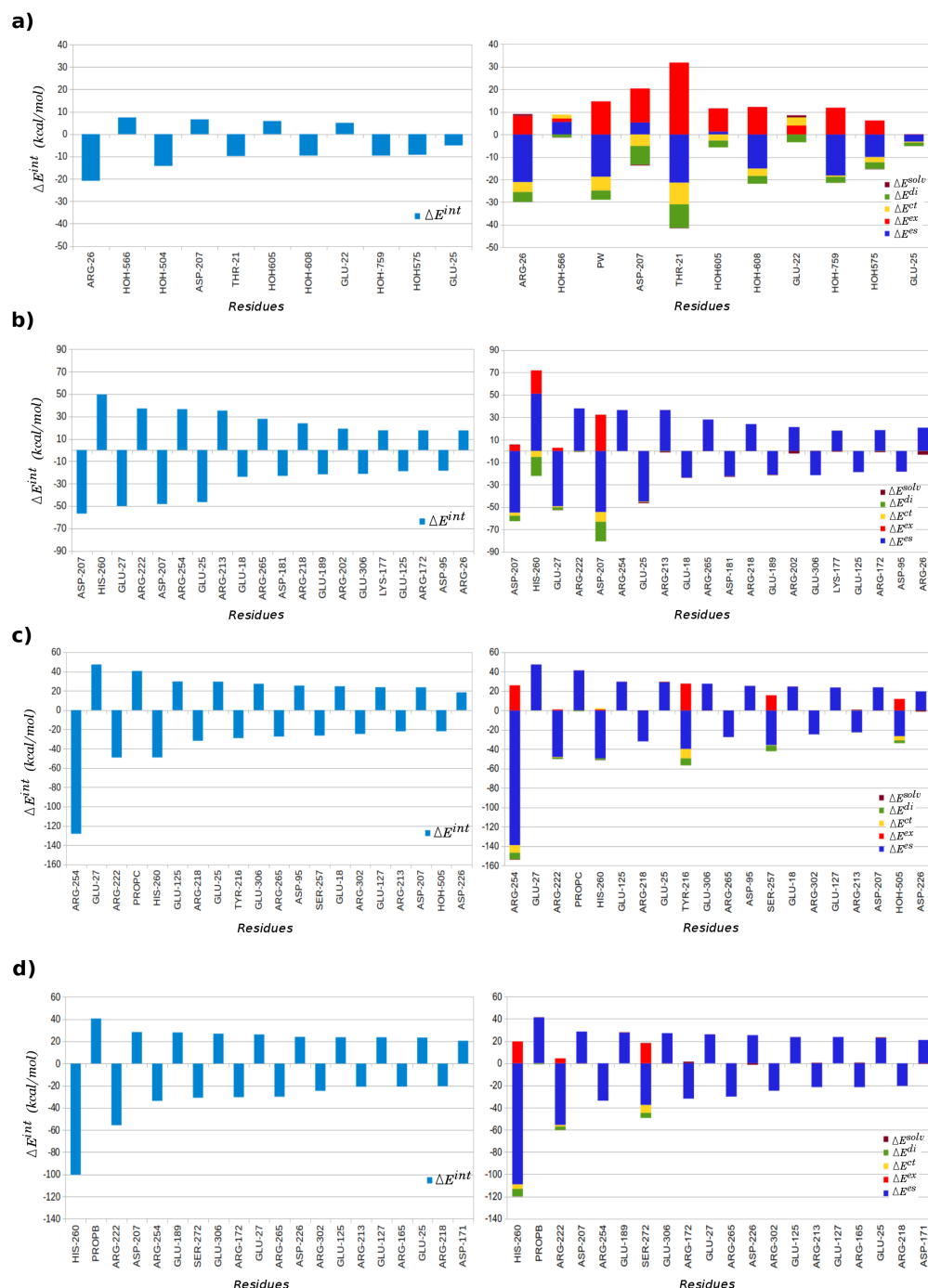


Figure 2.6: PIEs plot (left) and PIEDA plot (right) between BV-chromophore and protein of the DrBphP-hip structural model for a) ring-A-Cys b) rings-BCD c) propB and d) propC fragments. In PIEDA plots the energy terms: electrostatics, exchange repulsion, charge-transfer, dispersion and solvent screening are colored coded blue, red, yellow, green and red wine respectively.

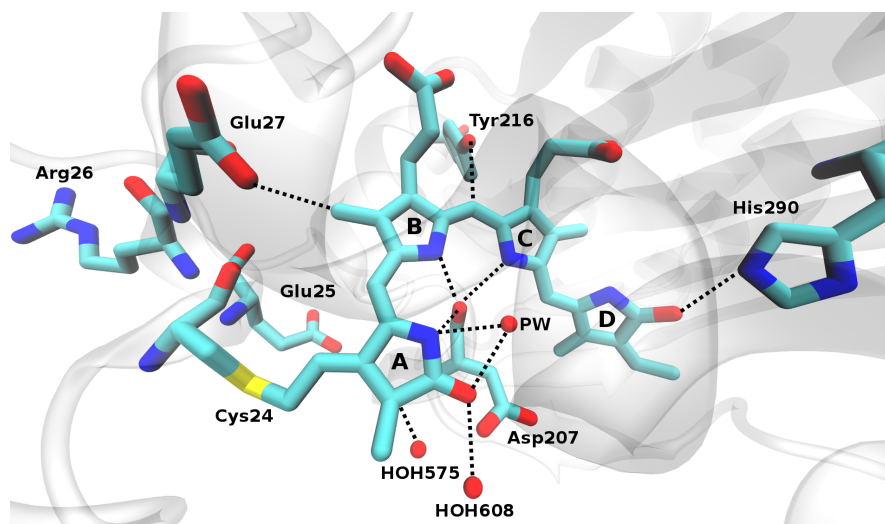


Figure 2.7: Structure of the chromophore binding pocket highlighting key residues and water molecules interacting with the BV tetrapyrrole in the DrBphP-hip structural model.

is clearly dominated by the electrostatic term. The second pair is a strong electrostatic interaction between propC and Arg222 (-55.63 kcal/mol). Another significant interaction between propC and Arg254 (-33.70 kcal/mol) was found by FMO (a purely electrostatic interaction, see Figure 2.6d). An additional H-bond was formed between propC and the side chain of the conserved Ser272 (-30.88 kcal/mol). Six important interactions driven by the attractive electrostatic term were identified between propC with Arg172 (-30.37 kcal/mol), Arg265 (-29.95 kcal/mol), Arg302 (-24.66 kcal/mol), Arg213 (-20.94 kcal/mol), Arg165 (-20.83 kcal/mol) and Arg218 (-20.39 kcal/mol).

Interestingly, the fragment *propC* is principally destabilized by the propionic side chain B. This interaction is dominated basically by repulsive electrostatic energy term, as in the case of the interactions with Glu125, Glu25, Glu306, Asp95, Glu18, Glu127, Asp207 and Asp171.

In addition, FMO identified His260 as a key residue in the conformational stability of the propionic side chain C, which is consistent with previous studies[57, 6, 2, 5]. However, it is well known that the state of protonation of histidine residues can be affected by slight changes in the pH. This could translate into the breaking of the H-bond between propC and His260, which might explain the high flexibility of the propionic side chain C inside the chromophore pocket. Interestingly, Schmidt et al.[80] reported the crystal structure dimer of Agp2 phytochrome in the Pfr state (PDB entry 6G1Y). This structure shows two different conformations of the propionic side chain C, which supports the idea of the high flexibility of propC. On the other hand, the salt bridge formed between propionic side chain B and the conserved Arg254 is a stronger interaction for breaking and is the strongest PIE according to FMO results. This might explain the very little conformational flexibility of propionic side chain B. Furthermore, the highly conserved Arg254 residue forms part of the trefoil knot that stabilizes this region of the protein, working as a structural bridge connecting the GAF and PAS domains[5].

New interactions in the chromophore binding pocket. FMO detected two non-classical hydrogen bonds (CH/O interactions)[81, 82] between the carbonyl oxygens of the Asp207 and the side chain of Tyr263 (-5.93 kcal/mol) (Figure 2.9). Interestingly, Wagner et al.[6] based on the crystal structure, suggested a polar interaction in the chromophore pocket

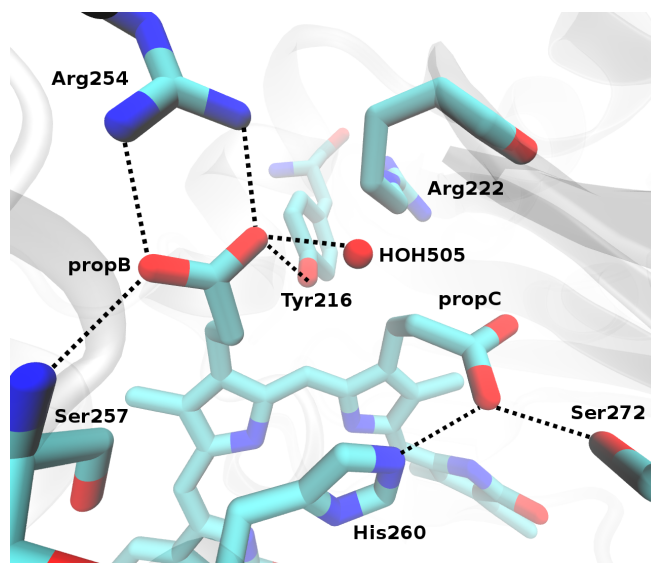


Figure 2.8: Zoom into the structure of the chromophore binding pocket highlighting key residues and water molecules interacting with propionic side chains B and C of the BV cofactor in the DrBphP-hip structural model.

involving the Tyr263 and Asp207, with their hydroxyl- and carboxyl- groups, respectively, separated by a distance of 2.88 Å. They argue that this polar interaction results from a slight rotation of each side chain relative to each other. However, upon geometry optimization of the DrBphP-Hip model, the interaction between these two residues is slightly weakened as reflected by the elongation of $\text{OH}_{\text{Y263}} \cdots \text{COO}^-_{\text{D207}}$ distance to 3.01 Å, which favors the formation of a OH/π interaction between the hydroxyl of Tyr263 and ring D of the BV-chromophore (-6.17 kcal/mol), separated 2.72 Å from each other (Figure 2.9). This interaction is dominated by the dispersion term (Figure S7). It is worth mentioning that, Wagner et al.[75] concluded from a mutagenesis study of combining several spectroscopic methods that Tyr263 and Tyr176 are necessary for the Pr to Pfr photoconversion. Thus, the OH/π interaction identified by FMO might play an important role in stabilizing ring D during isomerisation of the methine bridge double bond that undergoes the chromophore during the photocycle. The identification of this type of interaction in the chromophore pocket allows us to illustrate one of the main advantages of the FMO method.

2.4 Conclusions

In this work, the geometries of three structural models of the *Deinococcus radiodurans* DrBphP phytochrome in the Pr state were optimized (~ 6159 atoms) using the FMO method in combination with the DFTB3 approach[83]. The optimized structural models show only minor deviations from the experimental structure. Aided by PIE calculation we determined the protonation- and tautomeric state of His260 which confers maximum stability to the bilin chromophore of the DrBphP phytochrome in the Pr state. This corresponds to the cationic His260 form, protonated at both $\text{N}\epsilon$ and $\text{N}\delta$, which stabilizes the chromophore pocket mainly by forming hydrogen bond with the propionic side chain C and through strong electrostatic interactions with the propionic side chain B. Interestingly, the interaction of the H260 with the inner pyrrole rings is only indirect via the pyrrole water. Identification of the protonation

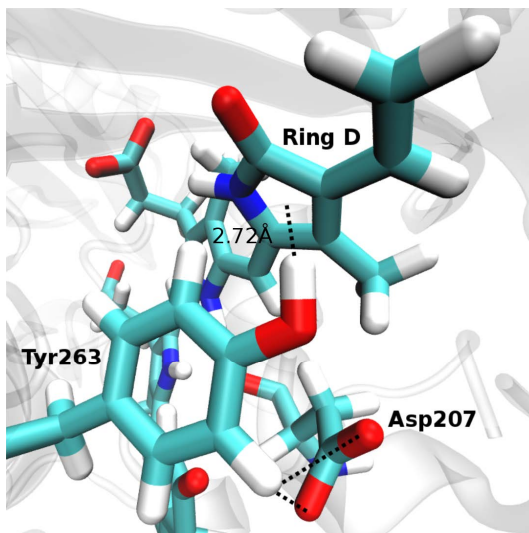


Figure 2.9: Interactions detected by FMO in the chromophore binding pocket: two non-classical H-bonds (CH/O interactions)[81, 82] between the carbonyl oxygens of the Asp207 and the side chain of Tyr263; one OH/ π interaction between the hydroxyl of Tyr263 and ring D of the BV-chromophore.

state of the conserved His260 residue and its influence on the structure of the chromophore binding site are essential requirements for any other quantum chemical calculation used for interpretation and analysis of experimental spectroscopic data.[12]

PIEDA analysis combined with the FMO approach demonstrated to be a powerful tool to quantify the strength and identify the nature of non-covalent interactions between chromophore and protein, using as target, bacterial phytochrome DrBphP in the Pr state. In particular, FMO identified Asp207 and Glu27 as key residues for the stabilization of the pyrrole rings B, C and D through formation of four H-bonds.

Interestingly, FMO also detected a OH/ π interaction between Tyr263 and ring D of the BV-chromophore, this interaction dominated by the dispersion term might play an important role in the stability of the ring D during the isomerisation of the methine bridge double bond that undergoes the chromophore during the photocycle. Furthermore, according to the FMO computations, the conserved Arg254 and His260 residues are key residue involved in the conformational stability of both propionic side chains B and C. For these two residues, very high PIE values ($|PIE| > 100$ kcal/mol) are predicted. The overestimation of PIEs for the case of very attractive or very repulsive contributions result from the current implicit solvent model. [84]. Fedorov demonstrated that drastic improvements in the estimation of PIEs, however, can be achieved by combining FMO with a new solvent screening model. [84]

In summary, the majority of chromophore-protein interactions identified by FMO in DrBphP are consistent with previous studies[57, 6, 2, 5], however, estimation of the strength of these interactions as well as the description of their chemical nature is one of the main contributions of this work to the phytochrome field.

3

The Fragment Molecular Orbital Method: Application to Phytochromes Structures

This chapter presents two applications of the fragment molecular orbital method to different phytochrome structures. In the first part, a brief summary of the following collaboration work is presented:

Fernandez Lopez, M., Nguyen, A. D., Velazquez Escobar, F., **González, R.**, Michael, N., Nogacz, Z., Piwowarski, P., Bartl, F., Siebert, F., Heise, I., Scheerer, P., Gaertner, W., Mroginiski, M.A. and Hildebrandt P. (2019). Role of the propionic side chains for the photoconversion of bacterial phytochromes. *Biochemistry*, 58(33), 3504-3519.

In the second part, the Raman spectrum of a cyanobacteriochrome photoreceptor obtained by using the FMO method within the frozen domain with dimers (FDD) approach is presented.

3.1 Role of the Propionic Side Chains for the Photoconversion of Bacterial Phytochromes

3.2 Introduction

The bacteriophytochromes Agp1 and Agp2 were studied by combining experimental and computational approaches, nevertheless, in this summary, we will focus mainly on the computational study based on the FMO method. The prototypical (Agp1) and bathy (Agp2) phytochromes were modified (esterified) at propionic side chain B (PsB) or C (PsC) of the BV molecule (see Figure 3.1) in order to study the local electrostatic effects in the chromophore binding pocket (CBP), which, in turn, control the mechanism and dynamics of the photoinduced reaction sequences of phytochromes. In the particular case of Agp2 phytochrome, the effects of including a methyl group in the BV molecule (BVM) are significant, since α -helix/ β -sheet secondary structure change of the "tongue" region is blocked, and the final product of the phototransformation slows down drastically. Regarding the Agp1 phytochrome, the result of introducing a BVM in the Pr-to-Pfr phototransformation is less severe, for example, the secondary structural transition of the tongue is not impaired and thermal reversion is even accelerated. The FMO calculations provide a detail analysis of the local electrostatic changes in the CBP of the Agp1 and Agp2 structures/variants. By using pair interaction energies (PIEs), the non-covalent interactions between the propionic side chains of the BV/BVM cofactors and protein residues have been identified and quantified.

3.3 Methods

Preparation of structural models. The structural models of Agp1 (PDB entry 5HSQ)[85] and Agp2 phytochromes (PDB entry 6G1Y)[86] as well as of the Meta-F state of the Agp2 variant PAiRFP2 (PDB entry 6G20)[86] were used as templates for generating the initial geometries of the respective variants. The chemical modifications of BV molecule into BVM-B (esterified at ring B) or BVM-C (esterified at ring C) were done manually (see Figure 3.1), additionally, an homology model of Agp2 in the Meta-F state was obtained with SWISS MODEL using the Agp2 variant PAiRFP2 as a template. In the particular case of Agp1 variants, the propionic side chains unaltered of BV were modeled with the deprotonated carboxylic groups. In contrast, both Agp2-BV and Agp2-BVM-B in the Pfr and the Meta-F state were modeled with the protonated propionic side chain of the ring C based on experimental spectroscopic data[87]. Concerning the model for the photoproduct of the esterified Agp2 adduct, the QM/MM-optimized structure of the Agp2 in Meta-F was used as the initial geometry with a deprotonated propionic side chain at ring C (All QM/MM-optimized structure were provided by Duc Nguyen). Additionally, the structure models for the enol tautomer[87] were constructed by manually moving the hydrogen atom on the pyrrole nitrogen (ring C or B) to the oxygen on ring D.

The protonation states of all titratable sites of the crystal structures and homology models were estimated by using the Karlsberg2+ software package[60]. Regarding the conserved

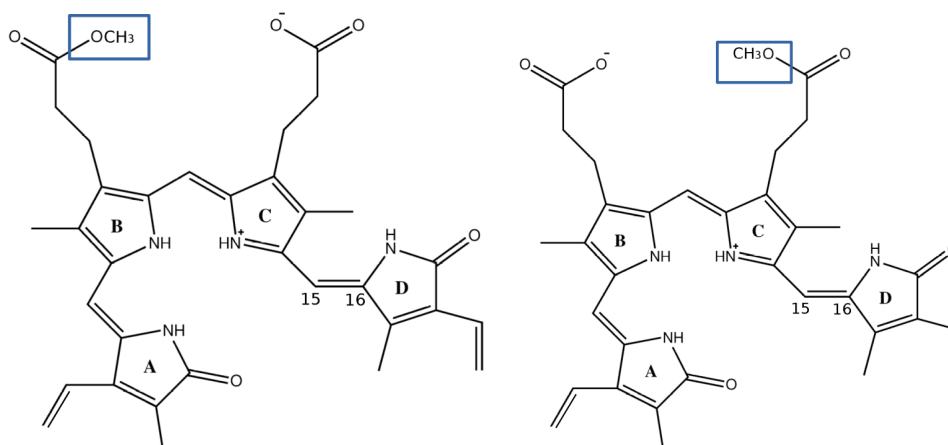


Figure 3.1: Chemical structures of the biliverdin monomethylesters with the ester function (blue) at the propionic side chain of ring B (BVM-B, left) and ring C (BVM-C, right). The chromophores are shown in the ZZZssa configuration of the Pr state; in Pfr, the chromophore adopts the ZZEssa configuration generated by rotation around the C(15)=C(16) bond (red). The chromophore is covalently bound to the protein via addition of a Cys side chain to the vinyl function of ring A (yellow).

residues His250 and His280 in Agp1 and the equivalent His248 and His278 residues in Agp2, their protonation states were modeled as charge neutral with a proton on N ϵ , however, in the Agp2-BVM-B models of the photoproduct, His278 was considered in both the cationic form and the charge-neutral (protonated at N δ) form. Geometry optimization of the BV molecule and the protein matrix was performed by using a well established QM/MM protocol[88]. In these calculations the B3LYP/6-31G* level of theory was applied to the BV cofactor and the Cys20 side chain, whilst the protein and solvent waters were described by the CHARMM32b2 force field[40].

Pair interaction energies. PIEs between the propionic side chains of the BV and/or BVM cofactors and protein residues of the QM/MM-optimized models were computed using the Fragment Molecular Orbital Method (FMO) implemented in FMO version 5.2 that is embedded in General Atomic and Molecular Electronic Structure System (GAMESS) code[69, 68]. PIEs were computed at the MP2/6-31Gd level of theory using the polarizable continuum model (PCM) to describe solvent effects. Fragmentation of the protein structures (one residue per fragment) was performed with the Facio program[73]. The BV chromophore was subdivided manually into four fragments comprising (i) ring A, (ii) rings B-D, (iii) propionic side chain B, and (iv) propionic side chain C.

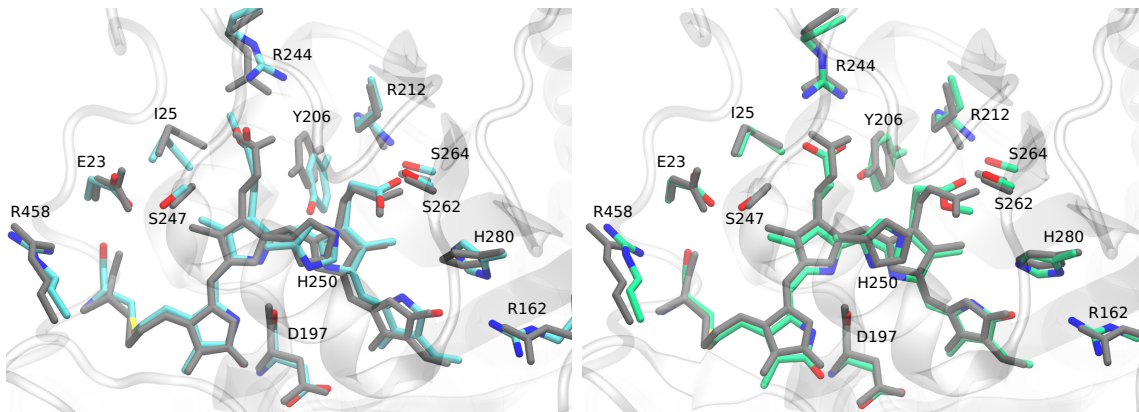


Figure 3.2: Chromophore binding pockets of (left) Agp1-BVM-B (cyan) and (right) Agp1-BVM-C (green). The optimized QM/MM structures of the two Agp1 variants are compared with the QM/MM structure of Agp1-BV (gray).

Table 3.1: BV chromophore and protein structural properties for Agp2-BV, Agp2-BVM-B, and Agp2-BV-PAiRFP2[77, 89, 90]

	stable dark state	intermediate state	final photoproduct
Agp2-BV			
	Pfr	Meta-F(P)	Pr
PsC	protonated	protonated	deprotonated
tautomer	keto	keto	keto/enol
tongue	α -helix	α -helix	β -sheet
Agp2-BVM-B			
	Pfr	Meta-F(P)	Meta-F(D)
PsC	protonated	protonated	deprotonated
tautomer	keto	keto	enol
tongue	α -helix	α -helix	degraded coiled loop
Agp2-BV-PAiRFP2			
	Pfr	Meta-F(P)	Meta-F(P)
PsC	protonated	protonated	deprotonated
tautomer	keto	keto	keto/enol
tongue	α -helix	degraded coiled loop	degraded coiled loop

3.4 Results and discussion.

Parent States. The root-mean-square-deviations (RMSDs) values for the heavy atoms of the BV chromophore (Agp1 and Agp2 optimized structures) compared to the crystal structures are relatively small (~ 0.2 Å) (Figure 3.2). For the Agp1 in the Pr state, the esterification of PsB or PsC has not an important effect in the structure of the chromophore, which is reflected by the relatively low RMSD predicted for the BV (~ 0.4 Å). Nevertheless, the Ile25 and Arg244 residues are reoriented as a result of the steric hindrance of the monomethylester at ring B, the strong salt bridge between the carboxylate and the Arg244 side chain is replaced by a hydrogen bond, and the hydrogen bond between Tyr206 and the carboxylate is broken.

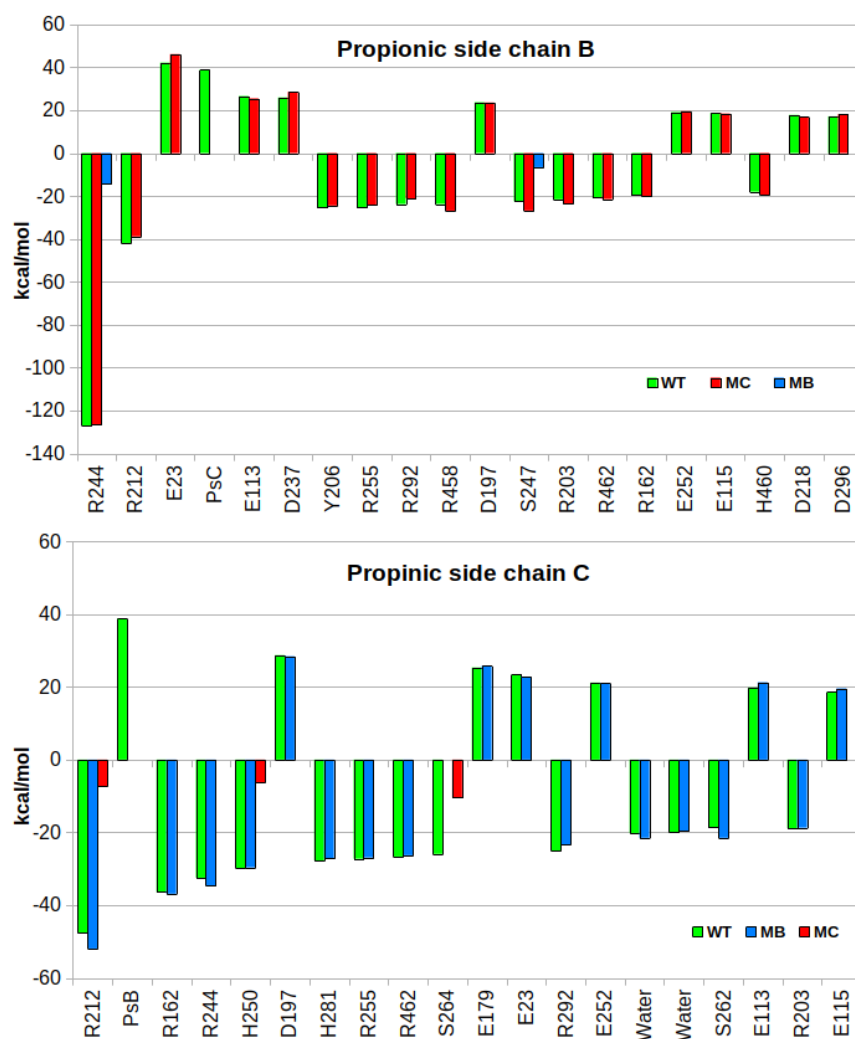


Figure 3.3: Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the propionic side chain B (top) and propionic side C (bottom) computed for models Agp1-BV (green), Agp1-BVM-B (blue) and Agp1-BVM-C (red).

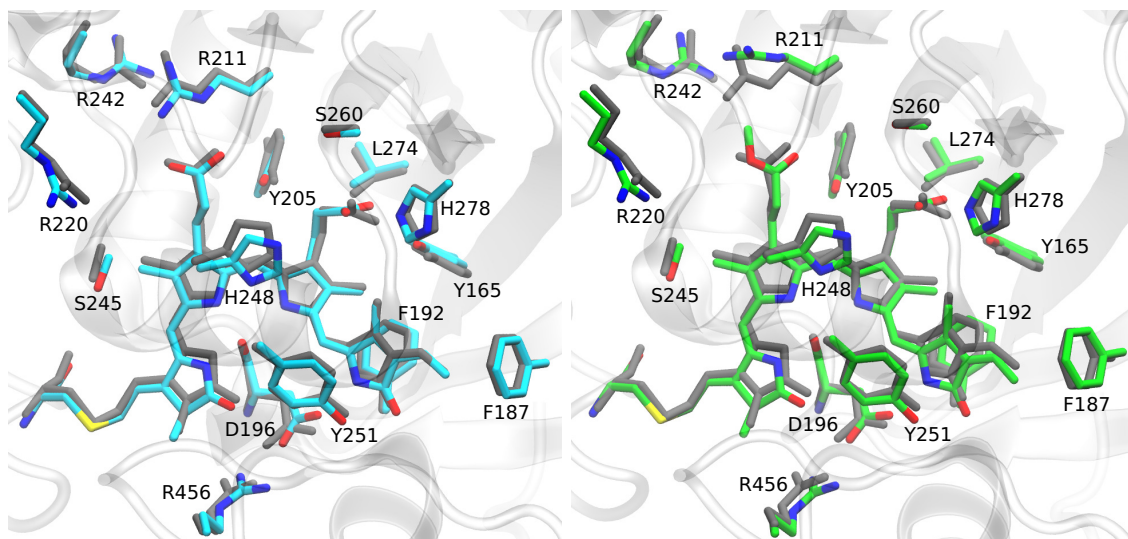


Figure 3.4: Structures of the chromophore binding pockets (left) Agp2-BV (cyan) and (right) Agp2-BVM-B (pink) in the Pfr state. The optimized QM/MM structures are compared with the crystal structure of Agp2-BV (gray, PDB entry 6G1Y).

Regarding the esterification process of PsC, similar changes in the chromophore binding pocket were observed.

PIEs plots (Figure 3.3) show that one of the main effects of including a methyl group at PsB or PsC is to quench the strong electrostatic interactions with charged residues such as Arg244, Arg212, Glu23, Asp197, and Arg162, among others, as well as the repulsion interaction between both propionic side chains. The absence of these strong electrostatic interactions contributes to the destabilization of the chromophore binding pocket. With respect to the Pfr state of Agp2, small deviations were observed for the Agp2-BVM-B optimized structure compared to the crystal structure (RMSD= ~ 0.24 Å for BV cofactor). In the case of Agp2-BV model, similar RMSD values were obtained, suggesting that the presence of the methyl group at the PsB has not an important effect on the conformation of the chromophore binding pocket (see Figure 3.4). The displacement of the Arg211 from methylated PsB is one of the most significant change compared to the WT model. As a result of the esterification of PsB, the interaction between the side chain of Arg211 and PsB is weakened by an increase in distance (from 3.8 to 5.5 Å).

In the particular case of Agp2 in the Pfr state, the esterification of PsB weakens the strong electrostatic interaction with Arg211 and removes the electrostatic interactions with Arg242 and Glu16, among others. In contrast, the hydrogen bond interactions involving PsC, are practically not affected by the esterification of PsB. These results are reflected by the PIEs plot in Figure 3.5.

Photoproducts of Agp2. Regarding the modeling of the photoconversion products of Agp2, the crystal structure of the photoconversion product of the Agp2 variant PAiRFP2, i.e., the Meta-F(P) (see Table 3.1) state was chosen as starting point. No significant structural changes were observed in the chromophore binding region upon the back-mutations of the

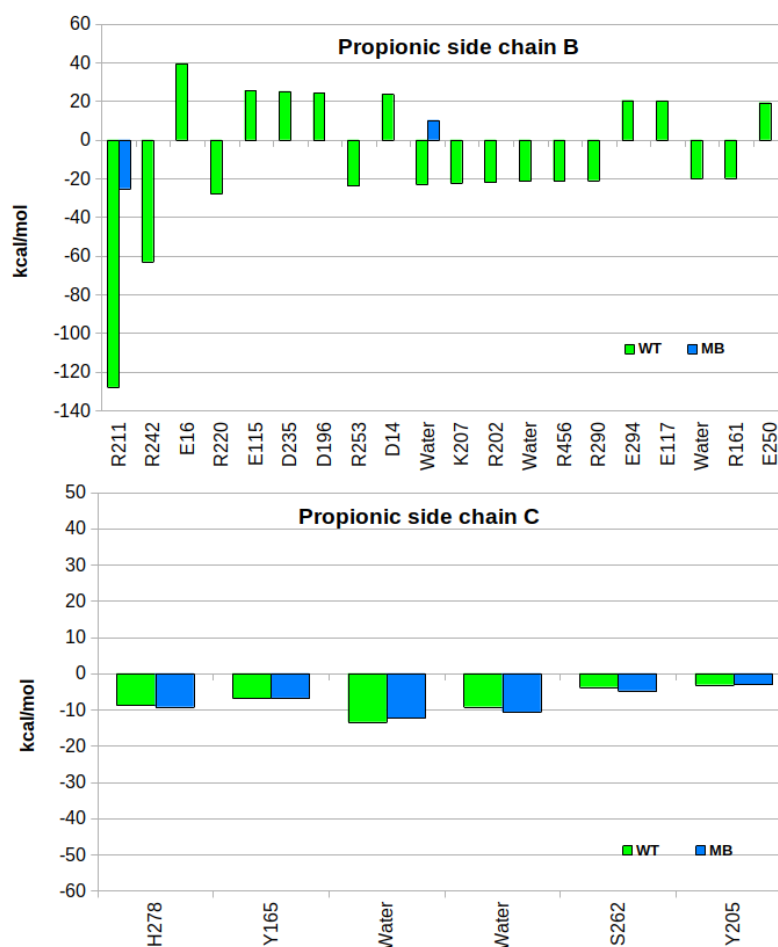


Figure 3.5: Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the PsB (top) and PsC (bottom) computed for models Agp2-BV (green) and Agp2-BVM-B (blue) in the Pfr state.

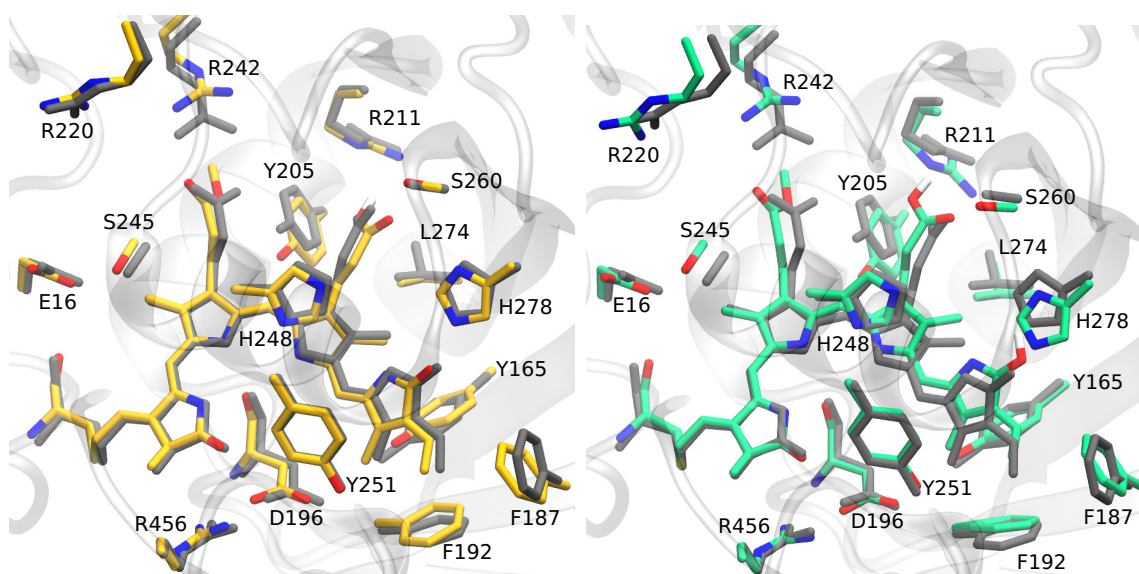


Figure 3.6: Structures of the chromophore binding pockets of the (left) Agp2-BVM-B-keto (orange) and (right) Agp2-BVM-B-enol (green) in the Meta-F(P) state. The optimized QM/MM structures are compared with the optimized QM/MM structure of Agp2-BV in the Meta-F(P) state (gray).

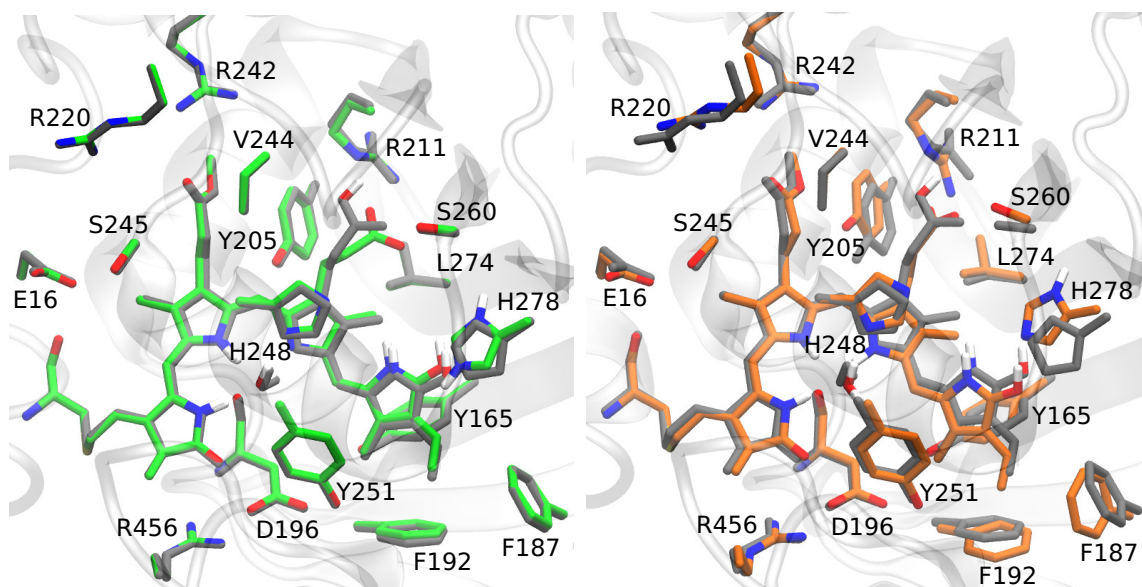


Figure 3.7: Structures of the chromophore binding pockets of photoproduct states of Agp2-BVM-B-enol with (left) cationic His278 (green) and (right) charge-neutral His278, protonated at N δ (orange). The two structural models for Meta-F(D) are compared with the QM/MM-optimized structure of Agp2-BVM-B (enol) in the Meta-F(P) state carrying a protonated PsC (gray).

24 substitutions of the Agp2-PAiRFP2 variant to the WT protein. In this case, a very small RMSD value (0.2 Å) for the BV cofactor was found for both the Agp2 WT model of Meta-F and the Agp2-PAiRFP2 mutant (Figure 3.6). The structural model that mimics the final photoproduct of Agp2-BVM-B is named here as the Meta-F(D) state. This model was modified according to the structural differences determined for the chromophore, in the case of BVM-B in its enolic form with a deprotonated PsC and His278 protonated (His278p) (Figure 3.7). Minor structural changes of the chromophore binding pocket in the Meta-F(D) (see Table 3.1) state were observed in comparison with the Meta-F(P) structural model. The reorientation of deprotonated PsC leads to the formation of a hydrogen bond with Ser262 and with a water molecule, separated by 2.59 and 2.64 Å, respectively.

The protonated His278 (positively charged) can modify the hydrogen-bonded water network in the CBP, this is reflected in the attractive electrostatic interaction between the cationic His278p and the anionic PsC, calculated by PIE ($\sim$ 30 kcal/mol). In contrast, a significant repulsion interaction is found between the PsC and ring D (Figure 3.9, red columns). It is worth noting that, in general, the modification of the total charge of PsC or PsB by esterification/protonation can distort drastically the intermolecular interactions within the protein matrix. Regarding the Meta-F(D) state of Agp2-BVM-B, the effect of the deprotonated PsC (negatively charged) and the cationic His278p in the CBP is reflected in the weakening of the electrostatic interaction with Asp196 (Figure 3.9).

In order to identify the essential role of the His278 in the CBP, a Meta-F(D) model of Agp2-BVM-B with the His278d in its neutral form was proposed (Figure 3.7). Unlike the model with His278p, no significant conformational changes were observed as a result of the formation of a strong salt bridge between Arg211 and PsC. The strength of the interaction energy between PsC and Arg211 is nearly doubled (Figure 3.9) as a result of the decreasing

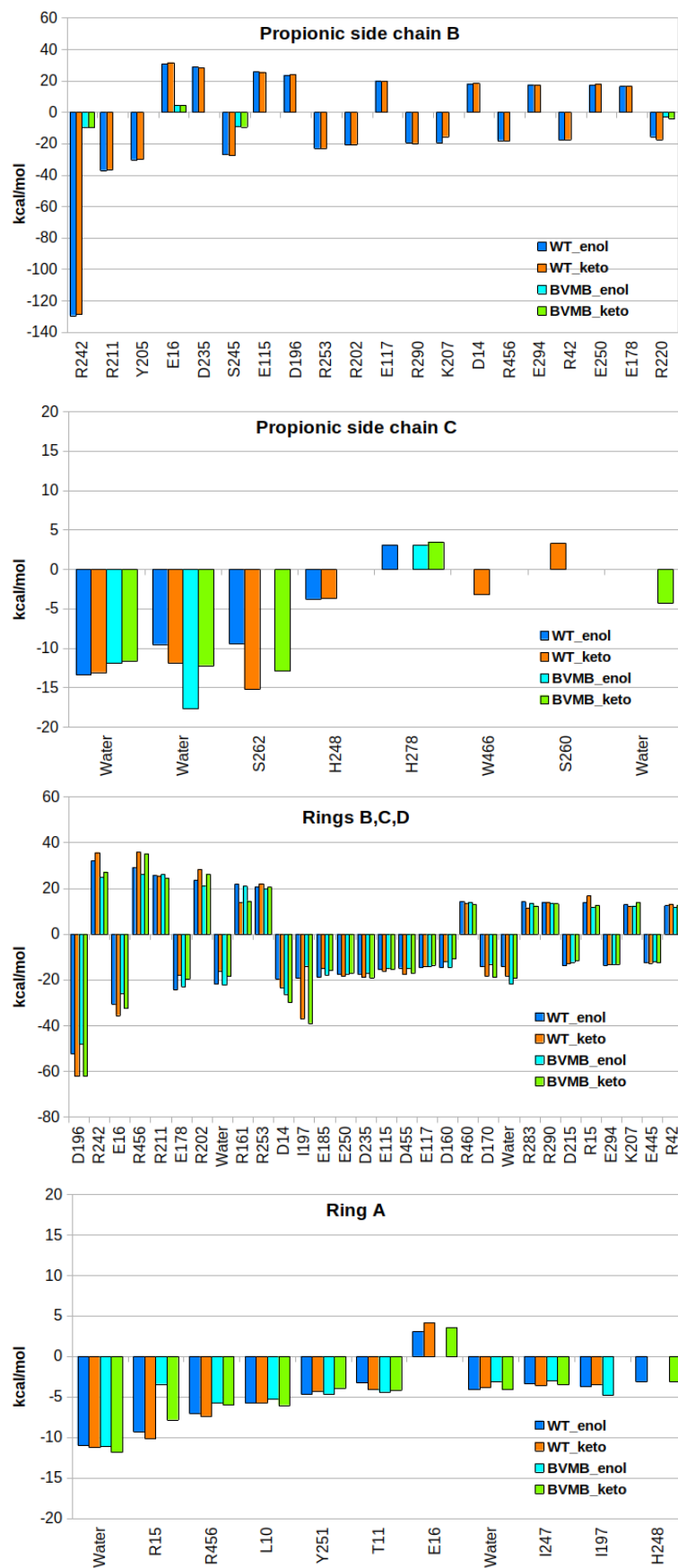


Figure 3.8: Pair Interaction Energies (PIE) in kcal/mol between protein matrix and rings B, C, D and ring A computed for the models Agp2-BV-enol (light green), Agp2-BV-keto (green), Agp2-BVM-B-enol (light blue) and Agp2-BVM-B-keto (blue).

of the O(PsC)···N(R211) distance (2.27 Å), compared to the initial optimized Meta-F(P) structure (5.52 Å). The interaction energy with the His278d was not significant.

Overall, the optimized enolic form of the Agp2-BVM-B model in the Meta-F(D) state with His278d, contributes to a greater destabilization of the chromophore binding pocket (Figure 3.7). This destabilization is reflected in the displacement of His278 that allows the formation of a hydrogen bond with the oxygen of ring D (2.70 Å), as well as movements of Tyr205, Ser262, and Leu274 as a result of the displacement of Arg211 toward PsC, and perturbations of the water network connecting ring D and PsC. Furthermore, a slight reorientation of the pyrrole water is observed due to the change in the protonation pattern of rings B and C.

Photoconversion of Pfr in Agp2-BVM-B. In general, the photoconversion process (Pfr → Meta-F → Pr transition) in bathy phytochromes involves the deprotonation of PsC, the protonation of His278, the secondary structure change of the tongue, and the formation of a keto-enol tautomeric equilibrium of the chromophore in the Pr state. From the Pr state, a thermal back-isomerization of the chromophore to Pfr is initiated[87].

Regarding the Agp2-BVM-B structure, the reaction sequence of the phototransformation is significantly altered. In the case of the intermediate Meta-F(P) (see Table 3.1) state, the BVM-B chromophore is found as an enol tautomer, this involves the migration of a proton of N-H rings B and C to the ring D carbonyl. The BV chromophore exclusively exists in the "normal" keto form of the ZZZssa configuration. This finding can be explained in terms of QM/MM calculations that afforded an energy lower by 13 kcal/mol for the enol than for the keto tautomer of BVM-B in the Meta-F(P) state (The QM/MM calculations were performed by Duc Nguyen). Interestingly, the tongue structure of Agp2-BV and Agp2-BVM-B is the same and exhibits a largely α -helical structure as in Pfr.

The reaction sequence from Meta-F(P) state to the final photoconversion products of both Agp2-BVM-B and Agp2-BV, involves the deprotonation of PsC. Nevertheless, Unlike the typical α -helical to the β -sheet transition found in Agp2-BV, in the case of Agp2-BVM-B, only the degradation of the coiled loop region of the tongue was observed[77]. Here, it is important to note that the neutralization of the charge on PsB leads to a drastic decrease of the pair interaction energy with the Arg242 and Glu16, whilst the interactions with Arg211, Tyr205 are removed, and even with His248 (Figure 3.9). Furthermore, the esterification of PsB has even a significant effect on the interaction of ring A and, to a smaller extent, of rings B, C, and D, with the protein environment. For instance, the interactions of ring A with the Arg15, Glu16, and His248 are significantly reduced. Based on the above, it can be concluded that not only the stabilization of the enolic form of the chromophore but also the failure to induce the complete secondary structure transformation of the tongue is due to the altered electrostatics in the chromophore binding pocket. This suggests that the long-range effects[91] could also play a significant role in the structural change of the tongue.

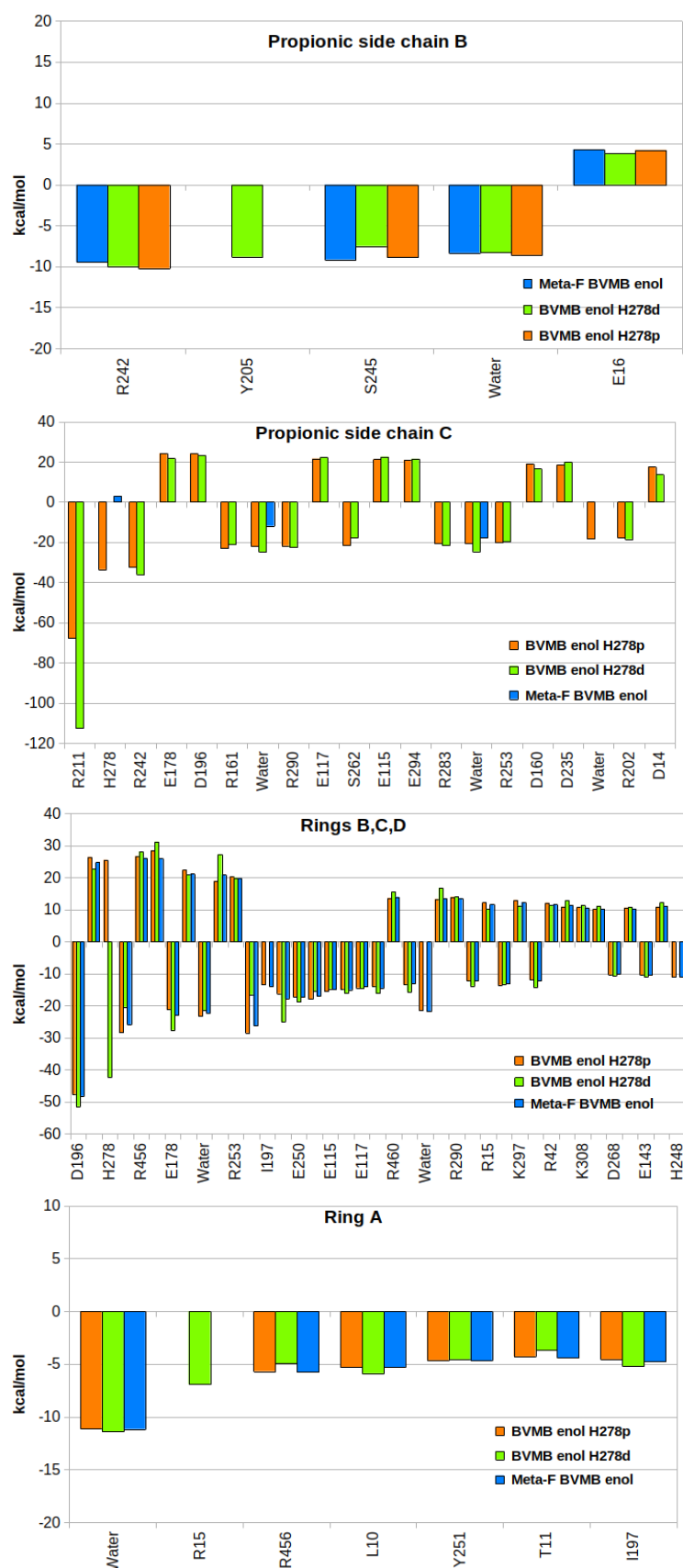


Figure 3.9: Pair Interaction Energies (PIE) in kcal/mol between protein matrix and the propionic side chains, rings B,C,D and A calculated for the models Agp2-BVM-B-enol in the Meta-F state with His278 doubly protonated and neutral His278d (protonated at delta position), compared to the Meta-f state (enol) (see Figure 3.7)

Chromophore Isomerization in Agp2-BVM-B. Regarding the final photoproduct of Agp2-BVM-B, i.e., Meta-F(D), significant retardation of the reversion to Pfr was found by Fernandez et al.[77]. Furthermore, based on a recently proposed mechanism for the dark reversion of bathy phytochromes[87], this suggests that the enolic state of the chromophore is best described by enol-His278d, which affects thermal isomerization of the C-D methine bridge. The theoretical calculations support this idea demonstrating the crucial role of PsC deprotonation and, as well as the protonation state of His278 for determining the reactivity of the chromophore. The protonation of His278 favors the reorientation of PsC toward ring D, triggering a rearrangement of the hydrogen bond network. This produces the lack of attractive interactions between His278p and ring D. In contrast, strong attractive interactions with ring D, are predicted for the neutral His278d as shown by the reversal of the PIEs (Figure 3.9). In this model, His278d forms a hydrogen bond with the ring D oxygen and favors the formation of a salt bridge between the deprotonated PsC and Arg211. The calculations predict an enol-His278d configuration more stable by 7 kcal/mol with respect to the enol-His278p, which would account for the experimental findings of a strongly hindered isomerization of the C-D methine bridge.

Photoinduced Processes in Agp1-BVM. For Agp1, the structural model does not reveal significant structural perturbations after the chemical modifications of the propionic side chains of ring B or C as reflected by the relatively low RMSD values predicted for the chromophore (~ 0.6 Å) (Figure 3.2). According to Fernandez et al.[77], in the particular case of Agp1, the esterification of either PsC or PsB allows the formation of the “normal” keto tautomer, however, only one fraction is capable of undergoing a photochemical ZZZ $\rightarrow$ ZZE transformation to the final state that is structurally very similar to the Pfr state of Agp1-BV. It is important to mention here that the role of the two propionic side chains is less clear since, in contrast to the Agp2, the PsC has not been identified as a site for transient protonation. Furthermore, regarding the secondary structure change of the tongue in Agp1-BVM, the alterations of the local electrostatics due to BV esterification have a less, compared to Agp2, effect on the reaction dynamics of the chromophore and the protein, since the structural change of the tongue is found much more advanced than in Agp2-BVM-B[77].

3.5 Conclusions

The esterification of the BV in prototypical (Agp1) and bathy (Agp2) phytochromes has relatively small consequences on the structure of the chromophore binding pocket. Nevertheless, local electrostatics changes in the chromophore binding pocket were identified by theoretical calculations as the main origin for the perturbations of the reactivity of the photosensors. In the case of Agp2, the Agp2-BVM-B adduct does not have the capacity to carry out the complete secondary structure change of the tongue, furthermore, the deprotonation of PsC along with the protonation state of His278 play a key role in the slowdown of the thermal back-reaction for the final photoproduct. For Agp1, the electrostatic changes have a less severe effect compared to the strong changes observed in Agp2.

3.6 Simulation of Raman spectra using the fragment molecular orbital method

Simulations of infrared (IR) and Raman spectra of chromophore molecules can provide valuable information about the structure of the chromophore and its interactions with the protein environment in the different states of phytochromes photocycle. However, the inclusion of the effect of the protein environment in the chromophore spectra is crucial for getting a more realistic description and a good agreement with the experimental spectra. In this work, The FMO method within the frozen domain with dimers (FDD) approach[33, 92] was applied to the cyanobacteriochrome (CBCR) structure (Slr1393) in the Pg state. CBCRs are photosensory proteins found in cyanobacteria that use a tetrapyrrole chromophore (bilin) to respond to changing light conditions. Furthermore, unlike phytochromes, CBCRs require only one GAF domain to bind the chromophore and photoconvert between two parent states (Pr $\rightarrow$ Pg). Therefore, the relatively small size of their functional units makes them attractive candidates for performing computational studies. To the best of our knowledge, this is the largest quantum chemical hessian and Raman calculation reported for a protein.

Computational details. In this study, The FMO/FDD method was used to compute: optimized geometry, the Hessian matrix, and Raman intensities of the Slr1393 cyanobacteriochrome in the Pg state (PDB entry 5M82[93]). Calculations were carried out with the FMO code version 5.3[68] as embedded in General Atomic and Molecular Electronic Structure System (GAMESS) which is a general *ab initio* quantum chemistry package[69]. The protonation states of all titratable residues at pH 7 were estimated by solving the linearized Poisson-Boltzmann equation of the electrostatic potential[42]. These computations were performed with the Karlsberg2+ software package[60].

Fragmentation of the Slr1393 cyanobacteriochrome in Pg state. The structure of the Slr1393 cyanobacteriochrome in the Pg state was divided into 230 fragments (one fragment per residue) using Fragit program[94]. The phycocyanobilin (PCB) chromophore consisting of two propionic side chains and four pyrrole rings was fragmented manually. In order to preserve the chemical identity of the chromophore and reduce the computational cost, we performed the fragmentation proposed in Ref[95]. As a result, both propionic side chains, B (propB) and C (propC) are treated as a single fragment. The high electron delocalization of PCB-molecule can be taken into account by joining rings B, C, and D in one fragment (rings-BCD) separated from ring A which, along with Cys528, form the fourth fragment (ring-A-Cys). Three methyl groups were additionally treated each one as a single fragment.

FMO/FDD is a multilayer approach that divides a large molecule into different layers. Figure 3.10 shows the Slr1393 cyanobacteriochrome structure divided into two layers, the first layer (L1) contains the frozen domain F and the second layer (L2) contains the active domain A and the polarizable buffer domain P. The active and polarizable buffer domains contain 622 and 2205 atoms, respectively. The rest of the atoms are assigned to the frozen domain. In all FMO/FDD calculations, STO-3G and 6-31G(d) will be used as the basis set for layers L1

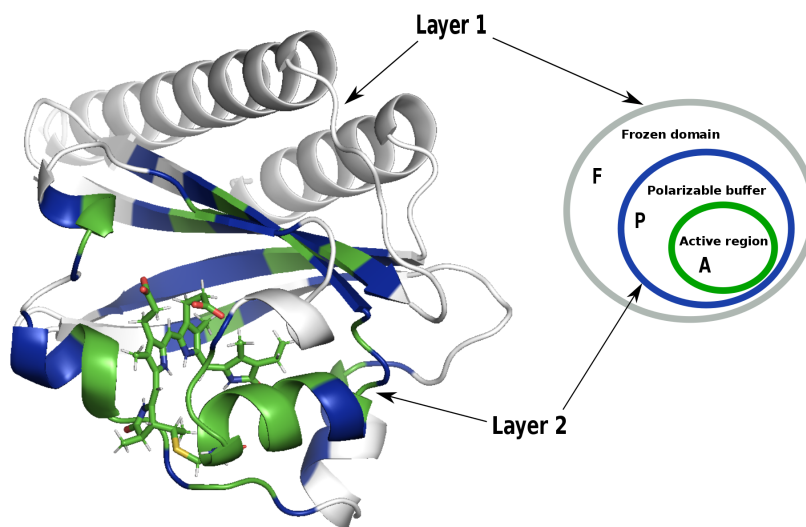


Figure 3.10: Cyanobacteriochrome structure divided into two layers (L1, and L2). The electronic state of fragments in domain F (Layer 1) is frozen during a geometry optimization. Fragments that belong to the polarizable and active region (Layer 2) share the same basis set, nevertheless, only fragments that belong to the active region are allowed to move during the optimization.

and L2, respectively. This method is denoted as FMO-HF-D/FDD/STO-3G:6-31G(d), where HF corresponds to Hartree–Fock method, and D denotes a dispersion model[96]. Hereinafter referred to as FMO/FDD. Regarding the analysis of the FMO/FDD spectra, the Raman active normal modes were localized by using the unitary transformation method[97], and the corresponding transformed vibrational modes and their frequencies were used for the peak assignment. The Movipac program[98] was used for the unitary transformations. However, to be able to use the GAMESS outputs for the analysis, force constant and dipole moment must be extracted in a file, then, a Perl script is executed to convert the GAMESS format into the Movipac format. The Perl script was provided by Hiroya Nakata¹. For the presentation of the FMO/FDD Raman spectrum, the band shapes of the individual modes were described by Lorentzian functions using the broadening parameter of 12 cm^{-1} for the line width. In order to compare with the experimental spectrum, frequencies were scaled by 0.8953[99].

Parallelization and scaling of FMO. One of the most significant benefits of the FMO method is that it can be used on PC clusters or petascale supercomputers. For example, it has been demonstrated that the FMO code in GAMESS program scales well up to 131072 processor (Blue Gene/P supercomputer) to obtain the MP2 atomic forces for a system with more than 4000 atoms, with the 6-31G(d) basis set, in approximately 7 min[100]. For the parallel execution, FMO can be used with the generalized distributed data interface[101], in which each subsystem (fragments, pairs, or triples) is calculated on a group of central processing units (CPU) cores. The number of groups depends on each subsystem. For the FMO/FDD calculations of the cyanobacteriochrome structure (2827 atoms), we used about 400 CPU cores with the computational power of the North German Association for High and Performance Computing (HLRN).

¹Tokyo Institute of technology

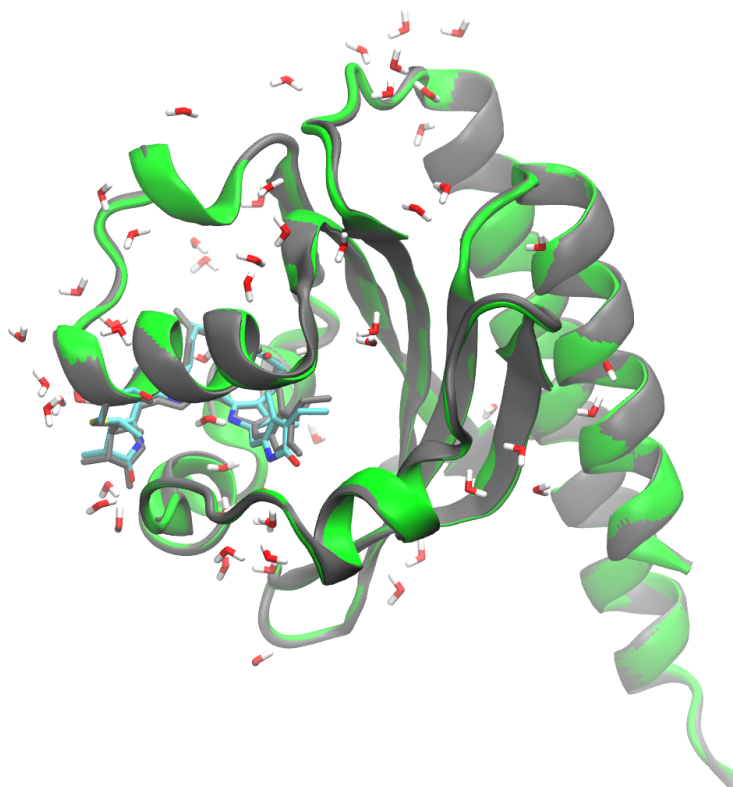


Figure 3.11: The optimized FMO/FDD structure of the cyanobacteriochrome structure (Slr1393, PDB entry 5M82[93]) in the Pg state (green) is compared with the crystal structure (gray). The rmsd was calculated after alignment of backbone heavy atoms of the optimized structural model with respect to the crystal structure (0.38 Å).

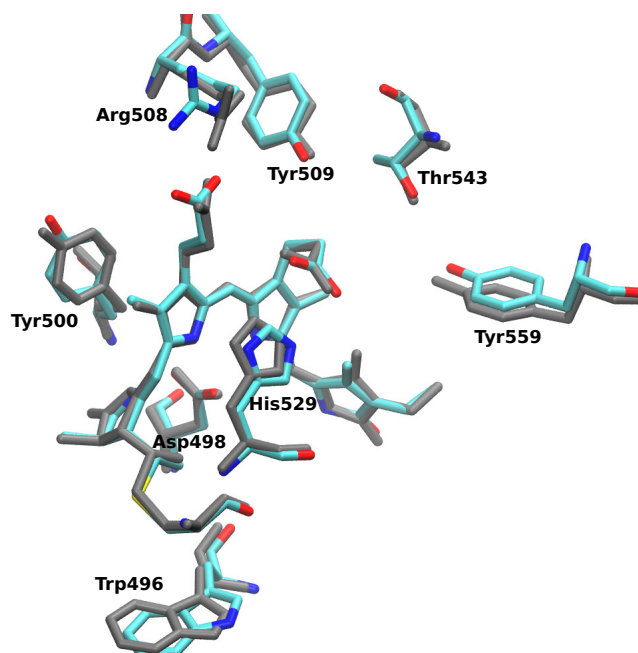


Figure 3.12: PCB molecule of Slr1393 in Pg state and residues in the vicinity of the PCB chromophore of the optimized structural model (color cyan) compared to the crystal structure (color gray). The rmsd for the PCB molecule was calculated for the heavy atoms compared to the PCB in the crystal structure (0.34 Å).

Results and discussion. The geometry of the structural model of the cyanobacteriochrome structure (Slr1393) in the Pg state was optimized using the FMO/FDD method, including 622 in the active region, which can change their atomic positions during the optimization. The root-mean-square deviations (rmsd) of the heavy atoms were computed for the FMO/FDD optimized structure after the alignment of backbone atoms of the protein with respect to the crystal structure. The rmsd value around 0.38 Å predicted for the optimized structural model (Slr1393) reflect only minor deviations of the atoms belonging active region (see Figure 3.11) with respect to the experimental crystal structure (PDB entry 5M82).

The root-mean-square-deviation (RMSD) value obtained for the heavy atoms of the PCB chromophore (optimized structure) with respect to the crystal structure is relatively small (0.34 Å). Figure 3.12 shows the PCB molecule and some residues located in the chromophore binding pocket of the optimized structural model compared to the crystal structure. The most significant changes observed correspond to the reorientation of the side chains of the Trp496, His259, Arg508, and Tyr559 residues. In particular, the movement of the guanidinium group of Arg508, since it has been demonstrated that this amino acid is responsible for the stabilization of the chromophore binding pocket[95]. The strong electrostatic interaction between Arg508 and propionic side chain B (Arg508···PsB 2.69 Å crystal structure) is weakened upon the geometry optimization (Arg508···PsB 2.86 Å optimized structure). Additionally, the hydrogen bond formed between the highly conserved residue His259 and propionic side chain C is weakened as a result of the optimization, the hydrogen bond distance is increased from 2.97 Å (His259···PsC crystal structure) to 3.10 Å (His259···PsC optimized structure).

In the particular case of Asp498, this residue forms three hydrogen bonds with the NH group of the pyrrole rings A, B, and C. The effect of including the chemical environment of PCB molecule in the geometry optimization is reflected in the decrease of the hydrogen bond distances between Asp498 and pyrrole rings A and B with respect to the crystal structure, from 2.83 Å (Asp498···ring A) and 2.89 Å (Asp498···ring B) to 2.88 Å (Asp498···ring A) and 3.02 Å (Asp498···ring B), respectively. In contrast, a stronger hydrogen bond interaction between Asp498 and pyrrole ring C is observed, varying from 2.87 Å (Asp498···ring C crystal structure) to 2.82 Å (Asp498···ring C optimized structure). In CBCRs, unlike phytochromes, the Asp498 residue plays an essential role in the stabilization of the pyrrole rings A, B, and C of the chromophore by forming three hydrogen bonds. It is important to mention that in phytochromes this role relies on the "pyrrole water". Interestingly, the geometry optimization has an important effect on the structure of the chromophore, which is reflected by the decrease of the dihedral angle value that involves the methine bridge around the double bond (C=C) with respect to the crystal structure (see Figure 3.12). This value varies from 38° in the crystal structure to 17.02° in the FMO/FDD optimized structure. This variation in the torsional angle is the result of the relaxation of the residues located in the chromophore binding pocket.

The experimental RR spectrum of the cyanobacteriochrome in the Pg state is compared to the one calculated with the FMO/FDD approach (see Figure 3.13). The experimental spectrum exhibits a broadband corresponding to the C=C stretching mode (1648 cm⁻¹), this

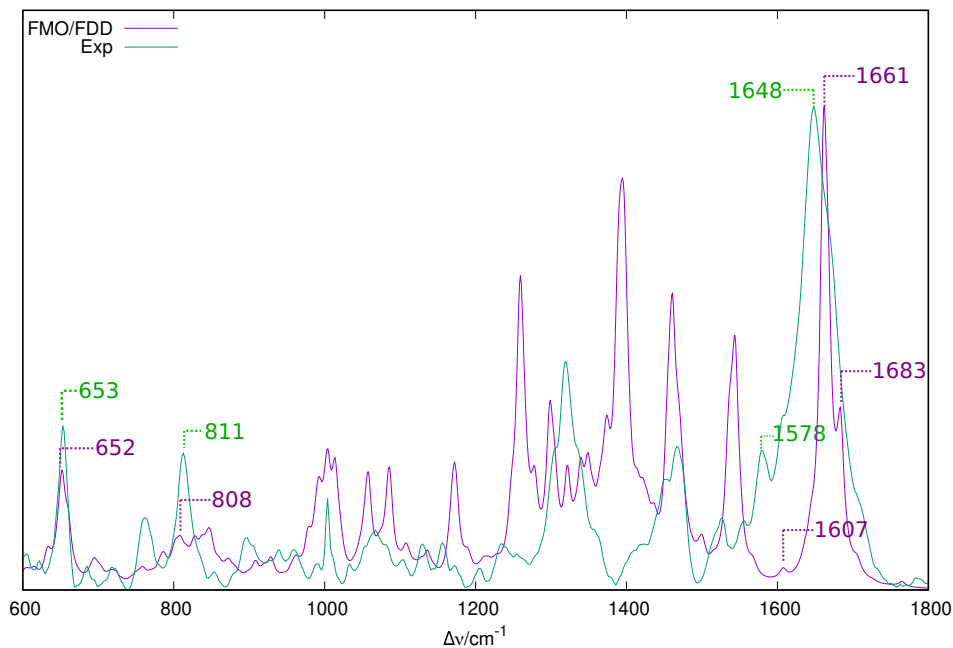


Figure 3.13: Experimental (green lines) and FMO/FDD calculated (purple lines) Raman spectra of Slr1393-g3 in the Pg state. The experimental spectrum was taken from the Ref.[102]

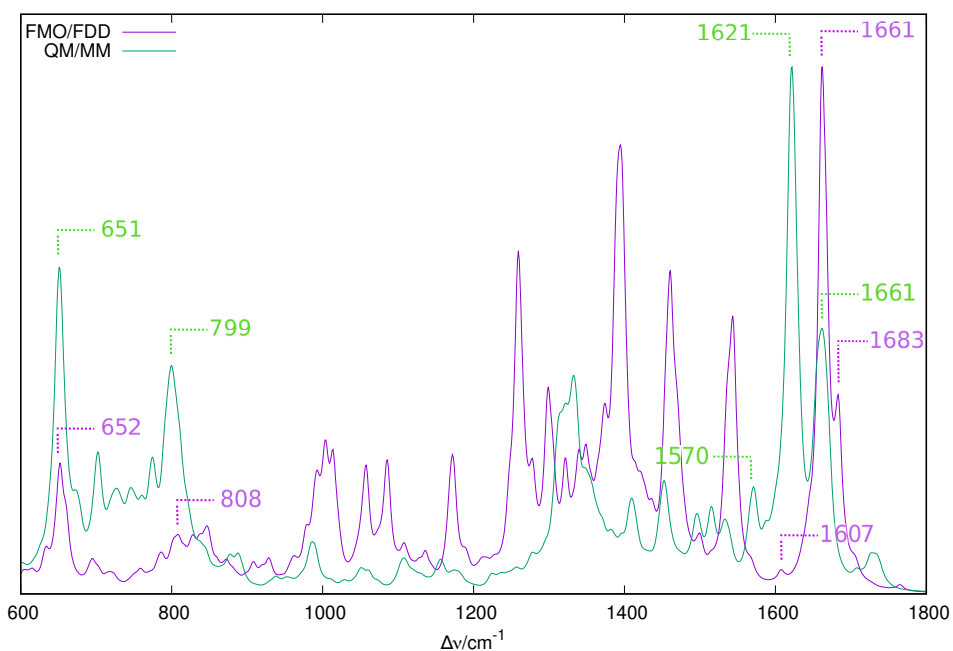


Figure 3.14: QM/MM (green lines) and FMO/FDD calculated (purple lines) Raman spectra of Slr1393-g3 in the Pg state. The QM/MM spectrum was provided by Giovanni Battocchio[102]

band is reflected in the FMO/FDD Raman spectrum located at 1661 cm^{-1} with a very similar intensity. Furthermore, according to the experimental spectrum, the HOOP mode and the ring D oop mode is located at 811 and 653 cm^{-1} , respectively. The calculated FMO/FDD spectrum reproduces very well the intensity and frequency of the ring D oop mode (652 cm^{-1}). With respect to the HOOP mode, the FMO/FDD spectrum shows a good agreement with the experimental frequency value (808 cm^{-1}), however, this band shows a minor intensity in the calculated spectrum. The band at 1578 cm^{-1} in the experimental spectrum is characteristic of a protonated chromophore in the Pg state. The band is located at 1607 cm^{-1} in the FMO/FDD spectrum. In general, the FMO/FDD spectrum shows a reasonable agreement with the experimental Raman spectrum.

In the Pr state, the ring A C=O group and conserved residue Trp496 form a hydrogen bond interaction. In contrast, this interaction is removed in the Pg state of CBCR, which is reflected in the optimized FMO/FDD structure (Figure 3.12). This new configuration of ring A (out-of-plane) contributes to decreasing the conjugation by uncoupling the A-B C=C stretching mode from the rest of the chromophore. This is reflected in the band located at 1683 cm^{-1} of the FMO/FDD spectrum. The decrease in the conjugation length has already been proposed by Wiebeler et al. based on a computational study of excitation energies of the cyanobacteriochrome (Slr1393)[103, 104].

The FMO/FDD calculated Raman spectrum of the cyanobacteriochrome in the Pg state is also compared to the one calculated with the QM/MM approach (see Figure 3.14). Both QM/MM and FMO/FDD calculated Raman spectra exhibit a broadband with a very similar intensity corresponding to the C=C stretching mode at 1621 cm^{-1} and 1661 cm^{-1} , respectively. Interestingly, the bands corresponding to the HOOP mode and the ring D oop mode calculated by both QM/MM and FMO/FDD approaches are in good agreement with the experimental Raman spectrum. However, these bands show a minor intensity in both the experimental RR spectrum and the FMO/FDD calculated spectrum. The characteristic band of a protonated chromophore in the Pg state is well reproduced by both QM/MM and FMO/FDD approaches, located at 1570 cm^{-1} and 1607 cm^{-1} , respectively. Overall, the FMO/FDD spectrum shows a reasonable agreement with the QM/MM calculated Raman spectrum, nevertheless, it is important to mention that the QM/MM calculated Raman spectrum was obtained by using density functional theory (DFT) calculations, therefore, the accuracy of the FMO/FDD calculated Raman spectrum can be improved by using DFT in combination with the FMO/FDD approach.

It is noteworthy that the experimental Raman spectrum takes into account the protein flexibility, thus, in order to include the conformational sampling in the Raman calculated spectra, the FMO/FDD approach can be combined with molecular dynamic (MD) simulations, taken several MD frames for evaluating the averaged spectrum. In general, the FMO/FDD spectrum shows a reasonable agreement with both the experimental and the QM/MM calculated Raman spectrum.

4

Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

4.1 Introduction

Molecular dynamics (MD) simulations have become an invaluable tool for studying the conformational fluctuations of proteins such as phytochrome structures. These photoreceptor proteins contain a bilin chromophore that allows the photoconversion between red-absorbing (Pr) and far-red-absorbing (Pfr) states. During the photocycle, the chromophore molecule and residues located in its surroundings undergo conformational changes which trigger functions of response. Thus, the study of the conformational fluctuations of the bilin chromophore becomes an essential step for achieving a better understanding of the phytochromes photocycle. However, the results of MD simulations based on molecular mechanics largely depend on the quality of force fields parameters, consequently, the study of the conformational fluctuations of bilin chromophore requires the optimization of accurate parameters. In order to address this issue, we derived new non-polarizable force fields parameters for the biliverdin (BV) chromophore in both Pfr and Meta-F states using the CGenFF parameterization protocol. The new CGenFF parameters of BV molecule can be used in conjunction with the highly optimized CHARMM all-atom force field parameters for proteins.

4.2 Methods

The optimization of biliverdin (BV) chromophore parameters followed the CGenFF parameterization protocol described by Vanommeslaeghe et al[105]. with the help of the FFParam-GUI python package[106], which is a tool designed for facilitating the

4. Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

parametrization of small molecules. Quantum mechanical (QM) calculations were performed with the Gaussian 09 program[107] while molecular mechanics (MM) calculations were done using the CHARMM program[108]. All molecular dynamics simulations were performed using the NAMD 2.13 code[109].

Parameterization details. The BV geometry was taken from Agp2 phytochrome structure in the Pfr state[86]. Nevertheless, it is important to mention that the BV chromophore adopts a ZZEssa stereochemistry in the Pfr state whereas in the Meta-F state adopts a ZZZssa stereochemistry. Although the BV molecule of Agp2 phytochrome in the Pfr state was chosen as a model compound, the aim is to develop BV parameters that enable the study of both Pfr and Meta-F states. Furthermore, the BV molecule is found attached to Cys13 via pyrrole ring A, therefore, it is also necessary the derivation of linking parameters between Cys13 (protein) and BV molecule.

Target Data. In order to reduce the computational cost the BV chromophore was divided into smaller fragments following the fragmentation schema proposed by Kaminski et al[65]. These fragments were used for the optimization of the bond lengths, bond angles, improper dihedral parameters and the nonbonded electrostatic parameters. Figures 4.1 and 4.2 show the division of chromophore into different fragments, additionally, six new atom types were created within CGenFF force field for the correct definition of torsional potentials of BV.

CGenFF program. The first step for the derivation of new parameters consists in the creation of a mol2 file of the molecule to parametrize. Then one can obtain CHARMM Additive parameters from the CGenFF web site[110] using automated algorithms[111, 112] as initial parameters. In the next step, the FFParam-GUI python package was used for generating all QM input files needed. The energy minimizations for all fragments were performed at the MP2/6-31G(d) level of theory using default tolerances.

Partial atomic charges optimization According to CHARMM protocol, partial charges need to be optimized in order to reproduce interactions with water molecules obtained from QM calculations. To this end, the FFParam-GUI package was used to generate QM input files and locate water molecules around each fragment in order to form hydrogen bonding interactions with all donor/acceptor groups. Afterwards, each fragment/water interaction distance was optimized at the HF/6-31G(d) level of theory. In the last step, partial charges were fitted using a Monte Carlo simulated annealing (MCSA) algorithm which was implemented in the FFParam-GUI package for the automated fitting of partial atomic charge, For each fragment an initial set of partial charges was assigned by analogy to those of similar CGenFF atom types using the CGenFF web site.

MD simulations with NAMD. The Agp2 phytochrome structures in both Pfr and Meta-F state were modeled with the plug-in of VMD and solvated in a TIP3P water box applying periodic boundary conditions. In this work, we used the CHARMM36 force field for protein, and the new CHARMM General Force Field (CGenFF) parameters derived for

the BV molecule. The MD simulations were performed with the software NAMD using a 2 fs time step and Langevin dynamics with friction constant $\beta = 1 \text{ ps}^{-1}$ at 300K temperature. The propionic side chain of ring C were protonated for the MD simulations in both Pfr and Meta-F state.

Table 4.1: Interaction Energies (kcal/mol) and Distances ($\text{\AA}$) between water and fragment RDV Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).

Interaction	E_{int}		ΔE_{int}	R_{min}		ΔR_{min}
	QM	MM		QM	MM	
O_B_WAT_LN1	-5.902	-6.432	-0.530	2.03	1.77	-0.26
O_B_WAT_LN2	-5.902	-6.432	-0.530	2.03	1.77	-0.26
O_B_WAT_LP2	-8.191	-7.892	0.299	1.98	1.75	-0.23
H_B_WAT_0	-6.022	-6.818	-0.796	2.03	1.87	-0.16
H_B_WAT_90	-6.023	-6.905	-0.882	2.04	1.87	-0.17

Table 4.2: Interaction Energies (kcal/mol) and Distances ($\text{\AA}$) between water and fragment RCD Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).

Interaction	E_{int}		ΔE_{int}	R_{min}		ΔR_{min}
	QM	MM		QM	MM	
RCD_H_A_WAT_0	-6.306	-5.345	0.961	2.06	1.91	-0.15
RCD_H_A_WAT_90	-6.125	-5.217	0.908	2.08	1.92	-0.16
RCD_N_B_WAT_LN2	0.478	1.049	0.571	3.38	3.00	-0.38
RCD_H_B_WAT_0	-5.687	-6.505	-0.818	2.04	1.86	-0.18
RCD_H_B_WAT_90	-5.751	-6.732	-0.981	2.05	1.86	-0.19
RCD_O_B_WAT_LP1	-5.725	-4.810	0.915	2.10	2.33	0.23
RCD_O_B_WAT_LP2	-8.629	-8.955	-0.326	1.97	1.74	-0.23

Table 4.3: Interaction Energies (kcal/mol) and Distances ($\text{\AA}$) between water and fragment RBC Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).

Interaction	E_{int}		ΔE_{int}	R_{min}		ΔR_{min}
	QM	MM		QM	MM	
RBC_H_D_WAT_0	-7.060	-10.342	-3.282	2.17	2.03	-0.14
RBC_H_D_WAT_90	-8.068	-10.606	-2.538	2.09	2.01	-0.08
RBC_H_A_WAT_0	-7.061	-10.354	-3.293	2.17	2.03	-0.14
RBC_H_A_WAT_90	-8.069	-10.619	-2.550	2.09	2.01	-0.08

Bonded Parameters An initial set of bonded parameters for each fragment was assigned by analogy using the CGenFF web site. Then, an analysis based on the differences between QM and MM data was done in order to identify those parameters to be optimized. Following the methodology proposed by Kumar et al[106], those parameters that showed differences (QM-MM) greater than $0.02 \text{ \AA}$ for bond lengths, 2° for angles involving non-hydrogen atoms and 4° for angles involving hydrogen atoms) were chosen to be optimized. The parametrization of the bond lengths and valence angles parameters was done iteratively using the equilibrium values obtained at the MP2/6-31G(d) level. For the case of dihedral angles, force constants, phase and multiplicities were optimized to minimize the difference between the MM and QM

4. Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

Table 4.4: Interaction Energies (kcal/mol) and Distances (Å) between water and fragment RAS Obtained at the HF/6-31G(d) and Empirical Levels (CGenFF).

Interaction	E_{int}		ΔE_{int}	R_{min}		ΔR_{min}
	QM	MM		QM	MM	
RAS_O_C_WAT_LP1	-6.862	-7.617	-0.755	1.99	1.76	-0.23
RAS_O_C_WAT_LP2	-7.209	-7.842	-0.633	1.97	1.75	-0.22
RAS_H_C_WAT_0	-5.072	-5.012	0.060	2.03	1.90	-0.13
RAS_H_C_WAT_90	-5.174	-5.125	0.049	2.03	1.89	-0.14

dihedral potential energy scans (PES). (see Figures 4.1-4.2). Standard CGenFF Lennard-Jones parameters were used for all atoms.

Table 4.5: Force constants (k), multiplicity (n) and phase (δ) of dihedral parameters for model compounds.

Dihedral parameter	k (kcal/mol)	n	δ (deg)
Ring-D			
CG2DC3 CG2DC1 CGCPY6 CGCPY5	0.85	1	180.0
CG2DC3 CG2DC1 CGCPY6 CGCPY5	1.13	2	180.0
CG2DC3 CG2DC1 CGCPY6 CGCPY5	0.19	3	0.0
CG2DC3 CG2DC1 CGCPY6 CGCPY5	0.18	4	0.0
Rings-C-D double bond			
CG251O CGCPY2 CG2R51 CGCPY5	3.39	2	180.0
Rings-C-D single bond			
CG2R51 CG251O CGCPY2 CG2R51	1.20	1	0.0
CG2R51 CG251O CGCPY2 CG2R51	2.58	2	180.0
CG2R51 CG251O CGCPY2 CG2R51	0.68	3	0.0
CG2R51 CG251O CGCPY2 CG2R51	0.40	4	0.0

4.3 Results and Discussion

4.3.1 Validation of Force Field Parameters.

The final chromophore parameters were validated by comparing the minimized geometries of each fragments obtained at the QM and MM level of theories (see figure 4.3). All geometry optimizations were done using the crystal structure as a starting point. Following the methodology proposed by Kaminski et al[65], both propionic side chains B and C were removed in order to avoid the formation of hydrogen bonds the BV backbone, since this could alter the validation of the new CGenFF parameters. Figure 4.3 shows the QM and MM optimized fragments of the BV molecule, as well as the rmsd values calculated for each fragment (only heavy atoms) after optimal superposition. Overall, all MM-optimized structures show a good agreement with respect to the QM-optimized structures, this is reflected by the

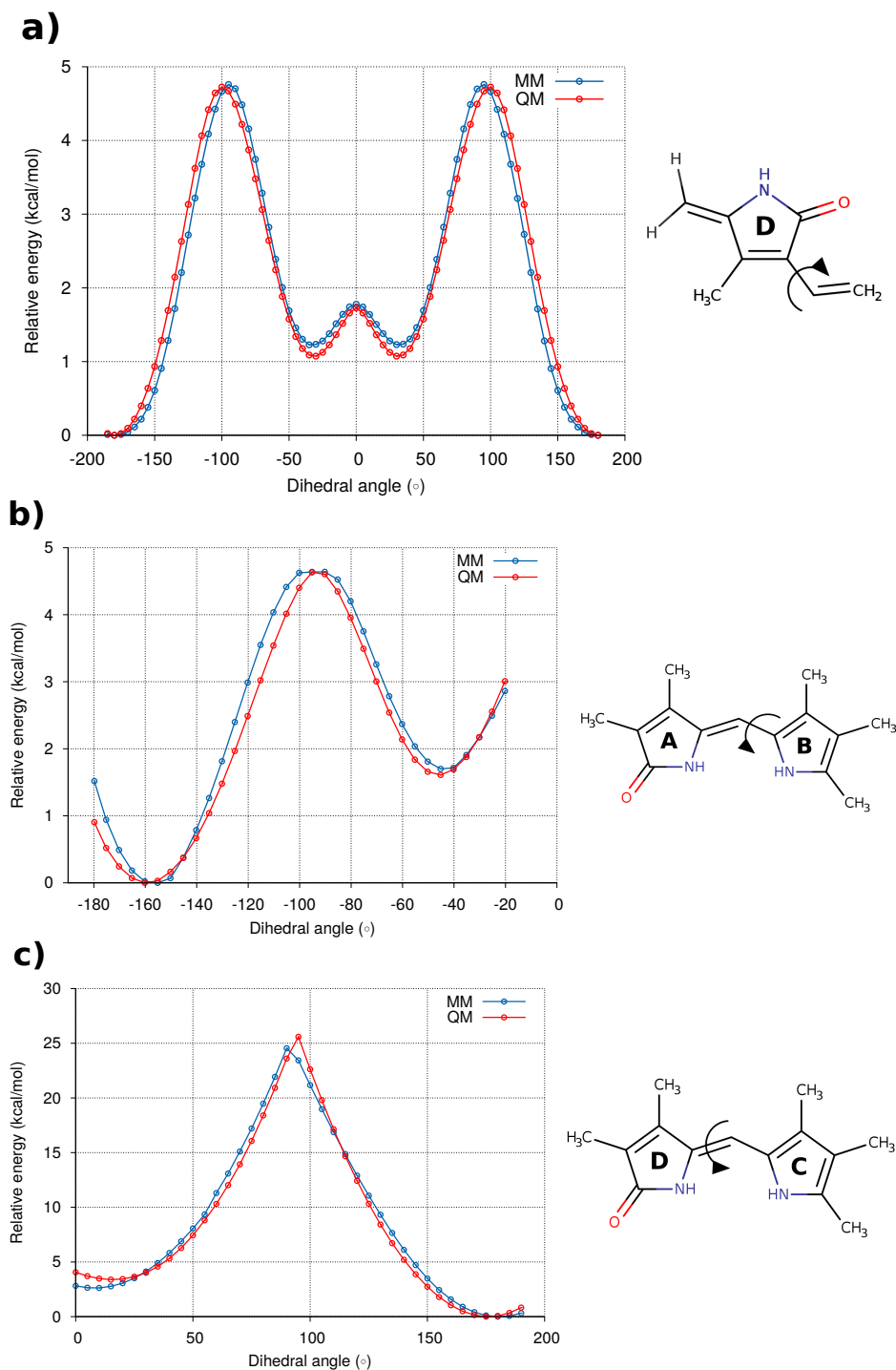


Figure 4.1: Potential energy scans for model compounds. The torsional potential were scanned in steps of 5° . The red curves represent the scans derived at the MP2/6-31G(d) level of theory, and the blue curves denote result from optimized torsion parameters (see Table).

4. Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

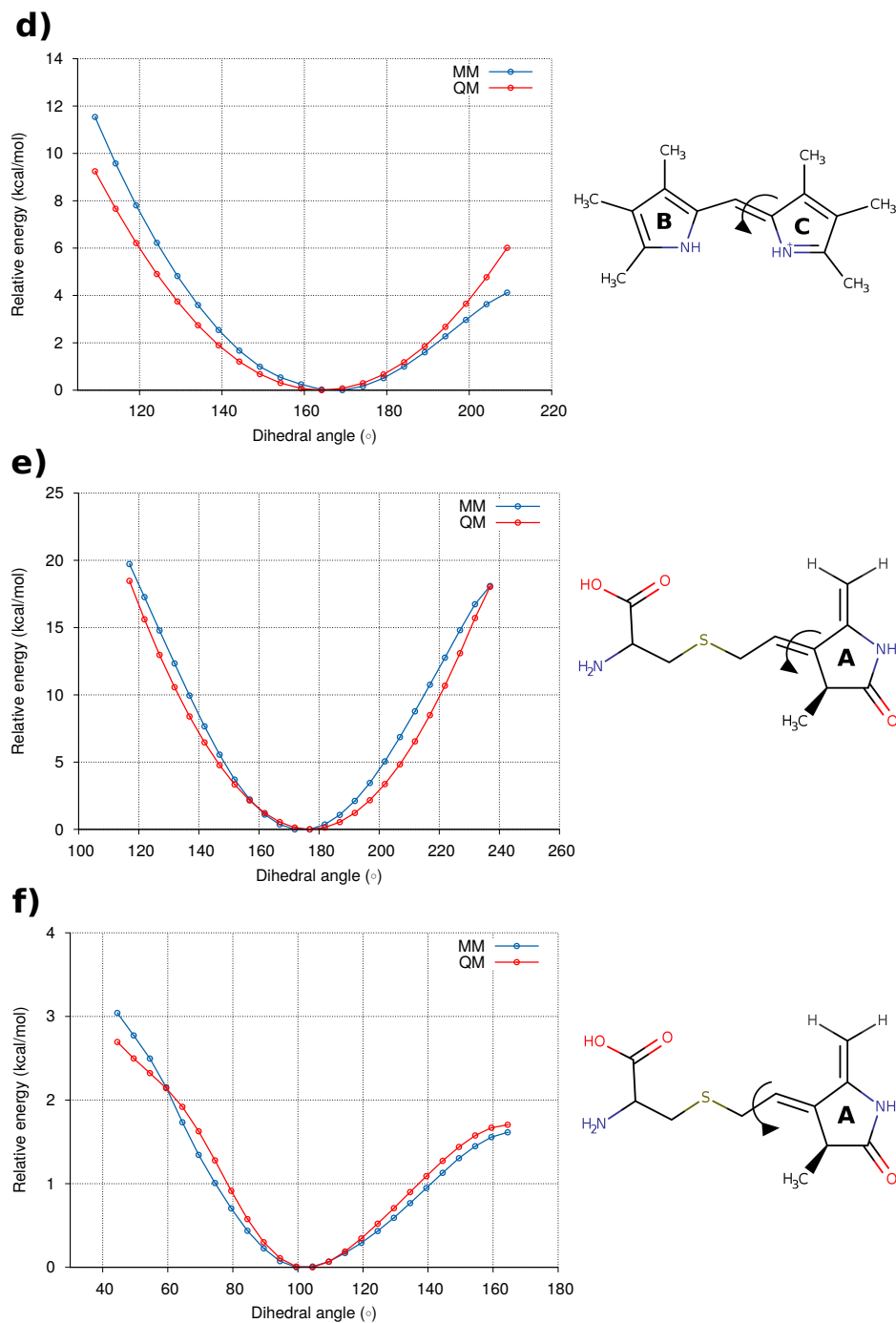
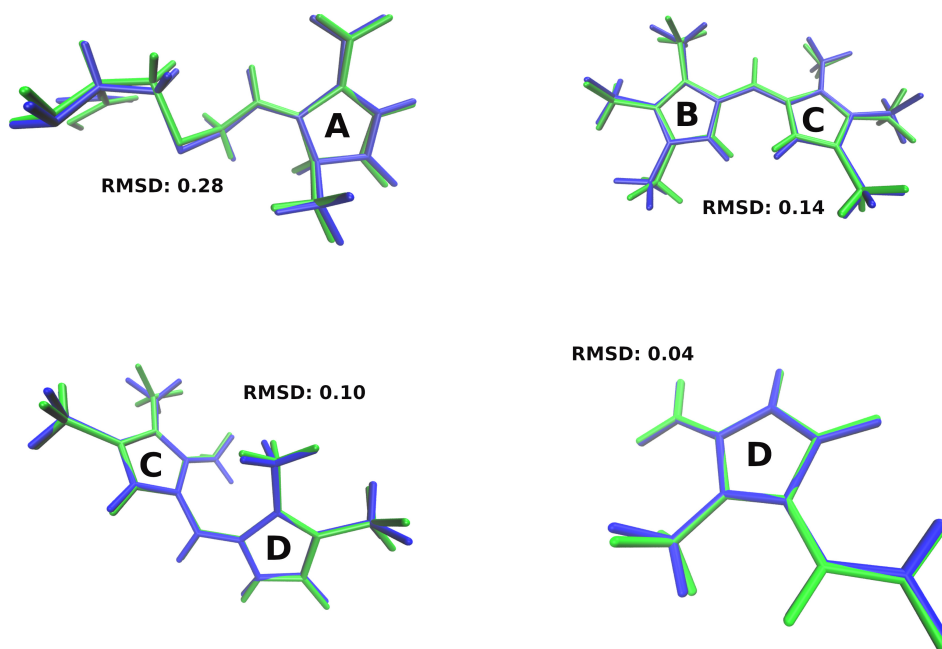


Figure 4.2: Potential energy scans for model compounds. The torsional potential were scanned in steps of 5° . The red curves represent the scans derived at the MP2/6-31G(d) level of theory, and the blue curves denote result from optimized torsion parameters (see Table).

Table 4.6: Force constants (k), multiplicity (n) and phase (δ) of dihedral parameters for model compounds.

Dihedral parameter				k (kcal/mol)	n	δ (deg)
Rings-B-C single bond						
CG251O	CG2DC1	CG2R51	CG2R51	2.25	2	180.0
CG251O	CG2DC1	CG2R51	CG2R51	0.19	3	0.0
Ring-A-Cys $CBC - -C4C$						
CG252O	CGCPY1	CG2DC1	CG321	0.75	1	0.0
CG252O	CGCPY1	CG2DC1	CG321	3.09	2	180.0
CG252O	CGCPY1	CG2DC1	CG321	1.00	3	0.0
Ring-A-Cys $C3C - -SG$						
CGCPY1	CG2DC1	CG321	SG311	0.60	2	0.0
CGCPY1	CG2DC1	CG321	SG311	0.52	3	180.0

**Figure 4.3:** Optimized fragments MP2/6-31G(d) (color blue) and empirical MM structures (color green) of the BV chromophore without propionic side chains. The rmsd values were calculated for heavy atoms (no hydrogen atoms) after optimal superposition.

4. Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

Table 4.7: Atom Types, Labels and Optimized Partial Charges of the BV Molecule

atom	atom type	partial charge	atom	atom type	partial charge
CAC	CG2DC1	-0.12	HV2	HGA5	0.22
HAC	HGA4	0.15	CBC	CG321	-0.33
C1C	CG2R53	0.56	HL1	HGA2	0.18
H2C	HGA1	0.02	HL2	HGA2	0.18
N_C	NG2R53	-0.39	CMC	CG331	-0.25
H_C	HGP1	0.35	HE1	HGA3	0.09
C4C	CG252O	-0.26	HE2	HGA3	0.09
C3C	CGCPY1	-0.19	HE3	HGA3	0.09
C2C	CG3C51	0.31	CMD	CG331	-0.24
O_C	OG2D1	-0.58	HD1	HGA3	0.09
CHD	CGCPY3	-0.16	HD2	HGA3	0.09
HHD	HGA4	0.11	HD3	HGA3	0.09
C1D	CG2R51	0.22	CMA	CG331	-0.24
N_D	NG2R51	-0.39	HA1	HGA3	0.09
H_D	HGP1	0.35	HA2	HGA3	0.09
C4D	CG2R51	0.04	HA3	HGA3	0.09
C3D	CG2R51	0.04	CMB	CG331	-0.24
C2D	CG2R51	0.06	HB1	HGA3	0.09
CHA	CG2DC1	-0.19	HB2	HGA3	0.09
HHA	HGA4	0.19	HB3	HGA3	0.09
C1A	CG251O	0.17	CBD	CG321	-0.28
N_A	NG2R51	-0.40	HO3	HGA2	0.09
H_A	HGP1	0.35	HO4	HGA2	0.09
C2A	CGCPY4	0.09	CGD	CG2O2	0.62
C3A	CG2R51	0.10	O2D	OG2D1	-0.76
C4A	CG251O	0.22	O1D	OG311	-0.76
C4B	CG2R51	0.77	CBA	CG321	-0.28
N_B	NG2R51	-0.45	HO1	HGA2	0.09
H_B	HGP1	0.33	HO2	HGA2	0.09
C1B	CG2R51	0.24	CGA	CG2O2	0.62
C2B	CGCPY5	-0.06	O1A	OG311	-0.76
C3B	CGCPY6	-0.04	O2A	OG2D1	-0.76
O_B	OG312	-0.59	CAA	CG321	-0.18
CHB	CGCPY2	-0.15	HO5	HGA2	0.09
HHB	HGA4	0.14	HO6	HGA2	0.09
CAB	CG2DC1	-0.20	CAD	CG321	-0.18
HAB	HGA4	0.15	HO7	HGA2	0.09
CBB	CG2DC3	-0.50	HO8	HGA2	0.09
HV1	HGA5	0.22			

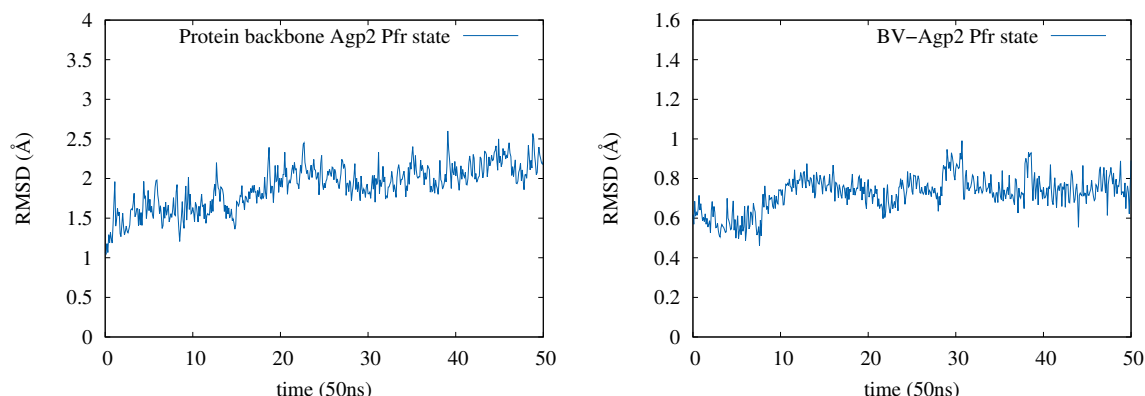


Figure 4.4: Left: root-mean-square deviation (RMSD) of backbone atoms for Agp2 phytochrome structure compared to the crystal structure (Pfr state) after 50ns. Right: RMSD of heavy atoms for the BV molecule compared to the crystal structure after.

very small rmsd values (0.04-0.28 Å) obtained for all MM-optimized structures using the new CGenFF parameters.

Table 4.7 shows the partial atomic charges obtained for the BV molecule. In order to evaluate the quality of the optimized partial charges of the BV molecule, the minimum interaction energies and water-chromophore distances were calculated at the ab initio level of theory (Tables 4.1-4.4). Overall, the QM-MM differences ($\Delta E_{int} \sim 2.0$ and $\Delta R_{min} \sim 0.2$) are in good agreement with the partial charges derived by Kaminski et al[65].

Pfr and Meta-F monomers. Figures 4.4 and 4.5 show the rmsd values calculated for all protein backbone atoms (without hydrogen atoms) during the MD simulation time using the crystal structures as the reference point. The rmsd values obtained for the Agp2 structural models (Pfr and Meta-F state) remains stable after 50 ns of MD simulation. Regarding the BV chromophore, Figures 4.4 and 4.5 show the rmsd calculated for all heavy atoms of BV molecule during the MD simulation using the crystal structures as the reference point. The rmsd of BV remains stable after 50 ns of MD simulations for both structural models (Pfr and Meta-F). The averaged rmsd calculated over the 50 ns of the production for all structural models led to ~ 1.9 Å for the protein backbone and ~ 0.7 Å for the BV molecule. These values are in good agreement with previous MD studies of phytochrome structures[65, 113]. However, the new CGenFF parameters led to a lower rmsd value for the BV molecule compared to the previous works.

Dynamic properties of the chromophore binding pocket. The analysis of the 50 ns production run of the MD trajectories (Agp2 Pfr and Meta-F state) led to the identification of several conformational changes in the chromophore binding pocket as is reflected in Figure 4.6. In the particular case of Agp2 in the Pfr state, the effect of the temperature is reflected in the breaking and formation of new hydrogen bonds between the BV molecule and protein. Since, unlike the crystal structure, the propionic side chain of ring C (PscC) forms a new hydrogen bond with the Ser262 residue (1.98 Å). Furthermore, His278 is reoriented in order to make a new hydrogen bond with Tyr165 (2.8 Å). Interestingly, the side chain of Arg242 undergoes a significant conformational change compared to the crystal structure. This suggests that the high flexibility of Arg242 can contribute to the formation of the salt

4. Optimized CGenFF force-field parameters for Biliverdin Chromophore in the Pfr and Meta-F states of Phytochromes

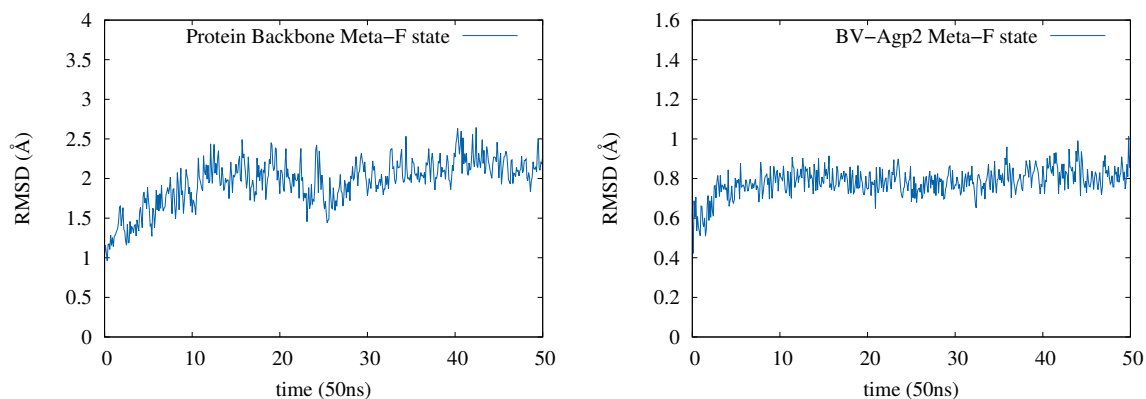


Figure 4.5: Left: root-mean-square deviation (RMSD) of backbone atoms for Agp2 phytochrome structure compared to the crystal structure (Meta-F state) after 50 ns. Right: RMSD of heavy atoms for the BV molecule compared to the crystal structure after 50 ns.

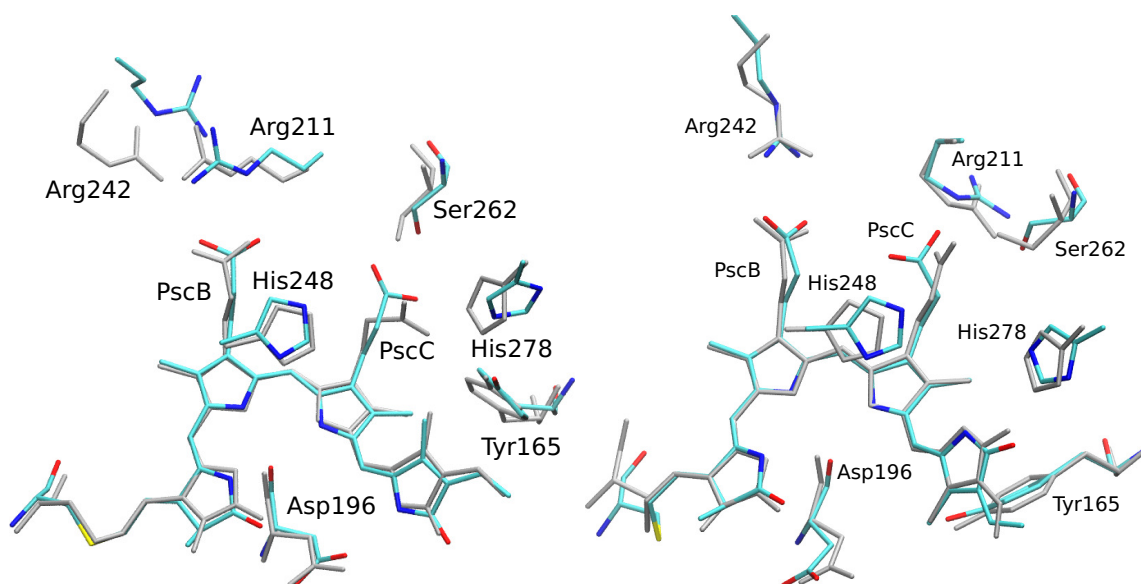


Figure 4.6: Left: Chromophore binding pocket (Agp2 in Pfr state). Crystal structure (color gray) superimposed with average structures after 50 ns MD simulation. Right: Chromophore binding pocket (Agp2 in Meta-F state). Crystal structure (color gray) superimposed with average structures after 50 ns MD simulation.

bridge with PscB in the Meta-F state. In contrast, Arg211, Asp196, and His248 remain stable after 50 ns of the MD simulation. This means that the Arg211, Asp196, and His248 residues contribute mainly to the stabilization of the chromophore binding pocket. As can be seen in Figure 4.6 the new CGenFF parameters maintain the structural identity of the BV molecule in the Pfr state after 50 ns of the MD simulation compared to the crystal structure.

Regarding the Agp2 phytochrome in the Meta-F state, Figure 4.6 shows small conformational changes in the chromophore binding pocket after 50 ns of the MD simulation. The salt bridge formed between the side chain of Arg242 and PscB (2.75 Å) remains stable during the 50 ns of the MD simulation. Unlike the Agp2 in the Pfr state, Arg242 is less flexible and contributes to the stabilization of the chromophore binding pocket. Furthermore, the isomerization process (rotation of the ring D) favors the formation of a hydrogen bond between His278 and carbonyl group (C=O) of ring D (2.03 Å). The Asp196 residue contributes to the

stabilization of the BV molecule by forming three hydrogen bonds with pyrrole rings A, B, and C in both Pfr and Meta-F states. With respect to the Arg211 residue, two new hydrogen bonds were formed with the Ser262 (2.07 Å) and PscC (2.24 Å), respectively. The formation of the hydrogen bond between Arg211 and PscC can contribute to the deprotonation of the PscC in the Pr state, since the deprotonation of PscC can lead to the formation of a strong salt bridge between Arg211 and PscC. In this case, the new CGenFF parameters also keep the structural identity of the BV molecule in the Meta-F state after 50 ns of the MD simulation compared to the crystal structure.

4.4 Conclusions

In this work, we derived accurate CGenFF parameters for the BV molecule. These parameters will enable the study of the dynamical properties of BV chromophore in different states of the phytochromes photocycle. The set of parameters obtained in this work successfully reproduce the potential energy surfaces of dihedral angles and intermolecular interactions of the BV chromophore with respect to the ab initio target data. Furthermore, the CGenFF parameters for the BV molecule were validated by performing MD simulations of the Agp2 phytochrome in both Pfr and Meta-F states. The MD results were in good agreement with respect to the experimental crystal structures based on the rmsd values calculated. The analysis of the MD trajectories of the Agp2 in both Pfr and Meta-F states led to the identification of different conformational changes in the chromophore binding pocket. New hydrogen bonds were formed between BV molecule with the Ser262 and Arg242, respectively. Furthermore, the conserved His278 makes a new hydrogen bond with Tyr165 in the chromophore binding pocket. Overall, the new CGenFF parameters keep the structural identity of the BV molecule in both Pfr and Meta-F states after 50 ns of the MD simulation compared to the crystal structures.

5

pKa Calculations in Bathy Phytochromes Structures

5.1 Introduction

The accurate evaluation of protonation states of residues located in the interior of a protein remains a challenging issue[114]. Typically, these amino acid residues show unusual pKa values in comparison with their values in aqueous solution. In the particular case of phytochromes, their photocycle involves proton transfer processes between the chromophore and key residues of the protein matrix [63, 115]. These proton transfer reactions are initiated by protonation changes of the chromophore or titratable residues of the protein. Thus, the precise determination of the pKa values of chromophore and titratable amino acids in the protein environment constitutes a fundamental step in the computational modeling of phytochromes.

In this work, we will focus on two approaches for computing pKa values of titratable residues in phytochromes. The first approach is based on the combination of electrostatic energy computations by solving the Poisson Boltzmann equation (PBE) with classical molecular dynamics (MD) simulations. The second one is the constant-pH molecular dynamics (CpHMD) method, which has been used for studying several important biological processes caused by changes in solution pH. One of the main advantages of the CpHMD approaches is that the protonation states may change during the conformational dynamics. Therefore, the dynamical processes coupled to a change in protonation states can be directly studied. These two methods allow taking into account protein flexibility, hydrogen-bond network rearrangements, side-chain reorientations and water molecules inside protein cavities.

These methodologies will be applied to the Agp2 phytochrome structure in order to understand the proton transfer process between the BV chromophore and protein. Interestingly, experimental spectroscopic data have shown that the propionic side chain at ring C (pscC) of the BV molecule of the Agp2 phytochrome is protonated in the Pfr and Meta-F states and remains protonated even up to pH of 11.60. In contrast, the pscC is deprotonated in the Pr state, indicating that the proton release takes place at the transition from Meta-F to Pr

state when the photoreceptor becomes activated. Recently, it has been demonstrated that the pscC deprotonation of BV chromophore is fundamental for triggering the secondary structure transition[11].

5.2 Methods

Atomic Partial Charges of Tetrapyrrole Chromophore. Initial coordinates of the tetrapyrrole chromophore of Agp2 phytochrome (Pfr state) were taken from PDB entry: 6G1Y. In the first step, we carried out geometry optimizations of the chromophore geometry divided into smaller fragments with the quantum chemical program Jaguar[116] using the B3LYP DFT functional and the 6-31G(d) basis set. Using the optimized chromophore fragments, we computed the electrostatic potential using the same DFT functional and basis set. Atomic partial charges were determined by means of the two-stage restraint-electrostatic-potential (RESP) method. These partial charges were used for computing pKa values of propionic side chain C with the Karlsberg2+ program.

MD simulations with NAMD. The Agp2 phytochrome structures (Pfr and Meta-F state) were modeled with the plug-in of VMD and solvated in a TIP3P water box applying periodic boundary conditions. We used the CHARMM36 force field for protein, and the CHARMM General Force Field (CGenFF optimized from chapter 4) for BV molecule. The MD simulations were performed with the software NAMD using a 2 fs time step and Langevin dynamics with friction constant $\beta = 1 \text{ ps}^{-1}$ at 300 K temperature. We performed three MD simulations for the Agp2 phytochrome structures in the Pfr and Meta-F states, with the propionic side chain (attached to ring C) deprotonated and protonated (see Figure 5.1).

Dowser waters The Dowser program was used to place internal water molecules inside hydrophobic cavities of the Agp2 phytochrome (Pfr) using proteinwater and waterwater interaction energy algorithms[117]. The DOWSER parameters for the BV chromophore molecule were taken from the Ref [118].

pKa computations with Karlsberg2+. The pKa values of titratable residues were carried out with a recent version of the Karlsberg2+ program[43]. The APBS software was used for the computation of the electrostatic energies by solving the Poisson Boltzmann equation (PBE). Electrostatic energy computations were performed for time frames of MD trajectories.

pKa calculations based on constant pH MD simulations. The CpHMD method was applied to the Agp2 phytochrome in both Pfr and Meta-F states. These calculations were performed by assigning the propionic side chain C of the BV, His248, His278, Tyr and Cys as titratable sites, as described by Cruzeiro et al[119]. All the Agp2 models were solvated in a water box and counter ions added to neutralize the total charge with the Amber tool leap program. All the MD simulations were performed with a 2 fs time step under periodic boundary conditions with the particle-mesh-Ewald method for electrostatic interactions, a

cutoff of 12 Å for the van der Waals interaction and hydrogens constrained with the SHAKE algorithm. After the equilibration 200 ns constant pH conventional MD was performed with the AMBER software. The pKa calculations were performed at pH=7.0 using the Monte Carlo protonation state change attempts every 100 fs. These calculations were done in collaboration with Giovanni Battocchio and also belong to his PhD thesis.

5.3 Results and Discussion

5.3.1 pKa calculations based on classical MD simulations and electrostatic energy calculations.

Pfr and Meta-F monomers. Figures 5.2 and 5.4 show the rmsd values calculated for all protein backbone atoms (no hydrogen atoms) over the simulation time using the crystal structures as the reference point. As can be seen in Figures 5.2 and 5.4, the rmsd of the Agp2 structural models (Pfr and Meta-F) remains stable after 50 ns of MD simulation, nevertheless, during the first ~ 40 ns the rmsd is increased progressively, therefore, the last 10 ns will be considered for the pKa analysis of propionic side chain. With respect to the BV chromophore, Figures 5.3 and 5.5 show the rmsd calculated for all heavy atoms of BV molecule during the MD simulation using the crystal structures as the reference point. The rmsd of BV molecule remains stable after 50 ns of MD simulations for all structural models (Pfr and Meta-F). The averaged rmsd for all structural models evaluated over the 50 ns of the production lead to ~ 1.9 Å for the protein backbone and ~ 0.7 Å for the BV molecule. These values are in good agreement with previous MD studies of phytochrome structures[65, 113].

Figure 5.6 shows the time evolution (50ns) of computed pKa values of pscC for Agp2 in the Pfr state. Three MD trajectories with different pscC protonation states were used for evaluation with the Karlsberg2+ program. In general, the pKa values of pscC for the two protonated MD trajectories are located in a range from -10 to 6.0, whilst the pKa values of pscC for the deprotonated MD trajectory are found in a range from -10 to -2.0. Regarding the Agp2 in Meta-F state, Figure 5.7 shows the time evolution (50ns) of computed pKa values of pscC located in a range from -10 to 5.0, however, unlike the Agp2 in the Pfr state, the pKa values of pscC in the Meta-F state show similar behavior in the three MD trajectories. Table 5.1 shows the pKa values obtained for propionic side chain pscC and for some titratable residues located in the chromophore binding pocket. According to the pKa values computed from MD trajectories, the pscC is found in their deprotonated form ($\text{pKa} < 7.0$), furthermore, the pKa values obtained for both His248 and His248 are lower than 7.0, this indicates that both His248 and His248 are found in their deprotonated form. With respect to the Tyr165 residue, the pKa value obtained is greater than 7.0, suggesting that the Tyr165 residue is found in its protonated form.

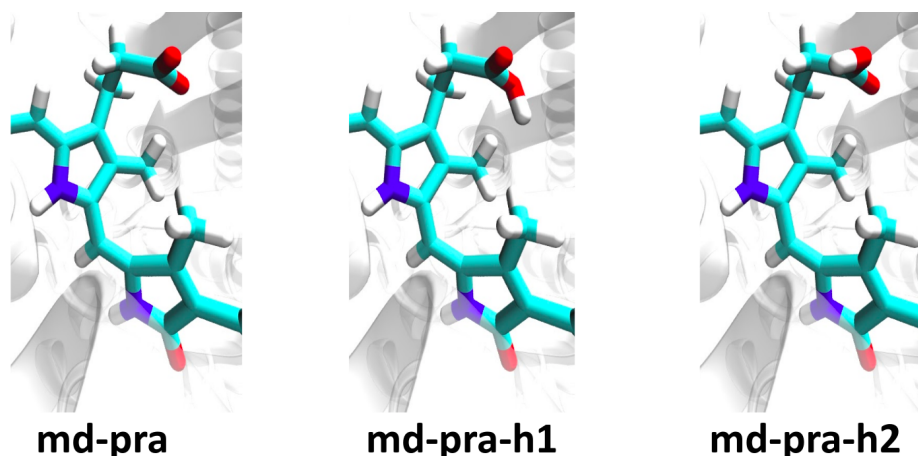


Figure 5.1: Three MD trajectories of the Agp2 phytochrome with propionic side chain (attached to ring C) deprotonated and protonated.

The pKa values calculated by Karlsberg2+ in combination with classical MD simulations led to a propionic side chain of ring C deprotonated in both Pfr and Meta-F states. However, these results are in complete contradiction with the experimental Raman spectra of Agp2 phytochrome. Thereby, in order to investigate the reasons for these contradictory results, we decided to compare the pKa values obtained by Karlsberg2+ with those obtained by the constant pH method. Figure 5.6 shows an interesting pKa evolution during the last 50 ns of the MD trajectory in the Pfr state (pscC in its protonated form). These results are closer to those calculated by the constant pH method, this suggests that the Karlsberg2+ program can reproduce pKa values with similar accuracy to the constant pH approach. Nevertheless, as can be seen in Figure 5.6, the MD trajectory corresponding to the deprotonated form of the pscC (md-pra) reproduces very low pKa values during the last 50 ns of the MD simulation. These values are the result of poor sampling of both conformational and protonation states. Therefore, the idea of combining different classical MD trajectories using specific protonation patterns shows, in the particular case of phytochromes, huge limitations. These MD trajectories are performed independently from each other, which in turn produces an “biased” sampling.

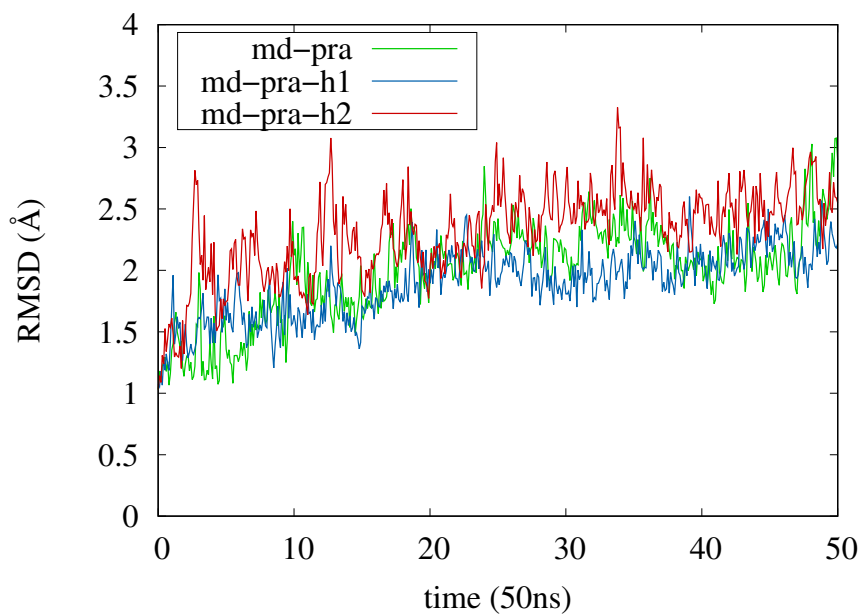


Figure 5.2: Root-mean-square deviation (RMSD) of backbone atoms compared to the Agp2 phytochrome crystal structure (Pfr state) after 50ns.

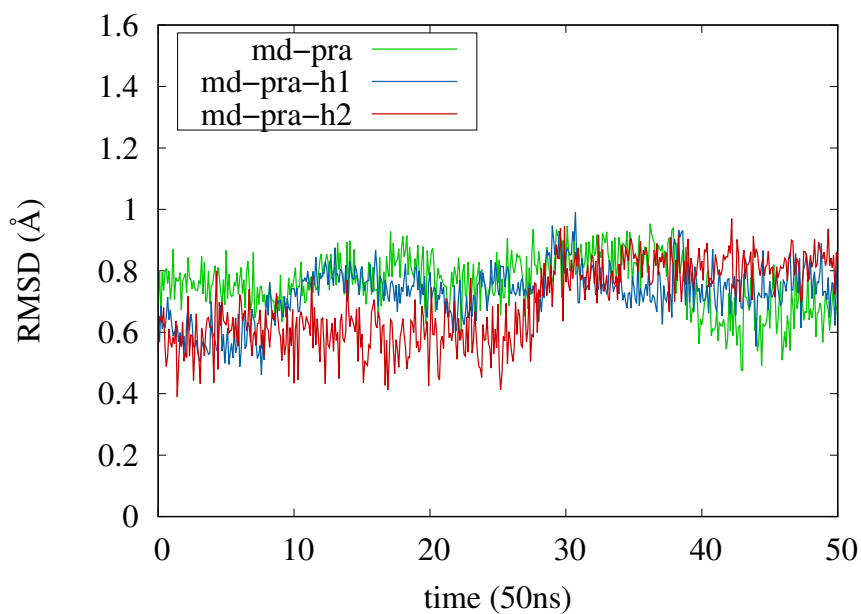


Figure 5.3: Root-mean-square deviation (RMSD) of heavy atoms of the BV molecule of Agp2 in Pfr state compared to the crystal structure after 50ns.

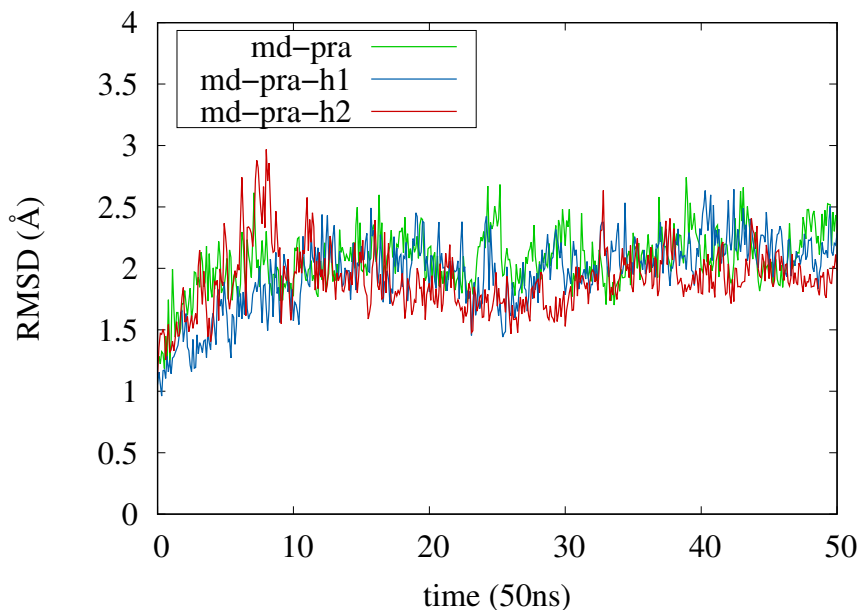


Figure 5.4: Root-mean-square deviation (RMSD) of backbone atoms of Agp2 phytochrome compared to the crystal structure (Meta-F state) after 50ns.

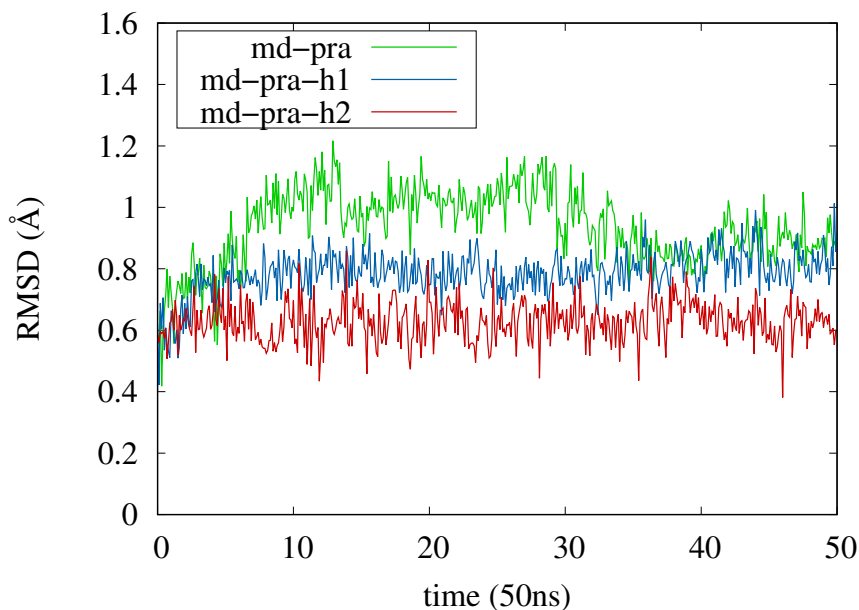


Figure 5.5: Root-mean-square deviation (RMSD) of heavy atoms of the BV molecule of Agp2 (Meta-F state) compared to the crystal structure after 50ns.

Regarding the Agp2 in Meta-F state, the accurate determination of the pKa value of pscC is even more challenging, since, unlike the Agp2 in the Pfr state, the chemical environment of pscC is more polarized. The transition from Pfr to Meta-F state triggers several conformational changes in the vicinity of the pscC. The most significant change corresponds to the isomerization process of the BV molecule, this rotation increases the polar environment of pscC by moving away the methyl group attached to the ring D (Pfr state, see Figure 5.8), whilst two polar groups (amino NH and carbonyl C=O) are closer to the pscC in the Meta-F state. Interestingly, the salt bridge formed between Arg211 and pscB (Pfr state) is open in order to make a new

hydrogen bond with the Ser262 (Meta-F state), this breaking allows the formation of a new salt bridge between Arg242 and pscB. As can be seen in Figure 5.8, the Arg211 is in direct contact with the pscC by forming a hydrogen bond (Meta-F state), thus, the reorientation of both Arg211 and Arg242 increases significantly the polar environment of pscC.

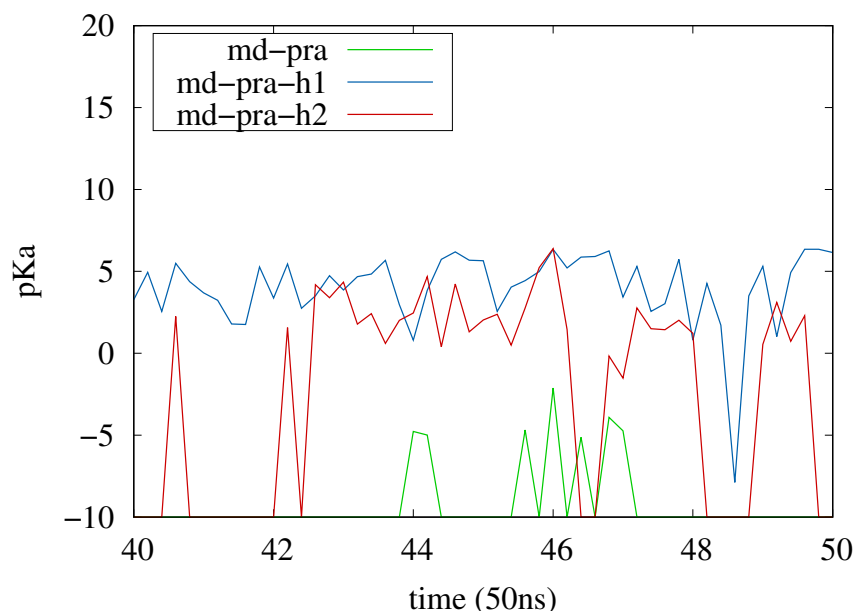


Figure 5.6: Time evolution (50ns) of computed pKa values of pscC for Agp2 in Pfr state. Three MD trajectories with different pscC protonation were used for evaluation with Karlsberg2+ program.

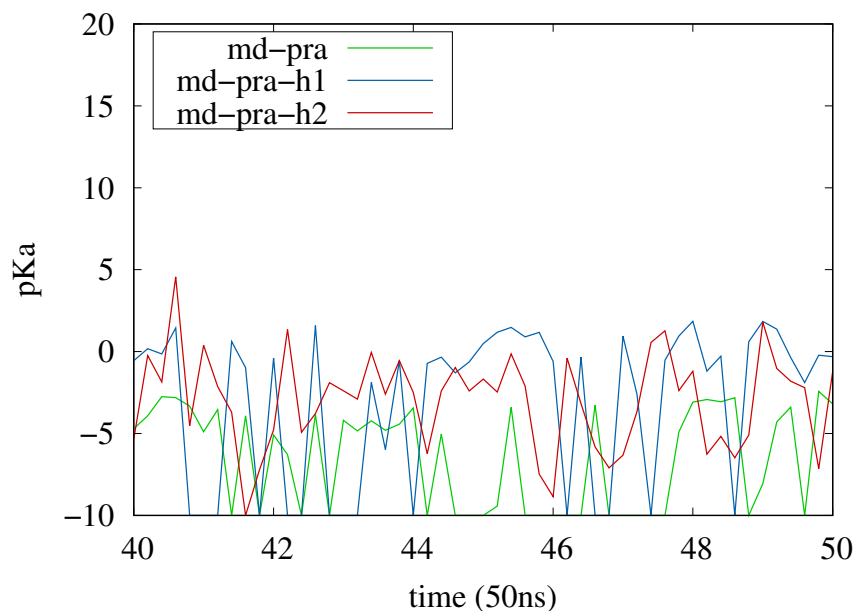


Figure 5.7: Time evolution (the last 10ns) of computed pKa values of pscC for Agp2 in Meta-F state. Three MD trajectories with different pscC protonation were used for evaluation with Karlsberg2+ program.

Table 5.1: pKa values calculated for pscC and some residues of the chromophore binding pocket with Karlsberg2+ in combination with classical MD simulations for the Agp2 in Pfr and Meta-F state.

Residue	Agp2 Pfr	Agp2 Meta-F
pscC	1.3	-10.0
His248	-10.0	-7.7
His278	-9.3	-7.0
Tyr165	20.0	19.6

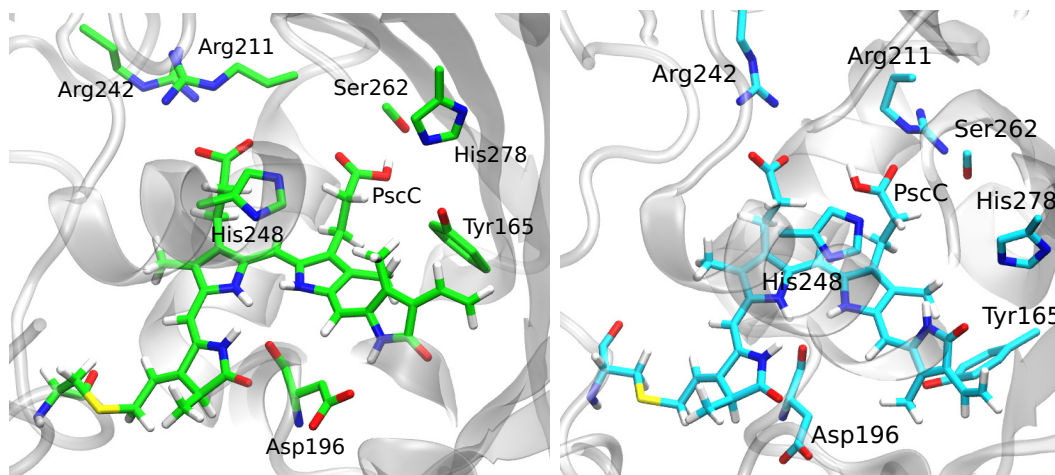


Figure 5.8: Left: Chromophore binding pocket (Agp2 in Pfr state) after 50 ns MD simulation. Right: Chromophore binding pocket (Agp2 in Meta-F state) after 50 ns MD simulation.

5.3.2 Dowser waters and Karlsberg2+

In order to investigate the role of water molecules inside the chromophore binding pocket and their effect on the pKa value of pscC, additional water molecules were located in the chromophore binding pocket of the Agp2 phytochrome in the Pfr state by using the Dowser program. 50 water molecules were located by Dowser in the whole protein, however, ~6 water molecules were located in the chromophore binding pocket. Usually, water molecules located in the interior of a protein are difficult to observe, partly due to the dynamic nature of internal water, which makes the crystallographic prediction of the water molecules, for example, placed in the chromophore binding pocket of phytochromes structures very challenging.

Figure 5.9 shows the pKa evolution during the last 20 ns of the MD simulation for two MD trajectories of Agp2 (Pfr), both trajectories contain the pscC protonated, however, one trajectory includes crystal water and internal water molecules located by Dowser program (md-cwd), in contrast, the second trajectory contains only crystal water molecules (md-cw). Interestingly, the pKa values calculated for the MD trajectory including the Dowser waters, shows fewer fluctuations during the last 20 ns of the MD simulation, on the contrary, the lack of the Dowser waters in the second MD trajectory increases the fluctuations on the pKa values, this suggests that internal water molecules placed in the chromophore binding pocket contribute to the stability of propionic side-chain C by forming several H-bonds, as well as reducing the flexibility of pscC.

Figure 5.10 shows the hydrogen bond network inside the chromophore binding pocket for both MD trajectories. According to the crystal structure (Pfr), the pscC forms a hydrogen bond with His278 (3.3 Å) and Tyr265 (2.9 Å), respectively. However, this conformation of the chromophore binding pocket leads to a very low pKa value of the pscC (-10.0). Interestingly, the inclusion of the Dowser waters leads to the formation and breaking of several hydrogen bonds in the chromophore binding pocket. As can be seen in Figure 5.10 (Right), the pscC forms a hydrogen bond with two water molecules, respectively. This new conformation leads to a higher pKa value of the pscC (6.1) suggesting that during the photocycle the proton release event of pscC might be assisted by water molecules located inside the chromophore binding pocket. Nevertheless, it is important to mention that the combination of the deprotonated and protonated MD trajectories (with Dowser waters) led to a pKa value lower than 7.0, which means that, in the particular case of phytochromes, classical MD trajectories produce “biased” results.

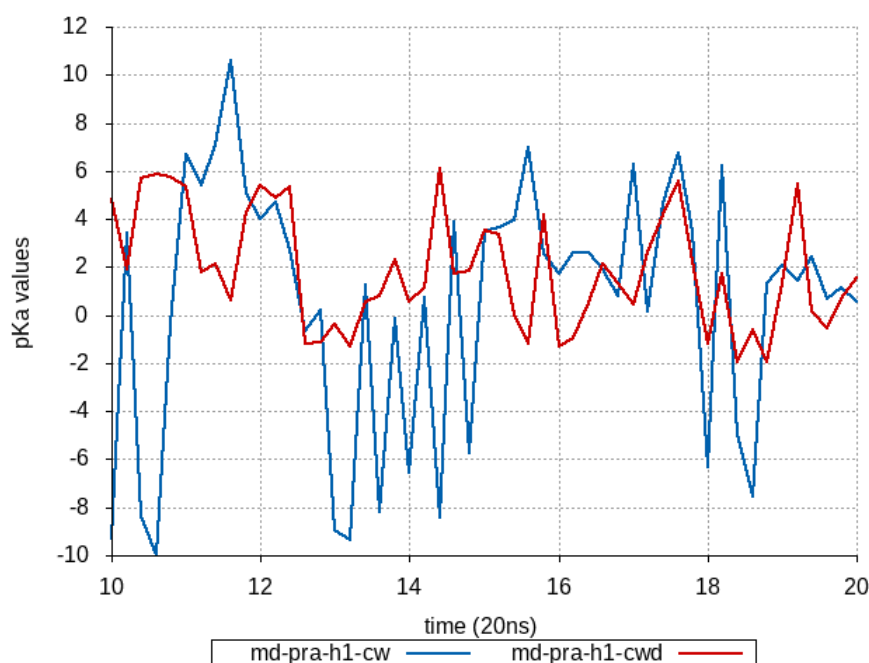


Figure 5.9: Time evolution (the last 20ns) of computed pKa values of pscC for Agp2 in Pfr state using classical MD simulations and Karlsberg2+ program. The “md-pra-h1-cw” trajectory includes crystal waters and the “md-pra-h1-cwd” trajectory includes crystal and Dowser waters.

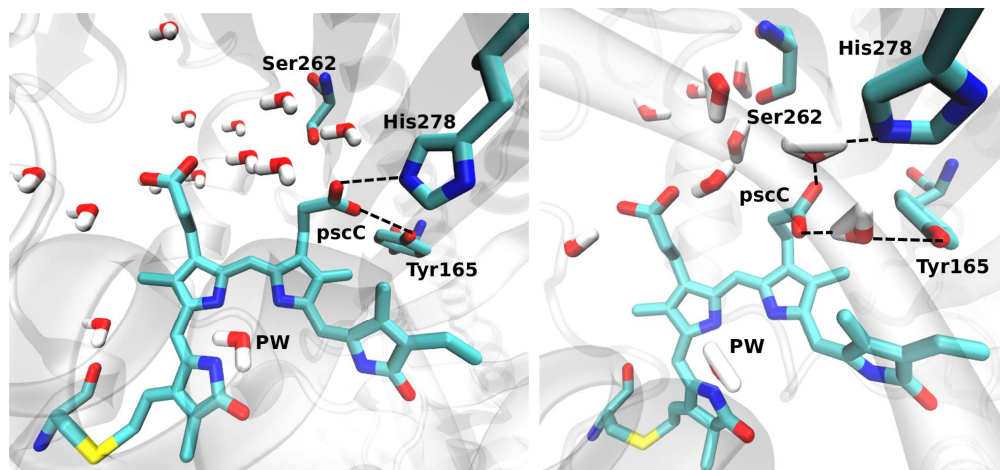


Figure 5.10: Left: Chromophore binding pocket and crystal water molecules of Agp2 in the Pfr state, for this MD frame the pscC has a $pK_a = -10$, Right: Chromophore binding pocket including crystal water and Dowser water molecules of Agp2 in the Pfr state, for this MD frame the pscC has a $pK_a = 6.1$.

5.3.3 pKa values using the constant-pH method

Since the pKa values obtained with the combination of electrostatics calculations and classical MD simulations led to results that are in contradiction with the experimental Raman spectra of Agp2 phytochrome[87, 11, 77], we decided to apply the CpHMD method to the Agp2 phytochrome in the Pfr and Meta-F states.

Table 5.2 shows the pKa values of pscC, His248, and His278 for Agp2 in both Pfr and Meta-F states. According to CpHMD results, the pscC is found in its protonated form in both Pfr and Meta-F states, which is in good agreement with the experimental evidence [11, 87]. Furthermore, the pKa value of pscC is lower in the Meta-F state (7.0) compared to the pKa value calculated in the Pfr state (7.9), this suggests that the polarized chemical environment of pscC (Meta-F) contributes significantly to the deprotonation of the pscC in the transition from Meta-F to Pr state. Moreover, CpHMD predictions led to a His248 deprotonated in both Pfr and Meta-F states ($pK_a < 7.0$), which is reflected by the high percentage of the population in the deprotonated form, this result suggests that the His248 is a very unlikely proton site acceptor for the deprotonation of the pscC. Interestingly, CpHMD predicts a His278 deprotonated in the Pfr state ($pK_a < 7.0$), whilst in the Meta-F state, His278 is found in its protonated form ($pK_a > 7.0$). The pKa shift of pscC in the transition from Pfr to Meta-F state reflects a strong correlation with the protonation state of the His278, this suggests that His278 is the most likely proton site acceptor in the chromophore binding pocket.

Due to the close proximity of His248 and His278 to the BV chromophore (see Figure 5.8) a greater emphasis has been given to the elucidation of its protonation state[12, 62, 63, 64], nevertheless, by using the CpHMD method both His248 and His278 can change their protonation states during the MD simulation, which leads to a more accurate sampling. Therefore, the CpHMD results not only reveal the role of His278 in the photocycle of Agp2, but also the importance of sampling the histidine residues that are in direct contact with the chromophore molecule. Interestingly, the conclusions derived for the Agp2 phytochrome can

be extended to all bathy phytochromes, since the residues located in the chromophore binding pocket are highly conserved.

Table 5.2: Mean pKa values and population distribution after 100 ns for the Agp2 in both Pfr and Meta-F states.

	Pfr pKa	Pfr Pop	Meta-F pKa	Meta-F Pop
Prop-C	7.9	D 30%, P 70%	7.0	D 50%, P 50%
His248	5.0	P 1%, D/E 99%	6.1	P 29%, D/E 71%
His278	6.1	P 14%, D/E 86%	7.3	P 67%, D/E 33%

5.4 Conclusions

In this work, we applied two different methods in order to study the protonation state of the propionic side chain of the ring C of Agp2 phytochrome in Pfr and Meta-F states. The first method combines the Karlsberg2+ program[43] with classical MD simulations, the results obtained with this method led to a propionic side chain of ring C deprotonated in both Pfr and Meta-F states. However, these results are in contradiction with the experimental Raman spectra of Agp2 phytochrome [87, 11, 77]. The second approach used in this work was the CpHMD method, the pKa values obtained with the CpHMD method are in good agreement with the experimental results[11]. Furthermore, the CpHMD method led to the identification of the His278 as a potential proton site acceptor, since the protonation state of pscC and His278 are highly correlated. In summary, by using the CpHMD approach, the protonation states of titratable sites may change in the course of conformational dynamics. Thus, the dynamical processes coupled to a change in protonation states of phytochromes can be directly studied. In contrast, by using classical MD simulations, the protonation states are fixed during the conformational dynamic, which leads to an unrealistic sampling.

6

Final Conclusions and Outlook

- **The fragment molecular orbital method (FMO)**

The present study has demonstrated the first successful application and potential of the fragment molecular orbital (FMO) method to investigate the geometries and the chromophore-protein interactions in phytochromes by using a fully quantum chemical treatment. The application of the FMO method in combination with the DFTB3 approach[83] led to the optimization of the geometries of three structural models of the *Deinococcus radiodurans* DrBphP phytochrome, each model contains ~ 6159 atoms. The optimized structural models show only minor deviations from the experimental structure.

The application of the pair interaction energy decomposition analysis (PIEDA) to bacterial phytochrome DrBphP led to the identification of the chemical nature and quantification of chromophore-protein interactions in DrBphP, Agp1 and Agp2 phytochromes. Moreover, new interactions were identified in the chromophore-binding pocket, two nonclassical H-bonds (CH/O interactions) and an OH/ π interaction, which might have photochemical relevance.

Furthermore, the FMO method was recently applied to the analysis of ligand binding in the SARS-CoV-2 M^{pro}-N3 complex system by using 1000 dynamically fluctuating structures[120]. Interestingly, the relative importance of each residue is modified by means of structural fluctuations. These results suggest that the ligand is stabilized in a dynamic way through collective interactions formed by multiple residues. Therefore, the combination of the FMO method with classical MD or CpHMD simulations might provide a new understanding of chromophore-protein interactions in phytochromes.

- **Optimized CGenFF force-field parameters for Biliverdin Chromophore**

Since the results of MD simulations based on molecular mechanics largely depend on the quality of force fields parameters, in this study, we derived accurate non-polarizable CGenFF parameters for the BV molecule. These parameters will enable the study of the dynamical properties of BV chromophore in different states of the phytochromes photocycle. Furthermore, the new CGenFF parameters of BV molecule can be used in conjunction with the highly optimized CHARMM all-atom force field parameters for proteins. Moreover, the potential energy surfaces of dihedral angles calculated can be used as starting point for the derivation of polarizable force field parameters for BV molecule[37], which can lead to more accurate results.

- **pKa Calculations in Bathy Phytochromes Structures**

In this work, we demonstrated that the study of the protonation state changes of the chromophore and key residues of the protein matrix in bathy phytochromes requires highly accurate methods. The calculation of pKa values in proteins has commonly relied on the combination of electrostatic energies with classical MD simulations. However, in this study, we found that classical MD simulations can lead to an “biased” sampling since the protonation states remain fixed during the conformational dynamic. To address this issue, the CpHMD method was applied to the Agp2 phytochrome in the Pfr and Meta-F states. One of the main advantages of the CpHMD approaches is that the protonation states may change during the conformational dynamics.

The pKa values obtained with the CpHMD method not only are in good agreement with the experimental results[11] but also led to the identification of the His278 as a potential proton site acceptor. Interestingly, this result can be extended to all bathy phytochromes since the residues located in the chromophore binding pocket are highly conserved. Furthermore, this work has shown the importance of carrying out a better sampling of the histidine residues that are in direct contact with the chromophore molecule. In summary, this thesis shows a new and different approach to the investigation of phytochromes. We hope that the results of this thesis contribute to the future studies in the phytochromes field.

References

- [1] Winslow R Briggs and John L Spudich. *Handbook of photosensory receptors*. John Wiley & Sons, 2005.
- [2] Nathan C Rockwell, Yi-Shin Su, and J Clark Lagarias. “Phytochrome structure and signaling mechanisms”. In: *Annu. Rev. Plant Biol.* 57 (2006), pp. 837–858.
- [3] Nathan C Rockwell and J Clark Lagarias. “A brief history of phytochromes”. In: *ChemPhysChem* 11.6 (2010), pp. 1172–1180.
- [4] David von Stetten et al. “Chromophore heterogeneity and photoconversion in phytochrome crystals and solution studied by resonance Raman spectroscopy”. In: *Angew. Chem. Int. Ed.* 47.25 (2008), pp. 4753–4755.
- [5] Jeremiah R Wagner et al. “A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome”. In: *Nature* 438.7066 (2005), p. 325.
- [6] Jeremiah R Wagner et al. “High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution”. In: *J. Biol. Chem.* 282.16 (2007), pp. 12298–12309.
- [7] Seong-Hee Bhoo et al. “Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore”. In: *Nature* 414.6865 (2001), pp. 776–779.
- [8] Baruch Karniol and Richard D Vierstra. “The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties”. In: *Proceedings of the National Academy of Sciences* 100.5 (2003), pp. 2807–2812.
- [9] Andrea Schmidt et al. “Structural snapshot of a bacterial phytochrome in its functional intermediate state”. In: *Nature communications* 9.1 (2018), pp. 1–13.
- [10] Peter H Quail. “Phytochrome photosensory signalling networks”. In: *Nature reviews Molecular cell biology* 3.2 (2002), pp. 85–93.
- [11] Anastasia Kraskov et al. “Intramolecular Proton Transfer Controls Protein Structural Changes in Phytochrome”. In: *Biochemistry* 59.9 (2020), pp. 1023–1037.
- [12] Francisco Velazquez Escobar et al. “Protonation-dependent structural heterogeneity in the chromophore binding site of cyanobacterial phytochrome Cph1”. In: *J. Phys. Chem. B* 121.1 (2016), pp. 47–57.
- [13] Francisco Velázquez Escobar et al. “Common structural elements in the chromophore binding pocket of the Pfr state of bathy phytochromes”. In: *Photochemistry and photobiology* 93.3 (2017), pp. 724–732.

REFERENCES

- [14] Alan D McNaught, Andrew Wilkinson, et al. *Compendium of chemical terminology*. Vol. 1669. Blackwell Science Oxford, 1997.
- [15] Berthold Borucki et al. “Light-induced proton release of phytochrome is coupled to the transient deprotonation of the tetrapyrrole chromophore”. In: *Journal of Biological Chemistry* 280.40 (2005), pp. 34358–34364.
- [16] Harald Foerstendorf et al. “FTIR studies of phytochrome photoreactions reveal the CO bands of the chromophore: consequences for its protonation states, conformation, and protein interaction”. In: *Biochemistry* 40.49 (2001), pp. 14952–14959.
- [17] SE Braslavsky, W Gärtner, and K Schaffner. “Phytochrome photoconversion”. In: *Plant, Cell & Environment* 20.6 (1997), pp. 700–706.
- [18] Qian-Zhao Xu et al. “Tongue Refolding in the Knotless Cyanobacterial Phytochrome All2699”. In: *Biochemistry* (2020).
- [19] Lars-Oliver Essen, Jo Mailliet, and Jon Hughes. “The structure of a complete phytochrome sensory module in the Pr ground state”. In: *Proceedings of the National Academy of Sciences* 105.38 (2008), pp. 14709–14714.
- [20] Max Born and Robert Oppenheimer. “Zur quantentheorie der molekeln”. In: *Annalen der physik* 389.20 (1927), pp. 457–484.
- [21] John C Slater. “A simplification of the Hartree-Fock method”. In: *Physical review* 81.3 (1951), p. 385.
- [22] Marcus Elstner et al. “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”. In: *Physical Review B* 58.11 (1998), p. 7260.
- [23] Kazuo Kitaura et al. “Fragment molecular orbital method: an approximate computational method for large molecules”. In: *Chem. Phys. Lett.* 313.3-4 (1999), pp. 701–706.
- [24] Yoshio Nishimoto, Dmitri G Fedorov, and Stephan Irle. “Density-functional tight-binding combined with the fragment molecular orbital method”. In: *Journal of chemical theory and computation* 10.11 (2014), pp. 4801–4812.
- [25] D.G. Fedorov. “The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications”. In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* e1322 (2017).
- [26] Iris Antes and Walter Thiel. “On the treatment of link atoms in hybrid methods”. In: ACS Publications, 1998.
- [27] Yoshinobu Akinaga et al. “Fragmentation at sp² carbon atoms in fragment molecular orbital method”. In: *Journal of Computational Chemistry* 41.15 (2020), pp. 1416–1420.
- [28] Dmitri G Fedorov and Kazuo Kitaura. “Energy decomposition analysis in solution based on the fragment molecular orbital method”. In: *J. Phys. Chem. A* 116.1 (2011), pp. 704–719.
- [29] Alexander Heifetz et al. “Fragment molecular orbital method applied to lead optimization of novel interleukin-2 inducible T-cell kinase (ITK) inhibitors”. In: *J. Med. Chem.* 59.9 (2016), pp. 4352–4363.

-
- [30] Dmitri G Fedorov and Kazuo Kitaura. "Second order Møller-Plesset perturbation theory based upon the fragment molecular orbital method". In: *J. Chem. Phys* 121.6 (2004), pp. 2483–2490.
- [31] Peifeng Su and Hui Li. "Continuous and smooth potential energy surface for conductorlike screening solvation model using fixed points with variable areas". In: *J. Chem. Phys* 130.7 (2009), p. 074109.
- [32] Hiroya Nakata et al. "Simulations of chemical reactions with the frozen domain formulation of the fragment molecular orbital method". In: *Journal of chemical theory and computation* 11.7 (2015), pp. 3053–3064.
- [33] Hiroya Nakata et al. "Simulations of Raman spectra using the fragment molecular orbital method". In: *Journal of Chemical Theory and Computation* 10.9 (2014), pp. 3689–3698.
- [34] Andrew Komornicki and James W McIver Jr. "An efficient abinitio method for computing infrared and Raman intensities: Application to ethylene". In: *The Journal of Chemical Physics* 70.4 (1979), pp. 2014–2016.
- [35] George B Bacskay, Svein Saebø, and Peter R Taylor. "On the calculation of dipole moment and polarizability derivatives by the analytical energy gradient method: Application to the formaldehyde molecule". In: *Chemical physics* 90.3-4 (1984), pp. 215–224.
- [36] Gongpu Zhao et al. "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics". In: *Nature* 497.7451 (2013), pp. 643–646.
- [37] Pedro EM Lopes et al. "Polarizable force field for peptides and proteins based on the classical drude oscillator". In: *Journal of chemical theory and computation* 9.12 (2013), pp. 5430–5449.
- [38] Yue Shi et al. "Polarizable atomic multipole-based AMOEBA force field for proteins". In: *Journal of chemical theory and computation* 9.9 (2013), pp. 4046–4063.
- [39] Adri CT Van Duin et al. "ReaxFF: a reactive force field for hydrocarbons". In: *The Journal of Physical Chemistry A* 105.41 (2001), pp. 9396–9409.
- [40] Robert B Best et al. "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles". In: *Journal of chemical theory and computation* 8.9 (2012), pp. 3257–3273.
- [41] Kenno Vanommeslaeghe et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". In: *Journal of computational chemistry* 31.4 (2010), pp. 671–690.
- [42] G Matthias Ullmann and Ernst-Walter Knapp. "Electrostatic models for computing protonation and redox equilibria in proteins". In: *Eur. Biophys. J.* 28.7 (1999), pp. 533–551.
- [43] Gernot Kieseritzky and Ernst-Walter Knapp. "Optimizing pKa computation in proteins with pH adapted conformations". In: *Proteins: Structure, Function, and Bioinformatics* 71.3 (2008), pp. 1335–1348.

REFERENCES

- [44] Bjoern Rabenstein and Ernst-Walter Knapp. “Calculated pH-dependent population and protonation of carbon-monooxy-myoglobin conformers”. In: *Biophysical Journal* 80.3 (2001), pp. 1141–1150.
- [45] Gernot Kieseritzky and Ernst-Walter Knapp. “Optimizing pKa computation in proteins with pH adapted conformations”. In: *Proteins: Structure, Function, and Bioinformatics* 71.3 (2008), pp. 1335–1348.
- [46] Tim Meyer and Ernst-Walter Knapp. “pKa Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories”. In: *Journal of chemical theory and computation* 11.6 (2015), pp. 2827–2840.
- [47] Thomas Simonson and Charles L Brooks. “Charge screening and the dielectric constant of proteins: insights from molecular dynamics”. In: *Journal of the American Chemical Society* 118.35 (1996), pp. 8452–8458.
- [48] Alexey V Aleksandrov, Benoit Roux, and Alexander D MacKerell. “pKa Calculations with the Polarizable Drude Force Field and Poisson-Boltzmann Solvation Model”. In: *Journal of Chemical Theory and Computation* (2020).
- [49] Nathan A Baker et al. “Electrostatics of nanosystems: application to microtubules and the ribosome”. In: *Proceedings of the National Academy of Sciences* 98.18 (2001), pp. 10037–10041.
- [50] Jack A Henderson et al. “How ligand protonation state controls water in protein–ligand binding”. In: *The journal of physical chemistry letters* 9.18 (2018), pp. 5440–5444.
- [51] Christopher R Ellis and Jana Shen. “pH-dependent population shift regulates BACE1 activity and inhibition”. In: *Journal of the American Chemical Society* 137.30 (2015), pp. 9543–9546.
- [52] Jana Khandogin, Jianhan Chen, and Charles L Brooks. “Exploring atomistic details of pH-dependent peptide folding”. In: *Proceedings of the National Academy of Sciences* 103.49 (2006), pp. 18546–18550.
- [53] Robert C Harris and Jana Shen. “GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: pKa Predictions with Single-pH Simulations”. In: *Journal of chemical information and modeling* 59.11 (2019), pp. 4821–4832.
- [54] Antonio M Baptista, Vitor H Teixeira, and Cláudio M Soares. “Constant-p H molecular dynamics using stochastic titration”. In: *The Journal of chemical physics* 117.9 (2002), pp. 4184–4200.
- [55] David A Case et al. “The Amber biomolecular simulation programs”. In: *Journal of computational chemistry* 26.16 (2005), pp. 1668–1688.
- [56] John Mongan, David A Case, and J ANDREW McCAMMON. “Constant pH molecular dynamics in generalized Born implicit solvent”. In: *Journal of computational chemistry* 25.16 (2004), pp. 2038–2048.
- [57] E Sethe Burgie et al. “Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion”. In: *J. Biol. Chem.* 289.35 (2014), pp. 24573–24587.

-
- [58] Yoshio Nishimoto and Dmitri G Fedorov. “The fragment molecular orbital method combined with density-functional tight-binding and the polarizable continuum model”. In: *Phys. Chem. Chem. Phys.* 18.32 (2016), pp. 22047–22061.
- [59] Eric F Pettersen et al. “UCSF Chimera-a visualization system for exploratory research and analysis”. In: *J. Comput. Chem.* 25.13 (2004), pp. 1605–1612.
- [60] Gernot Kieseritzky and Ernst-Walter Knapp. “Optimizing pKa computation in proteins with pH adapted conformations”. In: *Proteins: Struct. Funct. Bioinf* 71.3 (2008), pp. 1335–1348.
- [61] Jing Huang et al. “CHARMM36m: an improved force field for folded and intrinsically disordered proteins”. In: *Nat. Methods* 14.1 (2017), p. 71.
- [62] Aref Takiden et al. “Structural and vibrational characterization of the chromophore binding site of bacterial phytochrome agp1”. In: *Photochem. Photobiol.* 93.3 (2017), pp. 713–723.
- [63] Vaibhav Modi et al. “Protonation of the biliverdin IX α chromophore in the red and far-red photoactive states of a bacteriophytochrome”. In: *J. Phys. Chem. B* 123.10 (2019), pp. 2325–2334.
- [64] Olle Falklöf and Bo Durbeej. “Modeling of phytochrome absorption spectra”. In: *J. Comput. Chem.* 34.16 (2013), pp. 1363–1374.
- [65] Steve Kaminski, Grazia Daminelli, and Maria Andrea Mroginiski. “Molecular dynamics simulations of the chromophore binding site of *Deinococcus radiodurans* bacteriophytochrome using new force field parameters for the phytochromobilin chromophore”. In: *J. Phys. Chem. B* 113.4 (2009), pp. 945–958.
- [66] Chen Song et al. “Two ground state isoforms and a chromophore D-ring photoflip triggering extensive intramolecular changes in a canonical phytochrome”. In: *Proc. Natl. Acad. Sci. U.S.A.* 108.10 (2011), pp. 3842–3847.
- [67] Daniel Stöppler et al. “Dynamic nuclear polarization provides new insights into chromophore structure in phytochrome photoreceptors”. In: *Angew. Chem. Int. Ed.* 55.52 (2016), pp. 16017–16020.
- [68] Yuri Alexeev et al. “GAMESS as a free quantum-mechanical platform for drug research”. In: *Curr. Top. Med. Chem.* 12.18 (2012), pp. 2013–2033.
- [69] Michael W Schmidt et al. “General atomic and molecular electronic structure system”. In: *J. Comput. Chem.* 14.11 (1993), pp. 1347–1363.
- [70] Michael Gaus, Albrecht Goez, and Marcus Elstner. “Parametrization and benchmark of DFTB3 for organic molecules”. In: *J. Chem. Theory Comput.* 9.1 (2012), pp. 338–354.
- [71] Michael Gaus et al. “Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications”. In: *J. Chem. Theory Comput.* 10.4 (2014), pp. 1518–1537.
- [72] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD: visual molecular dynamics”. In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38.
- [73] URL: <http://zzzfelis.sakura.ne.jp/>.

REFERENCES

- [74] Andrew T Ulijasz and Richard D Vierstra. “Phytochrome structure and photochemistry: recent advances toward a complete molecular picture”. In: *Curr Opin Plant Biol.* 14.5 (2011), pp. 498–506.
- [75] Jeremiah R Wagner et al. “Mutational analysis of *Deinococcus radiodurans* bacterio-phytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes”. In: *J. Biol. Chem.* 283.18 (2008), pp. 12212–12226.
- [76] Olle Falklöf and Bo Durbeej. “Computational identification of pyrrole ring C as the preferred donor for excited-state proton transfer in bacteriophytochromes”. In: *ChemPhotoChem* 2.6 (2018), pp. 453–457.
- [77] Maria Fernandez Lopez et al. “The role of the propionic side chains for the photoconversion of bacterial phytochromes”. In: *Biochemistry* (2019).
- [78] Jessica A Rumfeldt et al. “UV-Vis spectroscopy reveals a correlation between Y263 and BV protonation states in bacteriophytochromes”. In: *Photochem. Photobiol.* (2019).
- [79] Maximillian JS Phipps et al. “Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns”. In: *Chem. Soc. Rev* 44.10 (2015), pp. 3177–3211.
- [80] Andrea Schmidt et al. “Structural snapshot of a bacterial phytochrome in its functional intermediate state”. In: *Nat. Commun.* 9.1 (2018), p. 4912.
- [81] Jovan Lj Dragelj et al. “C–H/O interactions of aromatic CH donors within proteins: a crystallographic study”. In: *Cryst. Growth Des.* 16.4 (2016), pp. 1948–1957.
- [82] Ryne C Johnston and Paul Ha-Yeon Cheong. “C—H···O non-classical hydrogen bonding in the stereomechanics of organic transformations: theory and recognition”. In: *Org. Biomol. Chem.* 11.31 (2013), pp. 5057–5064.
- [83] Anders S Christensen et al. “Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications”. In: *Chem. Rev.* 116.9 (2016), pp. 5301–5337.
- [84] Dmitri G Fedorov. “Solvent screening in zwitterions analyzed with the fragment molecular orbital method”. In: *J. Chem. Theory Comput.* (2019).
- [85] Soshichiro Nagano et al. “The crystal structures of the N-terminal photosensory core module of *Agrobacterium* phytochrome Agp1 as parallel and anti-parallel dimers”. In: *Journal of Biological Chemistry* 291.39 (2016), pp. 20674–20691.
- [86] Andrea Schmidt et al. “Structural snapshot of a bacterial phytochrome in its functional intermediate state”. In: *Nature communications* 9.1 (2018), pp. 1–13.
- [87] Francisco Velazquez Escobar et al. “A protonation-coupled feedback mechanism controls the signalling process in bathy phytochromes”. In: *Nature chemistry* 7.5 (2015), pp. 423–430.
- [88] Hans Martin Senn and Walter Thiel. “QM/MM methods for biomolecular systems”. In: *Angewandte Chemie International Edition* 48.7 (2009), pp. 1198–1229.
- [89] J. Sambrook and D. W. Russell. *Molecular Cloning. A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Plainview, NY., 2001.

-
- [90] Xiaojing Yang et al. "Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome". In: *Nature* 479.7373 (2011), pp. 428–432.
- [91] Hernan Biava et al. "Long-range modulations of electric fields in proteins". In: *The Journal of Physical Chemistry B* 122.35 (2018), pp. 8330–8342.
- [92] Hiroya Nakata et al. "Simulations of Raman spectra using the fragment molecular orbital method". In: *J. Chem. Theory Comput.* 10.9 (2014), pp. 3689–3698.
- [93] Xiuling Xu et al. "Structural elements regulating the photochromicity in a cyanobacteriochrome". In: *Proceedings of the National Academy of Sciences* 117.5 (2020), pp. 2432–2440.
- [94] Casper Steinmann et al. "FragIt: a tool to prepare input files for fragment based quantum chemical calculations". In: *PLoS One* 7.9 (2012), e44480.
- [95] Ronald González and Maria A Mroginiski. "Fully Quantum Chemical Treatment of Chromophore–Protein Interactions in Phytochromes". In: *The Journal of Physical Chemistry B* 123.46 (2019), pp. 9819–9830.
- [96] Stefan Grimme et al. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu". In: *The Journal of chemical physics* 132.15 (2010), p. 154104.
- [97] Christoph R Jacob, Sandra Luber, and Markus Reiher. "Analysis of secondary structure effects on the IR and Raman spectra of polypeptides in terms of localized vibrations". In: *The Journal of Physical Chemistry B* 113.18 (2009), pp. 6558–6573.
- [98] Thomas Weymuth et al. "MOVIPAC: Vibrational spectroscopy with a robust metaprogram for massively parallel standard and inverse calculations". In: *Journal of computational chemistry* 33.27 (2012), pp. 2186–2198.
- [99] Anthony P Scott and Leo Radom. "Harmonic vibrational frequencies: an evaluation of Hartree–Fock, Møller–Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors". In: *The Journal of Physical Chemistry* 100.41 (1996), pp. 16502–16513.
- [100] Graham D Fletcher et al. "Large-scale mp2 calculations on the blue gene architecture using the fragment molecular orbital method". In: *Journal of chemical theory and computation* 8.1 (2012), pp. 75–79.
- [101] Dmitri G Fedorov et al. "A new hierarchical parallelization scheme: Generalized distributed data interface (GDDI), and an application to the fragment molecular orbital method (FMO)". In: *Journal of Computational Chemistry* 25.6 (2004), pp. 872–880.
- [102] David Buhrke et al. "Red, orange, green: light-and temperature-dependent color tuning in a cyanobacteriochrome". In: *Biochemistry* 59.4 (2019), pp. 509–519.
- [103] Christian Wiebeler et al. "The effective conjugation length is responsible for the red/green spectral tuning in the cyanobacteriochrome Slr1393g3". In: *Angewandte Chemie International Edition* 58.7 (2019), pp. 1934–1938.
- [104] Christian Wiebeler and Igor Schapiro. "QM/MM benchmarking of cyanobacteriochrome Slr1393g3 absorption spectra". In: *Molecules* 24.9 (2019), p. 1720.

REFERENCES

- [105] Kenno Vanommeslaeghe et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". In: *Journal of computational chemistry* 31.4 (2010), pp. 671–690.
- [106] Anmol Kumar, Ozge Yoluk, and Alexander D MacKerell Jr. "FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules". In: *Journal of computational chemistry* 41.9 (2020), pp. 958–970.
- [107] M. J. Frisch et al. "Gaussian09 Revision E.01". In: (). Gaussian Inc. Wallingford CT 2009.
- [108] Bernard R Brooks et al. "CHARMM: the biomolecular simulation program". In: *Journal of computational chemistry* 30.10 (2009), pp. 1545–1614.
- [109] Mark T Nelson et al. "NAMD: a parallel, object-oriented molecular dynamics program". In: *The International Journal of Supercomputer Applications and High Performance Computing* 10.4 (1996), pp. 251–268.
- [110] URL: <https://cgenff.umaryland.edu/>.
- [111] Piero Procacci. "Primadorac: A free web interface for the assignment of partial charges, chemical topology, and bonded parameters in organic or drug molecules". In: *Journal of chemical information and modeling* 57.6 (2017), pp. 1240–1245.
- [112] Kenno Vanommeslaeghe, E Prabhu Raman, and Alexander D MacKerell Jr. "Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges". In: *Journal of chemical information and modeling* 52.12 (2012), pp. 3155–3168.
- [113] Giovanni Battocchio et al. "Dynamic Properties of the Photosensory Domain of Deinococcus radiodurans Bacteriophytochrome". In: *The Journal of Physical Chemistry B* 124.9 (2020), pp. 1740–1750.
- [114] Emil Alexov et al. "Progress in the prediction of pKa values in proteins". In: *Proteins: structure, function, and bioinformatics* 79.12 (2011), pp. 3260–3275.
- [115] O Anders Borg and Bo Durbeej. "Relative ground and excited-state p K a values of phytochromobilin in the photoactivation of phytochrome: A computational study". In: *The Journal of Physical Chemistry B* 111.39 (2007), pp. 11554–11565.
- [116] Art D Bochevarov et al. "Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences". In: *International Journal of Quantum Chemistry* 113.18 (2013), pp. 2110–2142.
- [117] Li Zhang and Jan Hermans. "Hydrophilicity of cavities in proteins". In: *Proteins: Structure, Function, and Bioinformatics* 24.4 (1996), pp. 433–438.
- [118] Vaibhav Modi et al. "Protonation of the biliverdin IX α chromophore in the red and far-red photoactive states of a bacteriophytochrome". In: *The Journal of Physical Chemistry B* 123.10 (2019), pp. 2325–2334.
- [119] Vinicius Wilian D Cruzeiro, Marcos S Amaral, and Adrian E Roitberg. "Redox potential replica exchange molecular dynamics at constant pH in AMBER: Implementation and validation". In: *The Journal of chemical physics* 149.7 (2018), p. 072338.

- [120] Shigenori Tanaka et al. “Dynamic Cooperativity of Ligand–Residue Interactions Evaluated with the Fragment Molecular Orbital Method”. In: *The Journal of Physical Chemistry B* (2021).