Thesis for the Degree of Doctor of Philosophy in the Natural Sciences

Structural Features of Bacteriophytochromes

Photoactivated Proteins Studied by Serial Femtosecond Crystallography

Petra Edlund

UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology

Gothenburg 2018

Thesis for the Doctor of Philosophy in the Natural Sciences

Structural Features of Bacteriophytochromes Photoactivated Proteins Studied by Serial Femtosecond Crystallography

Petra Edlund

Cover: The surface structure of the phytochrome chromophore binding domain (CBD) from *D.radiodurans* placed on photographs of CBD protein crystals (left macrocrystals and right microcrystals).

ISBN: 978-91-629-0472-2 (Print)
ISBN: 978-91-629-0473-9 (PDF)
Available online at via http://handle.net/2077/55645

Department of Chemistry and Molecular Biology
Division of biochemistry and structural biology
University of Gothenburg
SE-405 30 Göteborg, Sweden
Printed by BrandFactory AB
Göteborg Sweden, 2018

Till John, Ivar och Lillebror

Abstract

The key to life on earth is sunlight, which reaches the planet as an energy source. Nature has evolved different types of photoreceptor proteins to detect optimal light conditions for biochemical processes. A type of red light detecting photoreceptor proteins are called phytochromes and are present in plants, fungi and bacteria. A chromophore, converts the light signal into a structural change in the protein that alter its biochemical properties and thereby control developmental processes in the organism. A structural mechanism for signal transduction within the phytochrome protein is herein proposed.

The aim of the work presented in this thesis has been to elucidate the structural changes in bacteriophytochromes upon photoactivation. This has been done by the use of X-ray crystallographic methods that can provide a near-atomic resolution of the dynamic events. Crystallization strategies were developed to experimentally obtain novel structural information on bacteriophytochromes from both conventional crystallography and by Serial Femtosecond Crystallography at X-ray Free electron lasers. The method enable time-resolved structural studies with an ultrafast time-resolution due to the X-ray lasers short pulses.

Novel crystallization conditions for a bacteriophytochrome fragment yielded near-atomic resolution structures of both the wild type and a muted variant. The conditions could be modified for microcrystallization that provided microcrystals suitable for two different sample delivery systems at the world's two most prominent X-ray lasers. The obtained resting state structures and a preliminary data set of the excited state paves the way for future time resolved investigation on the early structural events in photoactivation of phytochromes. Furthermore, the microcrystallization strategies might be applicable to other proteins and are thereby contributing to method development within the emerging field.

The crystallographic structure of the mutated variant of the protein fragment supports IR-spectroscopy findings on the importance of the hydrogen bonding network around the chromophore. These results are in agreement with the excited state structural findings that waters might be of highest importance for the initial steps in the photoactivation of phytochromes.

Publications

This thesis consists of the following research papers together with an extended summary of my PhD work

Paper I: Petra Edlund, Heikki Takala, Janne.A Ihalainen, Sebastian Westenhoff. ”Structural Mechanism of Signaling in Bacteriophytochromes. “ Manuscript (2018)

Paper II: Petra Edlund,* Heikki Takala,* Elin Claesson,* Léocadie Henry, Robert Dods, Heli Lehtivuori, Matthijs Panman, Kanupriya Pande, Thomas White, Takanori Nakane, Oskar Berntsson, Emil Gustavsson, Petra Båth, Vaibhav Modi, Shatabdi Roy-Chowdhury, James Zook, Peter Berntsen, Suraj Pandey, Ishwor Poudyal, Jason Tenboer, Christopher Kupitz, Anton Barty, Petra Fromme, Jake D. Koralek, Tomoyuki Tanaka, John Spence, Mengning Liang, Mark S. Hunter, Sebastien Boutet, Eriko Nango, Keith Moffat, Gerrit Groenhof, Janne Ihalainen, Emina A. Stojković, Marius Schmidt & Sebastian Westenhoff. ‘‘The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography” Scientific Reports 6 35279 (2016) doi: 10.1038/srep35279 *Equally contribution.

Paper III: Nils Lenngren*, **Petra Edlund***, Heikki Takala*, Brigitte Stucki-Buchli, Ivan Peshev, Heikki Häkkänen, Sebastian Westenhoff, and Janne Ihalainen, “Coordination of the Biliverdin D-ring in Bacteriophytochromes”. *Submitted* (2018) *Equally contribution.

Paper IV: Nicole C. Woitowich, Andrei S. Halavaty, Patricia Waltz, Christopher Kupitz, Joseph Varela, Gregory Tracy, Kevin D. Gallagher, Elin Claesson, Takanori Nakane, Suraj Pandey, Garrett Nelson, Rie Tanaka, Eriko Nango, Eiichi Mizohata, Shigeki Owada, Kensure Tono, Yasumasa Joti, Angela C. Nugent, Hardik Patel, Ayesha Mapara, James Hopkins, Phu Duong, Dorina Bizhga, Svetlana E. Kovaleva, Rachael St. Peter, Cynthia N. Hernandez,Wesley B. Ozarowski, Shatabdi Roy-Chowdhuri, Jay-How Yang, **Petra Edlund**, Heikki Takala, Janne Ihalainen, Jennifer Scales, Tyler Norwood, Ishwor Poudyal, Petra Fromme, John Spence, Keith Moffat, Sebastian Westenhoff, Marius Schmidt, & Emina A.Stojković.”Structural basis for light control of cell development revealed by crystal structures of a Myxobacterial phytochrome” *Submitted* (2017)

Additional Publications

Publications to which I contributed during my PhD but that is not related to the work on bacteriophytochromes

Paper V: Sebastian Westenhoff, David Paleček, **Petra Edlund**, Philip Smith, and Donatas Zigmantas.”Coherent Picosecond Exciton Dynamics in a Photosynthetic Reaction Center” JACS 40 16484-16487 (2012) doi: 10.1021/ja3065478

Paper VI: David Palecek, **Petra Edlund**, Sebastian Westenhoff, Donatas Zigmantas “Quantum coherence as a witness of vibronically hot energy transfer in bacterial reaction center” Science Advances 3 9 2017 doi:10.1126/sciadv.1603141

Paper VII: Egle Bukarte, David Palecek, **Petra Edlund**, Sebastian Westenhoff, Donatas Zigmantas. Revealing the precursor state to charge separation in bacterial reaction centers. (Manuscript 2018)

Paper VIII: Petra Edlund, Erin M Tranfield, Vera van Noort, Karen Siu Ting, Sofia Tapani, Johanna Hoog “Gender balance in time-keeping at life science conferences. Submitted. (2018)

Contribution report

Paper I: I took the main responsibility for the literature search and for compiling the information into a review. I wrote the major part of the paper and made all the figures.

Paper II: I purified the protein and developed crystallization conditions. I designed the strategies to obtain microcrystals and I participated in data collection at ESRF, LCLS and SACLA. I processed parts of the data and took part in solving the structures. I contributed to the writing of the paper and the making of figures

Paper III: I crystallized the protein and fished the crystals. I took part in refining the structure. I wrote parts of the paper and made figures.

Paper IV: I purified and the crystallized protein and collected part of the data at SACLA. I commented on the paper.

Abbreviations

Here follows a list of abbreviations used in the thesis

BphP(s) Bacteriophytochromes (phytochrome protein from bacteria)
BV Biliverdin (chromophore)
CBD **C**hromophore **b**inding **d**omain (PAS-GAF together)
CCD **C**harge-**C**oupled Device (type of detector)
CSPAD **C**ornell-**S**LAC **P**ixel **A**rray **D**etector (type of detector)
***Dr*BphP** Bacteriophytochrome from *Deinococcus radiodurans*
EM **E**lectron **M**icroscopy
ESRF **E**uropean **S**ynchrotron **R**adiation **F**acility (X-ray source in France)
FID **F**ree **I**nterface **D**iffusion (crystallization method)
GAF c**G**MP phosphodiesterase/**a**denylate cyclase/**F**hlA transcriptional activator (protein domain)
GDVN **G**as-**D**ynamic **V**irtual **N**ozzle (injector equipment)
HK **H**istidine **k**inase (protein domain that phosphorylates)
IPTG **I**so**p**ropyl β-D-1-**t**hio**g**alactopyranoside (chemical that induces protein expression)
LCLS **L**inac **C**oherent **L**ight **S**ource (X-ray source in Stanford US)
LCP **L**ipidic **C**ubic **P**hase (crystallization method)
OPM **O**ut**p**ut domai**n** (signaling protein domain)
PAS Per-ARNT-Sim (protein domain)
PDB **P**rotein **d**ata **b**ank (online databank for protein structures)
PEG Polyethylene glycol
PHY Phytochrome-specific (protein domain)
PSM **P**hoto**s**ensory domai**n** (protein domain)
RMSD **R**oot **m**ean **s**quare **d**eviation
SACLA **S**Pring-8 **A**ngstrom **C**ompact free electron **La**ser (X-ay source in japan)
***Sa*BphP** Bacteriophytochrome from *Stigmatella aurantiaca*
SFX **S**erial **F**emtosecond **C**rystallography
TR-SFX **T**ime **r**esolved **S**erial **F**emtosecond **C**rystallography
WT **W**ild type (original protein without mutations)
XFEL **X**-ray **f**ree **e**lectron **l**aser
Å **Å**ngström (measurement of distance corresponding to 0.1nm)

Content

CHAPTER 1

1. Introduction

1.1. Light is crucial for all life

The sunlight is crucial for all life on earth. Nature has evolved ways for organisms to harvest the light energy and store it as chemical energy, to use it for physiological vision and to detect the optimal light conditions for survival. Photoreceptors are proteins attached to one or several molecules, so called chromophores, with the ability to absorb light photons. A number of different photoreceptors have developed to detect photons of different energy. Together they cover a great part of the visible light spectrum. The absorption properties of the chromophore are fine-tuned by the protein environment that surrounds it. Whereas some proteins in photosynthetic organisms (plants, cyanobacteria) are able to use the absorbed energy for electron transfer and store it as chemical energy in the cell, other photoreceptors are signaling proteins that react to a photon as a trigger for further signaling in the cell to control developmental processes. My PhD work during the last years has included investigations on both types of photoactive proteins but the thesis will focus on the work on phytochromes, which are photoreceptors involved in signaling in the cell in response to the environment's light conditions. They are for example, the reason that all the trees in a forest grow to the same height. The lower trees are shaded by higher trees and this is sensed by the phytochromes, that signal to the plant to grow higher to reach the sunlight.[1]

1.2. Protein structure and function

Proteins are complex macromolecules that are responsible for the majority of the essentially biochemical processes that occur in the cells. From DNA replication, metabolism of nutrients, building up new molecules, to transport of molecules and signaling. A protein´s function is dependent on its structure that builds up from the primary amino acid sequence. There are 20 natural occurring amino acids with structurally and chemically different side chains and the particular amino acid composition will give the protein its properties. Amino acids give the protein its structure and ability to work as catalysts for chemical reactions and are involved in all kind of interactions ranging from the ability to bind cofactors, such as chromophores, or to form the proteins quaternary structure or protein-protein interactions.

The structural information for all proteins is embedded in DNA of the organism. The DNA is read and the code is transferred as messenger RNA to a ribosome. The ribosome translates the code into an amino acid sequence and builds the protein chain. The protein often spontaneously folds into its native confirmation that is encoded in the primary amino acid sequence. The sequence is often arranged so that hydrophobic parts of the protein are embedded in the core of the protein whereas more hydrophilic parts are exposed to the surrounding solvent. The correct fold of the

protein is essential for its function and the aim for the protein is to reach the lowest energy conformation.[2] The protein's energy landscape can be seen as a map where the protein's conformation coordinates corresponds to states with different free energy. The proteins struggle to find the correct fold can be described as rolling a ball at this map. The ball would eventually fall down in the global energy minima, which represents the protein's correct fold. A one dimensional representation of this can be seen in Figure 1.1. The same concept can be used to describe proteins dynamics. If the protein can adopt different conformations the energy landscape's global minima can have several minima representing the substates (Figure 1.1 right.) Dependent of the protein, different amount of activation energy might be needed to switch in between the two.[3,4]

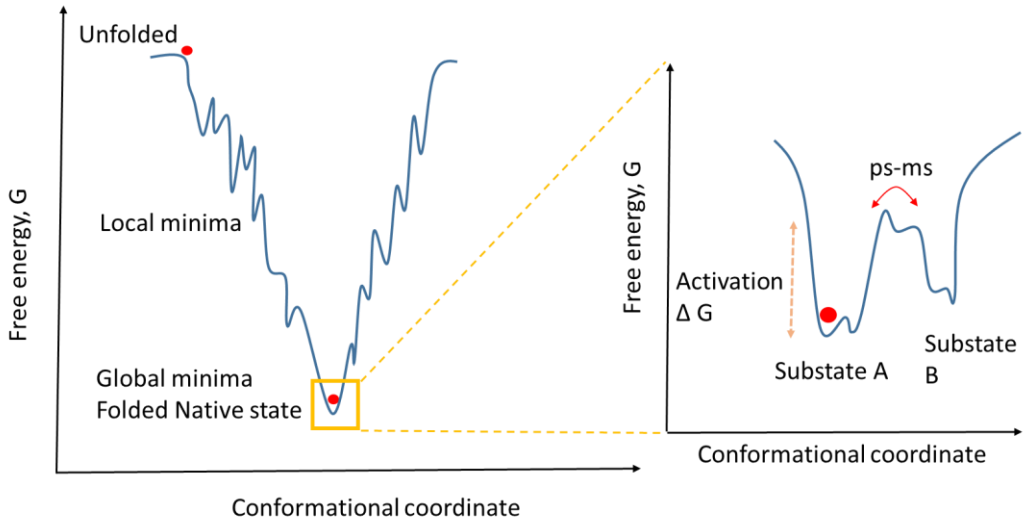

Figure 1.1. The relation between a proteins conformational state and its free energy. One dimensional representation of the energy funnel for protein folding (left) and a representation of two substates A and B in the proteins native state (right). By the addition of the needed activated energy the protein I substate A can undergo conformational changes and adopt the conformation of substate B in a ps to ms timescale.

Proteins work as catalysts of biochemical reactions, meaning that they lower the activation energy needed to conduct the reaction. Proteins can perform reactions at milder conditions and with several orders of magnitudes greater reaction rates compared to normal conditions.[5] The function of the protein is closely related to its structure and dynamics.[6] Structural changes in proteins can be both large and small, ranging from movements on the ångström, Å (0.1nm) to nanometer scale. The dynamic events can include global rearrangements like unfolding and refolding of secondary structure elements. As well as being small changes such as a rotation or slight spatial shift of a single amino acid side chain. The timescales for the dynamics can be of a great variety ranging from femtoseconds to seconds.[3,7] Although mutational studies of conserved amino acids can identify residues responsible for the proteins function, three dimensional structural information of the protein is of highest importance in identifying the functional mechanism in the protein of interest.

1.3. Structural determination of proteins

Even though proteins are large complexes and can contain hundreds of thousands of atoms, they are too small to be visualized under a microscope. Therefore the use of X-ray radiation has evolved as the most common method for structure determination of proteins. Nucleic Magnetic Resonance Spectroscopy (NMR), can also be used but is limited to smaller proteins due to its complex analysis. Cryo-electron microscopy is a growing field for protein structure determination, but does not yet reach as good resolution as X-ray crystallography. The first protein structure to be solved by X-ray crystallography was Myoglobin in 1958.[8] Since then the accumulated structures deposited in the protein data bank (PDB) (www.pdb.org) has passed 120 000 (solved by different methods).

1.3.1. X-ray crystallography

X-ray crystallography is a method that accounts for the majority of solved structures of proteins. Briefly, it is performed by making crystals of protein and illuminate them with X-rays to record diffraction patterns as a fingerprint of the specific protein structure. The use of crystals is needed because the diffraction pattern is a consequence of constructive interference of all the well-ordered molecules within the crystal.[9] The X-rays needed are often produced by a type of particle accelerator called synchrotron. The use of synchrotron radiation for structure determination has shown to be extremely successful over the last decades, collecting diffraction data from all kind of macromolecules.[10] The use of cryo temperatures to reduce radiation damage and the possibility to focus the beam to collect data on small crystals (10-15 µm) have further improved the method.[11-13] However, the method has its limitations. First, it requires large, well diffracting crystals which is the bottleneck in crystallography. Some proteins lack the ability to form large crystals and membrane protein are difficult to crystallize due to their hydrophobic nature. Second, radiation damage cannot be fully avoided and can hamper the native structural determination.

1.3.2. Serial femtosecond crystallography

The limitations with conventional crystallography along with the development of extremely powerful X-ray Free-electron lasers (XFELs) has led to the development of serial femtosecond crystallography, (SFX). XFELs provide exceptionally brilliant, micro focused X-ray pulses with an ultra-short (femtosecond) duration. SFX uses the XFEL pulses for probing of micro (to nano) crystals in a serial way.[14] This means that each crystal interacting with an X-ray pulse provides a diffraction pattern. With enough collected diffraction patterns, these can be merged together to comprise a data set, informative enough to solve the protein structure. The technique evades the need for large crystals and the ultrashort pulses and replenishment of new sample for each diffraction pattern eliminate the radiation damage problem.

SFX also opens up for time resolved studies of ultrafast structural dynamics thanks to the extremely short XFEL pulses. The first protein structure solved by SFX was the large membrane photosystem complex I in 2011.[15] Now the number of proteins structures

solved by SFX is increasing for each year. The amount of entries in the PDB were close to 150 in March 2017.[16] One way to take the structural studies one step further is to study the structural dynamics in proteins in a time-resolved manner. The dynamics of the proteins are the structural changes that a protein undergo while moving in the energy landscape as discussed previously. To alter between structural states a trigger that corresponds to the activation energy is needed for the transition. For photoactivated proteins it can be a short laser pulse. Furthermore, a method that can probe the structural dynamics with a time resolution that is fast enough to capture the structural events is required. This is possible with SFX due to the ultrashort X-ray pulses and it provides a method to make 'molecular movies' of structural dynamics in proteins.[17]

1.4. Phytochromes

Phytochromes are photosensory proteins activated by red light and are present in plants, fungi and bacteria. They are crucial for the organism physiological responses to the light environment in which they live. In the following section phytochromes function and structural features are discussed. A special emphasis is put on phytochromes present in bacteria, bacteriophytochromes (BphPs), which have been the proteins of interest in this thesis.

1.4.1. Discovery and function

The word phytochrome means 'plant color' and was given to the unknown component in plants, which enabled them to respond to changes in daylight almost 100 years ago.[18] The responsible protein for red light detection was discovered in the 1950s by the illumination of seeds with red light.[19] The response was found to be reversed to by far red light illumination. This suggested the presence of a photo reversible pigment, which was confirmed by absorption spectroscopy 1959.[20] In 1983 the protein was purified[21] and two years later the amino acid primary sequence was revealed.[22] Later, phytochromes have also been discovered in funghi[23] and in prokaryotes. Initially in photosynthetic cyanobacteria[24,25] and later also in non-photosynthetic bacteria.[26]

In plants, phytochromes have been shown to be important for various developmental processes such as, for example, shade avoidance, flowering time, and stem elongation.[27-29] Plants grown in sunlight are lower and bushier than plants grown in the shadow, which are taller and have smaller leaves (Figure 1.2a). The great interest in phytochromes comes from the attempt to better understand and to control the development of plants, especially crops.[30]

In later years the phytochromes have become more interesting in the fields of optogenetic and deep tissue imaging due to their simple modular architecture and red light absorption properties.[31-35]

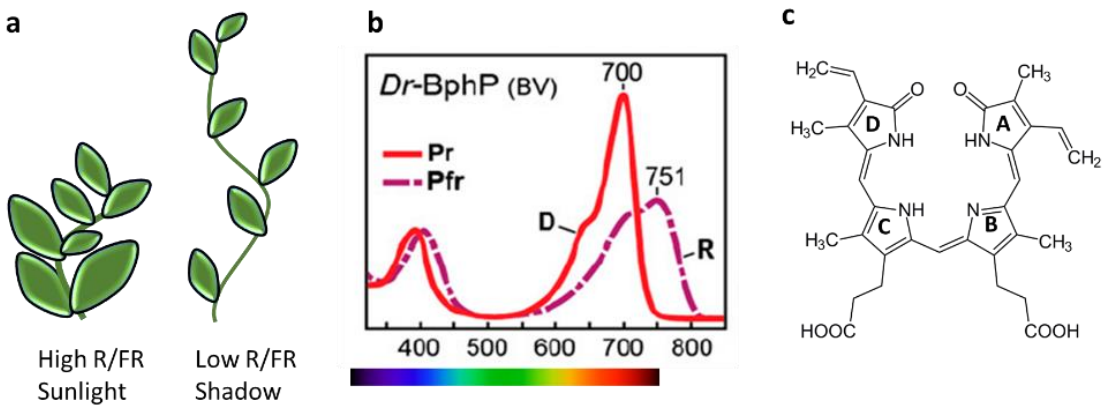

Figure 1.2. Plant development and light detection in phytochromes. a) The different physiological appearance of plants grown in light versus shadow. b) The absorption spectra of a bacterial phytochromes two metastable substates called Pr (red light absorbing) and Pfr (far-red light absorbing). Imported from[36] c) The structure of the biliverdin chromophore and its conjugated system of double bonds that are responsible for photon absorption in phytochromes.

The phytochromes role in plants is established, but their physiological function in non-photosynthetic bacteria remains unknown. Although BphPs function might be distinct from plants, they greatly resemble plant phytochromes in structure and absorption properties.[36-38] The possibility to produce large amounts of bacteriophytochromes, by recombinant protein expression, enables a simple way to conveniently study and crystallize BphPs. Gaining knowledge about them and their plant relatives by proxy.[39,40]

Phytochromes work as light controlled switches by being able to adopt two metastable states that are named after their absorption properties, Pr for the red light absorbing (A_{MAX} ~700 nm) and Pfr for the far red light absorbing (A_{MAX} ~750 nm). The absorption maxima of the Pr and Pfr states vary between species. However, the difference in absorption maxima between Pr and Pfr is usually 50 ‑60 nm for BphPs with a biliverdin chromophore.[36] For the absorption spectra of the two states of *Deinoccous radiodurans* bacteriophytochrome *Dr*BphP see Figure 1.2b. The responsible molecule for the light sensing ability in phytochromes is a covalently bond bilin molecule (Figure 1.2c). The bilin is a heme-derived linear tetrapyrrole with a conjugated double bond system, which enables it to absorb light. It consists of four rings named A, B, C and D. Upon light absorption the excited biliverdin molecule can undergo a cis to trans isomerization over its double bond between ring C and D (Pr, *ZZZssa* and Pfr *ZZEssa*).[41,42] The isomerization involves a rotation of the D-ring that creates rearrangements in the protein structure and alternation of the absorption properties, hence the switch to the Pfr state.

1.4.2. Phytochromes role in cell signaling and their modular structure

Phytochromes are built up from an N-terminal photosensory module (PSM), which detects the light signal and an output domain (OPM), which transfers the signal further on in the cell (Figure 1.3a). The OPM is often a histidine kinase (HK). Phytochromes

harboring a HK are part of a so called two component signaling system, which they form together with their partner, the response regulator (RR), protein. The RR can be phosphorylated by the HK and is able to interact with DNA to control gene expression (Figure 1.3a).[43] Phytochromes can also hold other types of OPMs and be involved in other signaling pathways.[37,44] The variety of the bond OPMs is proposed to be nature's way to create a signal variability although using the same type of detection mechanism by the PSM.[45]

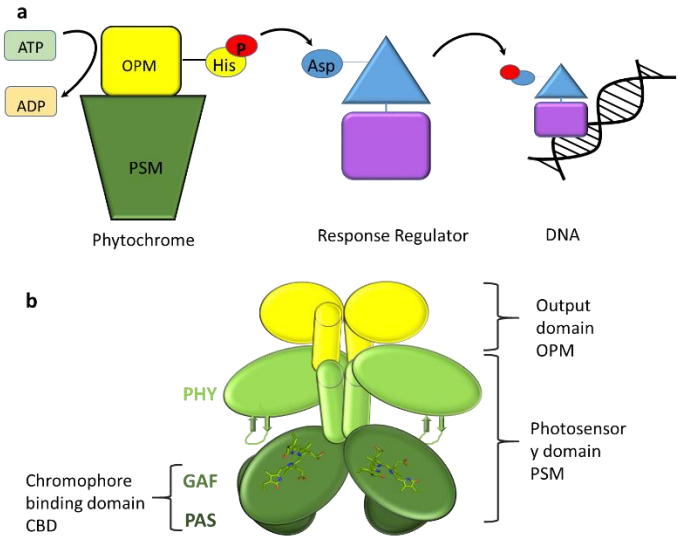

Figure 1.3 The mode of action of phytochromes. a) The bacteriophytochromes role in a cells two components signaling system is to phosphorylate a response regulator, which can interact with the cell´s DNA. b) The homodimeric parallel modular architecture of phytochromes with the photosensory domain (PSM) in green and the output domain (OPM) in yellow. The PAS and the GAF domain together form the chromophore binding domain, CBD.

BphPs, classified as group I phytochromes (plant, cyanobacteria and BphPs) contains a PSM with a conserved architecture with three different, but structurally related domains (Figure 1.3b).[46] The N-terminal PAS (Per-ARNT-Sim) and following GAF (cGMP phosphodiesterase/adenylate cyclase/FhlA transcriptional activator) domains together forms the chromophore binding domain (CBD). The signal transduction PHY- (phytochrome-specific) domain extends out from the CBD and binds to the OPM.[37,47] BphPs are normally homodimeric in solution and typically adopt a parallel head to head arrangement with the two sister monomers twisting around each other. The dimerization interactions are located between the two CBDs and the OPMs, as visualized by cryo-electron microscopy, EM.[48]

1.4.3. General structural features of BphPs

A great number of crystals structure of BphPs fragments of different lengths and from different species have been revealed.[38,42,44,49-67] The first structure of a phytochrome fragment to be solved was the CBD fragment from *Deinococcus radiodurans* in its Pr (dark adapted) state.[49] Two years later a mutated variant with improved atomic resolution was released.[49,51] Later the same fragment from the same and other species

have been crystallized multiple of times[49,51-56,68,69] providing information on the chromophore configuration and interactions in the chromophore binding pocket. In bacteria the tetrapyrrole biliverdin (BV) (Figure 1.2c) is covalently attached to a conserved cysteine residue in the PAS domain and in cyanobacteria the closely related Phycocyanobilin (PCB) is attached to the GAF-domain.[37] Although the bilin attachment is in different parts of the amino acids sequence (PAS for bacteria and GAF for cyanobacteria) the spatial position is the same, embedded in the GAF domain (Figure 1.4a.).[47] The CBD structures revealed an unusual structural motif of a figure eight knot, between the PAS and GAF domains. The knots function is unknown.

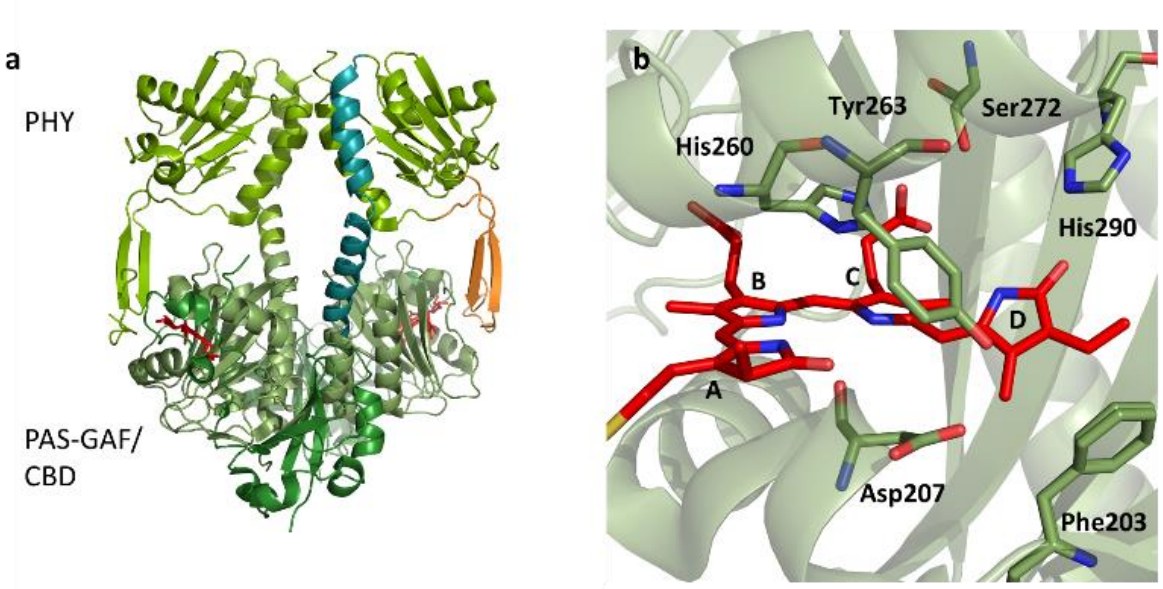

Figure 1.4. Structural features of BphPs. a) The general Pr structure of a bacteriophytochrome PSM shows a head-to-head parallel dimer arrangement (PDB id 4O01). The PAS, GAF and PHY domain are colored in different shades of green. The biliverdin is colored red. In the right monomer the PHY-tongue is colored orange and the helical spine teal. b) Zoom in view of the chromophore (red) and selected surrounding amino acids shown important for signal transduction (PDB id 4Q0H).

The high resolution structures of the CBD can provide detailed information about the chromophore's interactions with surrounding amino acids. (Figure 1.4b). This enables structural interpretations of mutagenesis studies of the conserved residues. It has been determined that the incorporation of the chromophore is quite robust, whereas the ability to properly photo convert is often impaired by single site mutations of the highly conserved residues around the chromophore.[70] A lot interest has been focused on the so called DIP-motif (conserved aspartate, isoleucine and proline). The mutation of the aspartate (D207 in *Dr*BphP) makes the protein unable to photo convert and instead forms a fluorescent variant.[51,53] Other identified important residues are a histidine (260 in *Dr*BphP) and a tyrosine (263 in *Dr*BphP). Another is a histidine (H290 in *Dr*BphP) that forms a hydrogen bond with the carbonyl of the D-ring and is proposed to stabilize the Pr state (Figure 1.4b).

The first crystal structures of the complete photosensory module, (PSM) containing the CBD and the PHY-domain were solved in 2008. The Cph1 from cyanobacterium

Synechocystis sp. PCC6803.[58] and *Pa*BphP from *P.aeruginosa*. [57] Both structures show dimeric assembly. However, the Cph1 structure was crystallized in an antiparallel arrangement of the dimer whereas the *Pa*BphP showed a parallel arrangement. In later structures the parallel organization was identified as the most common and general arrangement among phytochromes (Figure 1.4a). The PHY-domain holds a long helical spine that extrudes from the CBD and ends in globular shape with a β-sheet and α-helices. From this the so called PHY-tongue stretches back to the GAF domain to interact with the chromophore binding region (Figure1.4a).

The OPM is positioned on the PHY-domain as confirmed by cryo-EM[48] and solution scattering experiments of *Dr*BphP.[71] At the time of writing, there is still no available crystal structure of a full length phytochrome with a HK attached. However, three BphPs with other OPMs have been crystallized.[65-67] Two of those exhibit a parallel dimer organization[66,67] and one has a confirmed enzymatic activity.[67]

1.4.4. The photocycle of BphPs

The BphP photocycle (Figure 1.5) is well studied by vibrational spectroscopy (IR or Raman) or time resolved transient spectroscopy. The methods have identified intermediates and the kinetics of their formation and thereby provided information about the protein function long before any structures were available.[72-79] The spectroscopical investigations have revealed three intermediates upon photoactivation, before it reaches the Pfr state, named Lumi-R, Meta-Ra and Meta-Rc (Figure 1.5). The formation of the Lumi-R state is fast and occurs on a picosecond timescale.[75-77,80-82] The formations of the Meta-states are slower and involves a deprotonation and a reprotonation step before the protein can relax into the Pfr state.[83] BphPs can then revert back to its relaxed Pr state either via thermal dark reversion or illumination by far-red light. The Pfr to Pr transition follow a similar but not identical pathway (Figure 1.5).[84,85] Thermal dark reversion is slow and the protein often exists in a mixture of the two states.

Under dark conditions prototypical BphPs relax to the Pr state through thermal dark reversion. These are called canonical phytochromes. However, some BphPs show reversed dark adaption, relaxing to Pfr as their resting state. These are called bathy phytochromes (Figure 1.5.) Hence, the difference between canonical and bathy phytochromes lies in the direction of the thermal dark reversion and neither the mechanism nor the reason for the evolvement of the two types is fully understood.

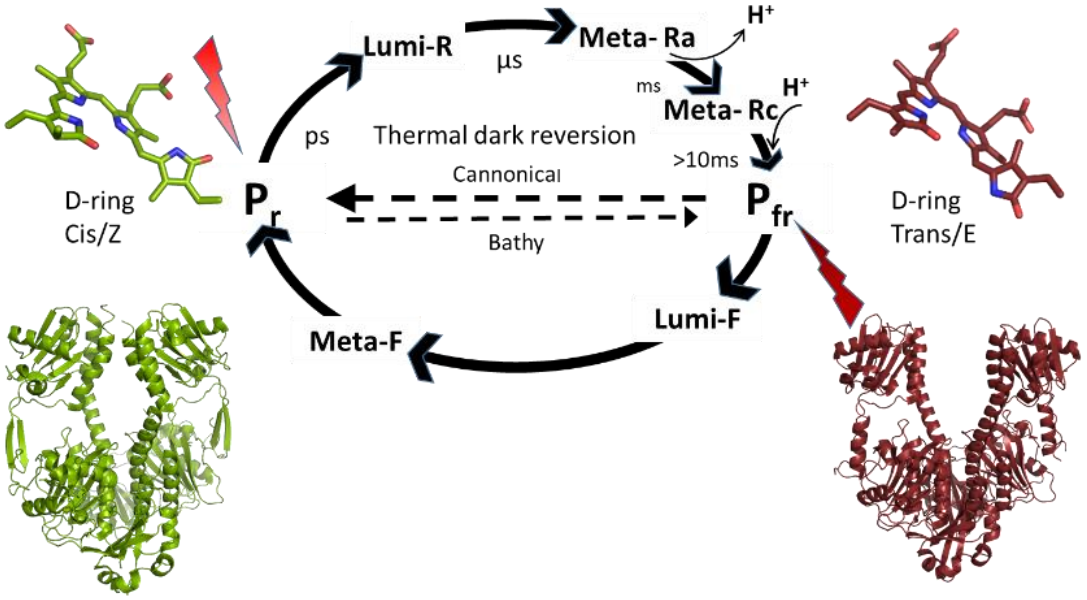

Figure 1.5. The photocycle of BphPs. The Pr state can be converted to the Pfr state with illumination of red light passing by Lumi and Meta intermediates states and the Pfr can by illumination by far-red light be converted to the Pr state. For the two states the chromophore configuration and the structural changes in the PSM are demonstrated, Pr in green (PDB 4Q0P) and Pfr in ruby (PDB 4O01). The directions for the thermal dark relaxation in canonical vs bathy phytochromes are demonstrated by opposing arrows.

1.4.5. Signaling mechanism in BphPs

The structures of BphPs in their different states connected to the photocycle is extensively described in **Paper I,** but is briefly introduced here together with the signaling structural mechanism that we propose in the paper.

The early PSM structures describe the dark state from both a prototypical [58,61] and a bathy phytochrome,[57,59] which enabled the comparison of structural features between the Pr and Pfr states.[86] However, the fact that the structures originated from different species made it difficult to determine if the structural features originated from the state of the protein or variances between species. The most striking structural difference between the two was the distinct fold of the PHY-tongue that showed a β-sheet structure in the canonical phytochromes and an α-helical structure in the bathy ones. The refolding of the PHY-tongue was confirmed for the Pr to Pfr transition when the PSM from *Dr*BphP was crystallized in both its dark state and a light enriched state (Figure 1.5).[63] The model was later improved when the light state was crystallized with a higher Pfr content due to a mutation that impedes thermal back reversion.[42] The structures revealed that the photoactivation of *Dr*BphP PSM leads to a refolding of the PHY-tongue and opening of the PHY-domains by several nm (Figure 1.5). The large structural change was further confirmed to occur also in solution by solution scattering experiments.[63] A similar change was later proposed to take place in several other PSM of BphPs.[87]

An overview of the dynamics of the Pr to Pfr transition is presented in Figure 1.6. It is now accepted that the Pr to Pfr conversion involve the cis-trans isomerization of double bond between the C and D-ring in the BV. This leads to a rearrangement of the chromophore and alternations of its interactions with the protein matrix. How this happens is not yet fully understood but it is translated into to a refolding of the tongue and a straightening of the helical spine.

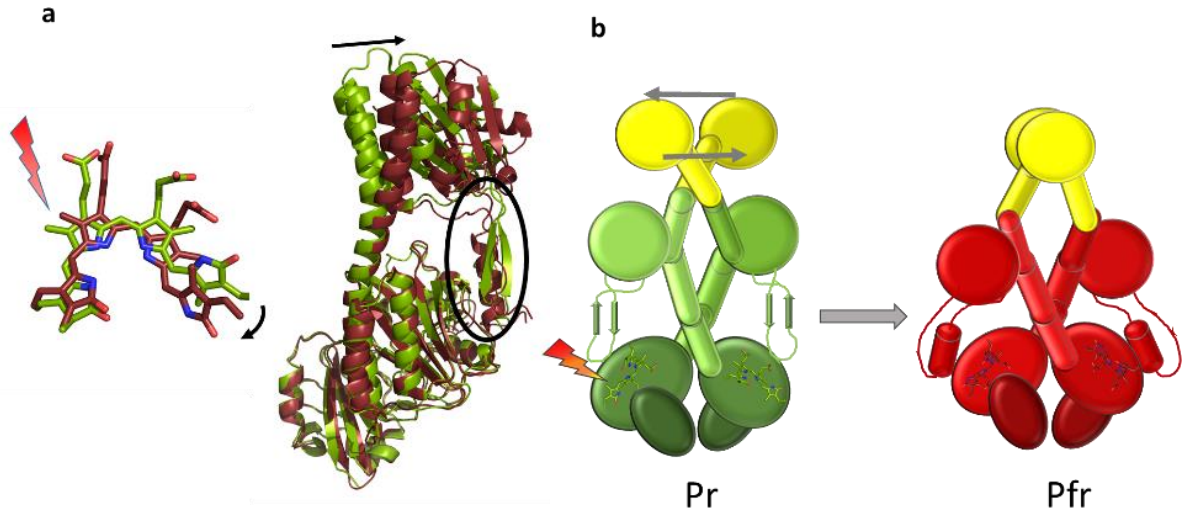

Figure 1.6. Photoactivation of bacteriophytochromes induces structural changes in the protein. a) The absorption by the chromophore leads to an isomerization of the D-ring. The structural change is relayed to the protein leading to a straightening of the helical spine and a displacement of the PHY-domain. Pr is shown in green and Pfr in ruby. b) The global structural changes upon light activation in the full-length BphP. The changes and the straightening of the helical spine in the PHY-domain is translated into a rotational movement of the OPM. Coloring as in a.

In the PSM fragment the tongue refolding/helical spine straightening leads to a rearrangement of the PHY-domains that make them move in opposite directions (Figure 1.4). The helical spine conformation can indeed be associated with the state of the protein, Pr having a bent/kinked structure and Pfr a more straight conformation.[64] However, the PHY-domain adopts many different orientations in the solved PSM structures and the Pfr structure of *Pa*BphP from *Pseudomonas aeruginosa* does not crystallize with separated PHY-domains.[57,59] The orientation of the PHY-domain cannot be generally associated with the proteins state.[67] Instead the many orientations in the crystal structures indicates a high plasticity of the PHY-domain. This might reflect the difficulties in obtaining well diffracting protein crystals of these fragments.

Based on the opening of the PHY-domain in the PSM crystal structures from *D.radiodurans* together with negative staining electron microscopy, Burgie et al.[54] proposed a structural mechanism where the HK monomers in the OPM follows the PHY-domains movement and dissociate upon light activation.[54] In **Paper I** we propose an alternative structural mechanism where the HK stays intact and the induced strain in the PHY-domain instead makes the HK domains rotate in relation to the PSM. (Figure 1.6b) This is based on time resolved solution scattering experiments on both the full

length *Dr*BphP[71] and a chimeric sensor histidine kinase YF1,[88] which identifies a rotation of the OPM upon light activation. The rotation mechanism is further strengthened by crystal structures of other sensor histidine kinases[89] together with the findings that HKs need dimerization interactions for their activity.[90]

1.4.6. Bacteriophytochrome from *Deinococcus radiodurans*

D. radiodurans is a polyextremophilic bacterium that can survive cold, dehydration, vacuum, acidity and foremost radiation. In fact, it is one of the most radiation resistant organisms known.[91] It is found in environments rich in organic materials such as soil, feces, dust and food.[91,92]*D. radiodurans* is not only known to tolerate high doses of radiation it is also has a unique ability to repair DNA. The Bacteriophytochrome from *D. radiodurans*, *Dr*BphP, is one of the most studied bacteriophytochromes and its CBD was the first phytochrome fragment that was structurally determined.[49]. However, the physiological function of *Dr*BphP remains unknown. In was earlier proposed that the phytochrome might be involved in carotenoid production in the bacterium[26] but this not been confirmed.

1.4.7. Bacteriophytochrome from *Stigmatella aurantiaca*

S. aurantiaca is a soil bacterium included in the family of myxobacteria that have the ability to group together into so-called fruiting bodies, and move around together, as a big lump of bacteria.[26,93] Fruiting bodies can vary in size between 50-500 µm and can be observed under a microscope.[94,95] *S. aurantiaca* expresses two bacteriophytochromes called *Sa*BphP1 and *Sa*BphP2, which differ slightly from each other. The *Sa*BphP2 acts as a prototypical phytochrome but the *Sa*BphP1 variant lacks a very well conserved histidine (289, 290 In *D.radiodurans*)[96] Instead it harbors a threonine at this position which results in incomplete conversion to the Pfr state upon illumination but the ability is obtained by introduction of a histidine.[87]

1.5. Scope of the thesis

In this thesis, the structure determination of phytochrome protein fragments was performed by X-ray crystallography both at a synchrotron and at XFELs. The aim was to elucidate structural changes connected with the photoactivation of the protein.

Paper I is a review of the current structural knowledge on bacteriophytochromes related to their state in the photocycle. It discusses the structures obtained from X-ray crystallography in combination with other methods to get an overview of important structural features related to photoactivation and signal transduction within the protein. Based on current knowledge, we propose a structural mechanism for the signal transduction from the photosensory module to the output module.

To structurally study structural features of phytochromes well diffraction crystals are needed. Furthermore SFX put additional demands on crystallization by the requirements of microcrystals.

Chapter 2 introduces the methods for protein production, crystallization and structural determination used in this thesis work.

Chapter 3 describes the conventional crystallization and structure determination of the wild type CBD and a H290T mutant presented in **Paper II** and **Paper III** respectively**. Paper III** investigate the role of a conserved histidine in the protein by vibrational spectroscopy. The findings are supported by the crystal structure of the H290T mutant.

Chapter 4 describes how the crystal conditions in **Paper II** were modified to obtain microcrystals for SFX experiments by two different methods also presented in **paper II**. It also describe how several microcrystallization strategies for the phytochrome PSM from different species were developed. It finally lead to a SFX structure of the *Sa*BphP1 PSM presented in **Paper IV**.

Chapter 5 presents the primarily results from TR-SFX studies of the CBD and a possible excited state structure of the protein.

CHAPTER 2

2. Methodology

2.1. Protein expression and purification

2.1.1. Recombinant protein expression in *E-coli*

To be able to structurally study the protein of interest it must be available in enough amounts for crystallization. Furthermore it should be of highest purity and have a low content of other macromolecules to form homogenous crystals. Conventional crystallization often requires a minimum of micro to milligrams of protein. However, by exploring different crystallization conditions and the need for numerous optimization cycles can require endless amounts. Additionally, SFX experiments can consume extremely high amounts of samples due to low hit rate and high consumption by time unit. The majority of structural biology targets are expressed in very low amounts in their host cell. Fortunate, the method of recombinant protein expression has provided a solution.[97,98] The gene of the protein of interest is cloned and inserted into a DNA-vector especially designed for expression purposes. The vector is further transformed in to an *E-coli* strain (or another expression host) where the protein is expressed in high yields.[99,100]

Often the vector of choice in recombinant expression in *E.coli* contains the so called *lac* operon which can control the expression by the regulation of allolactose concentration. Under normal conditions a DNA binding protein, called lac repressor, binds to the *lac* operon and inhibits protein translation. The *E.coli* cells are grown to the optical density measured at 600 nm and then protein expression is induced by the addition of Isopropyl β-D-1-thiogalactopyranoside, IPTG. IPTG is a chemical compound that mimics allolactose and binds to the *lac* repressor and thereby free the DNA for translation of the protein.[101] The temperature and length of the cell cultivation together with IPTG concentration might be varied for an optimal yield of protein.

2.1.2. Cell lysis and biliverdin incorporation

The first step in protein purification is cell lysis that involves disruption of the cell wall, either by chemicals such as lysozyme or more traditionally by mechanical force by either high pressure or by high frequency sound waves. The mechanical pressure methods conducted by french press or emulsiflex makes the cell go through a very small passage and the cell breaks open. The high operating pressures, however, result in a rise in temperature. Hence, the pressure cells are cooled (4°C) prior to use. The cell lysis is followed by centrifugation to remove unbroken cells and cell debris. Since the phytochrome studied is expressed as an apoprotein without its chromophore, it is added in excess to the supernatant after centrifugation and covalently binds to the protein spontaneously.

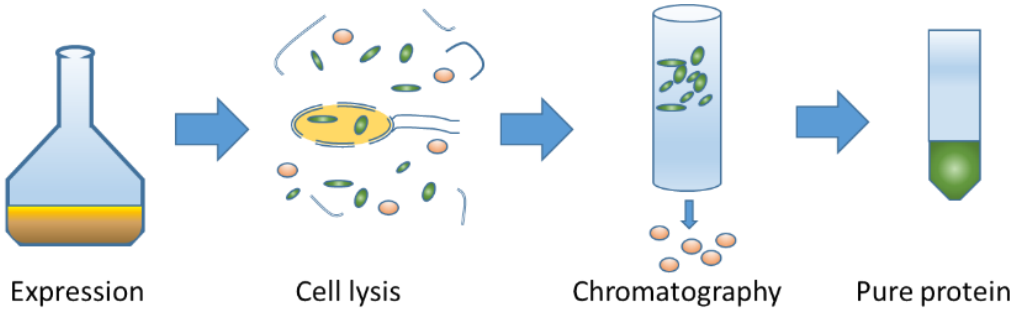

Figure 2.1. Schematic flowchart of general protein purification. The protein of interest is expressed by *E.coli* cells in shaking flasks. The cells are lysed to extract the protein into solution. The proteins are separated by different types of chromatography and the protein of interest can be obtained as a homogenous pure sample for biochemical investigations

2.1.3. Chromatography

Proteins in solution are generally purified using different types of chromatography. Proteins are separated due to their specific physical properties such as size, charge or binding affinity. However, the general principle for all types of chromatography is the same. The separation is based on having two phases, a stationary and a mobile. For protein purification column chromatography is used, where the stationary phase is a resin packed in a column, and the mobile phase is a buffer that passes the column. The proteins in the mobile phase interact with the stationary phase in different ways depending on their specific properties. Due to these interactions they are retained in the column to different extent. The separated protein is eluted with the mobile phase in fractions.[102] With the use of recombinant protein expression comes also the ability to genetically modify the protein sequence by adding an affinity tag to the protein to facilitate protein purification. Most common is the addition of a polyhistidine-tag (usually 6-10 residues) in either the N-terminal or C-terminal of the protein. The his-tag has a high affinity for nickel ions that are bound to the resin of a Ni-affinity column. The protein interacts strongly with the stationary phase and is eluted with a gradient of increasing imidazole concentration, which share the same molecular structure as histidine.[103] The final step in purification is often size exclusion chromatography (SEC), which separates the proteins according to size. The smaller proteins enter the tiny pores in the column material deeper and therefore have a longer retention time compared to larger molecules. Gel filtration removes salt and aggregated protein. It can be used to get an indication about sample purity and that the protein is in its right quaternary arrangement (e.g. monomeric/dimeric etc.).

2.1.4. Protein characterization and functional validation

A major advantage of working with colored proteins is the easy way to follow that the purification process follows the correct route. On the other hand, a disadvantage of working with light sensitive proteins is that one has to work in the dark to minimize that the protein adopts different states (e.g. Pr and Pfr) that leads to heterogeneous sample in for example crystallization.

To confirm that the protein of interest is the one purified and the purity is good enough one often performs a gel electrophoresis (e.g. SDS-PAGE.) The protein loading buffer denatures and reduces the proteins disulfide bonds and gives it a negative net charge that is almost proportional to the molecular mass. An electrical field is applied and the proteins go through the gel at velocities dependent on molecular weight. Small proteins wander further than large proteins. One band on the gel means that, there is only one protein in the sample, hence it has a high purity. The molecular weight of a protein band can be determined through comparison with a so called ladder. The ladder is loaded onto the same gel and contains a mixture of proteins of known molecular weights.

The protein concentration is measured by the absorption at 280 nm where aromatic amino acids absorb light. Every protein has to have its individual extinction coefficient, ε, determined and then the concentration, c, can be calculated by Beer-Lamberts law.

$$A_{280} = \varepsilon c l$$

Where l is the path length of the light passing the sample.

Since phytochromes works as photo switches, even more information can be obtained from its spectroscopic properties. To assert the functionality of the protein the UV-VIS spectra can be determined for the two states by switching between them with LED lasers. Also the incorporation level of the BV can be examined by comparing the 700/280 absorption ratio.

2.2. Protein crystallization

A single protein molecule is normally in the range of 30-200 Å (3-20 nm) and is too small to be visualized by an optical microscope that usually has a resolution limit of around 200nm. For visualization of smaller objects the need of radiation with shorter wavelength arises. X-rays have the right wavelength (0.01-10 nm) to interact with the electrons in the molecule and has been used to determine the three dimensional structure of various types of molecules since their discovery by Wilhelm Conrad Röntgen in 1985.[104] Although X-rays have the right wavelength they cannot be used like visible light in a microscope since they cannot be focused by lenses and mirrors to recreate a visible image of the studied sample. Instead the method relies in the crystal properties that make the individual molecules arrange in a highly repetitive manner in the crystal. This means that the X-rays interacting with the electrons in are diffracted in the same way from all molecules and create a constructive interference that can be detected as a diffraction pattern. This put high demands on the crystal being homogenous and well-ordered since any misalignment of the molecules would cancel out the interference. This reflects the requirement of highly homogenous sample with high purity to build up well ordered crystals.

Crystallization involves a phase change from liquid to solid in an ordered manner, unlike precipitation that is unordered. Hence, crystals are solids that are built up of

atoms, molecules or ions, in a repeatable ordered manner extending in three dimensions. They have an inherent symmetry and can adopt different crystal systems and lattices. Crystals consists of unit cells, which are defined as the smallest unit from which the whole crystal can be built up through translation. The unit cell can be further divided into the so called asymmetric unit, which is the smallest part of a crystal that can be rotated or translated by symmetry operations to build up the unit cell.[105]

2.2.1. Protein crystal growth

The optimal crystallization conditions are specific for individual proteins and several conditions can yield crystals of variable quality with different interactions (crystal contacts). The screening for crystal conditions and crystal optimization can be highly iterative and both time and sample consuming. There is never a guarantee for successful crystallization. High conformational flexibility or lack of sufficient crystallization contacts of the protein can lead to failure in finding crystallization conditions at all. Even if crystals appear, they might need endless rounds of optimization where conditions are slightly altered, results evaluated, and conditions further optimized.

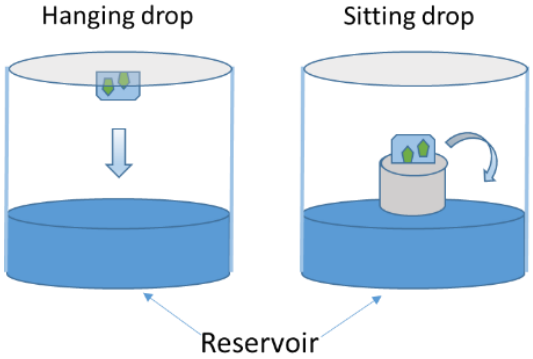

Figure 2.2. Crystallization set ups for vapor diffusion crystallization. The drop can be placed either hanging on the sealing glass (left) or sitting on an area separated from the reservoir (right).

2.2.2. Protein crystallization by vapor diffusion

Crystallization trials are often set up in drops in the microliter (µl) size. In the drop the protein is mixed with a precipitant solution containing a precipitating agent. Usually crystallization is performed by vapor diffusion meaning that the drop is enclosed in a sealed environment together with an excess of so called reservoir solution. The drop can be placed in for example hanging or sitting position (Figure 2.2). The sitting drop can allow for greater drops whereas hanging drops are sample saving and facilitate crystal harvest. The reservoir solution contains the same precipitant as the drop but in a higher concentration. This leads to a vapor diffusion (e.g. water migration) from the drop to the reservoir until an equilibrium within the sealed area is reached.[106]. Since crystallization is dependent upon the oversaturation of molecules in the solution, the solubility diagram of the drop conditions is determinant in the crystallization process (Figure 2.3).

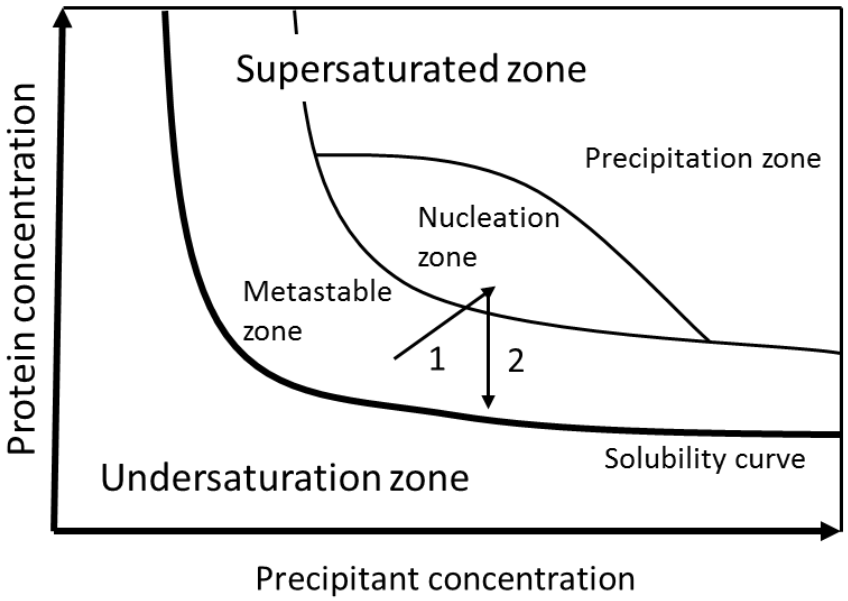

Figure 2.3 Solubility phase diagram for crystallization. By varying concentrations of protein and precipitant agent in the crystallization drop the aim is to reach the supersaturated zone. The protein and the precipitant concentration increase as an effect of water evaporation from the drop (arrow 1). When the nucleation zone is reached crystals starts to form and the concentration of free protein in the drop decreases (arrow 2).

The evaporation of water from the drop lead to that less water molecules are available for hydration of the proteins and their solubility is reduced. If the protein and/or precipitant concentration is too high it will reach the precipitant zone. In that case protein form an amorphous precipitate instead of a well ordered crystal. On the other hand, if the concentrations are too low, the drop will be in the undersaturation zone and the protein will remain in solution. Ideally the starting conditions in the drop lies somewhere in the metastable zone. When water content is reduced both protein and precipitant concentration in the drop is increased according to arrow 1 in Figure 2.3. When the protein reach the nucleation zone, small nuclei are starting to form and protein crystals starts to build up. When crystals grow the protein concentration drops according to arrow 2 in Figure 2.3, until it reaches the solubility curve and crystal growth ceases.[107]

The growth of crystals is not only dependent on protein and precipitant concentrations. Other factors such as type of precipitant, salt, pH, buffer, temperature and of course protein purity are decisive for optimal crystal growth.[106] Since the initial test for crystallization often follow a trial and error approach, they are often performed with crystallization robots and commercially available screens containing chemicals shown to yield protein crystals in many cases. Once a promising condition has been found, the condition can be optimized by for example varying precipitant concentration, pH or adding additive molecules to enhance crystal formation. The crystal growth may also vary in time from a couple of minutes to months and should be monitored regularly under a microscope. The crystal quality is examined by

diffraction screening with X-rays and poor diffracting crystals needs to be optimized to be suitable for structure determination.

2.3. X-ray diffraction

2.3.1. X-ray diffraction theory

When a crystal is illuminated by X-ray radiation the electron cloud around the atoms scatter the X-rays in all directions. Most of the scattered waves cancel out each other but some are reinforced by constructive interference creating a diffraction peak pattern. Bragg's Law describes the conditions essential for constructive interference

$$n\lambda = 2d \sin\theta$$

Where, n is an integer describing the order of diffraction, λ is the wavelength of the X-rays, d is the distance between the planes in the crystal lattice and θ is the angle between the incident X-ray beam and the scattering planes. Braggs Law is illustrated in Figure 2.4.

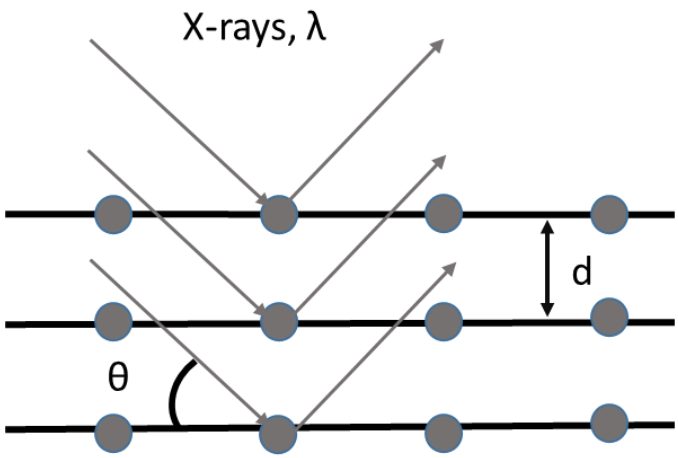

Figure 2. 4. Illustration of Bragg's law. The X-rays with wavelength λ, are scattered by the atoms (represented dots) arranged in lattice plans separated by a distance d. The angle θ is the angle between the incoming X-ray beam and the lattice plane.

Reflections only depend on the crystal lattice and only a few will fulfill the conditions. A Bragg peak (h,k,l) corresponds to diffraction from a set of crystal planes that are defined by the Miller indices hkl. The diffraction from a crystal can be described as a reciprocal lattice that represents the Fourier transform of the crystals real lattice (Bravais lattice). This relation between reciprocal space and the diffraction criterion can be explained with the Ewald sphere construction (Figure 2.5.) The reciprocal lattice points are values where the Bragg diffraction condition is fulfilled and the scattering vector must be equal to a reciprocal lattice vector for diffraction to occur.

In Figure 2.5. the incoming vector on the crystal is K_i, with the length $2\pi/\lambda$ where λ is the X-ray wavelength. K_d is the scattering vector. The red vector in Figure 2.5 represent the vector, ΔK which is the difference between K_i and K_d. For elastic scattering, (kinetic

energy of the photons are conserved, but direction is altered) the length of the K_d is equal to K_i and thereby the scattering vector lies on a sphere with the radius $2\pi/\lambda$.

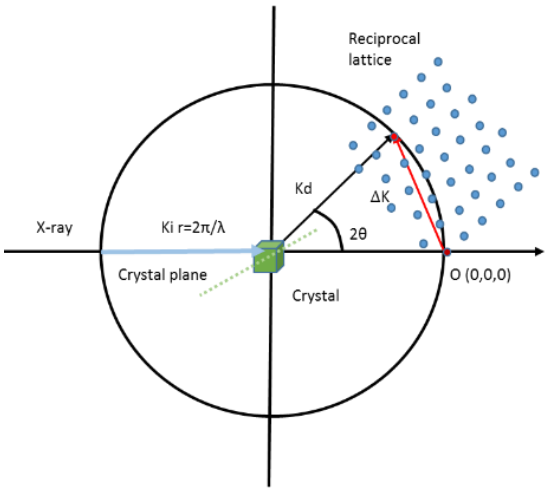

Figure 2.5. The Ewald sphere construction. The reciprocal lattice is rotated when the crystal is rotated in the X-ray beam. When the reciprocal lattice points touch the Ewald sphere the conditions for a Bragg peak is fulfilled and a diffraction spot can be recorded. See text for more details

The reciprocal lattice points are the values of momentum transfer where the Bragg diffraction condition is satisfied and for diffraction to occur the scattering vector must be equal to a reciprocal lattice vector. Geometrically this means that if the origin (O) of the reciprocal lattice is set at the intersection of the sphere and the beam that has passed through the crystal as in Figure 2.5. Diffraction will occur at lattice points on the surface of the Ewald sphere. When the crystal is rotated in the X-ray beam the reciprocal lattice also rotates and other lattice points will lie on the surface and hence show diffraction. The obtained diffraction pattern represent the reciprocal space and from this the crystal lattice can be calculated. The intensities of the Bragg peaks contain information about the unit cell content i.e. the electron density, which in turn provides information about atomic arrangement.

The electron density ρ(x,y,z) at position x,y,z in the unit cell can be obtained by applying Fourier transform on the structure factors according to the equation:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)}$$

Where V is the volume of the unit cell and F_{hkl} is the absolute value of the structure factor for the measured hkl reflection. The structure factor F(hkl) contains information of both the amplitude and the phase of the wave but experimentally only the diffraction spot intensities ($|Fhkl|^2$) can be determined whereas the phases (iα(hkl)) of the scattering waves are lost. In crystallography this is referred to as the phase problem.

2.3.2. Solving the protein structure and structure validation

The obtained crystallographic data set from synchrotrons contain the diffraction patterns with the Bragg peaks needed for solving the structure. A few images are used to determine the unit cell dimensions and the space group, which is then used to index all the reflections in the data set. These steps are performed by computer programs such as XDS or MOSFLM. [108] Following this the data needs to be scaled and merged together to a single file containing the structure factors obtained from the Bragg peaks positions and intensities. This is performed in a crystallographic software like ccp4 that contain the necessary scripts (SCALA) to treat the data.[109] The resulting file contains the structure factors but lacks the phases.

To regain the phases and obtain an electron density map several methods have been developed. It can be done either by experimental techniques, such as heavy atom incorporation or anomalous diffraction, or by so called molecular replacement which uses the phases from a previously determined structure similar to the protein of interest.

Molecular replacement is by far the most common method since it is fast and easy whereas experimental methods are more complex and time consuming to perform. For molecular replacement to be applicable the protein from which the phases are calculated must have a sequence homology of at least 30%. The molecular replacement is performed by computational programs such as Phaser[110] also included in the ccp4 package. When using molecular replacement one must bear in mind that the use of phases from another protein incorporate some bias into the electron density. To evaluate this and create a biased free map an omit map can be constructed. Parts of the model are omitted and a map is calculated. A good quality map should show the full model, including the parts that were omitted. Computational programs such as Phenix[111] can calculate a number of omits map and combine these to a so called composite omit map, which cover the full area of interest without biased phases.

The calculated phases needs to be refined further to better fit the experimental data. By providing a sequence of the protein a model of the protein structure can be built into the electron density in a program like coot.[112] The model is manually altered to fit the electrons density in real space. This is followed by automated refinement against the reflection intensities using computer programs such as REFMAC.[113] This cycle is repeated and need to be validated to not impose bias in the model building. The agreement between the model and the experimental electron density is measured by a factor called R, calculated by the equation:

$$R = \frac{\sum_{hkl}\left||F_{obs}(hkl)| - k|F_{calc}(hkl)|\right|}{\sum_{hkl}|F_{obs}(hkl)|}$$

If the model would agree perfectly with the electron density it would mean that the R factor would equal zero. Hence, a low R-value is considered as a good validation of the

built model tor represent the actual data. The R- value is often given as percent and a value of about 10 times the resolution is considered as acceptable. Since part of the refinement is performed manually it can introduce some bias to the built model. To avoid this a cross validation of the refinement can be performed by excluding 5-10% of the reflections and used for calculation of the so called R_{Free} value.[114,115] The refined R value should not differ substantially from the R_{Free} value since it would mean that the data is over interpreted and the model is not a valid representation of the data.

Another way to validate the model building is to generate a Ramachandran plot in the end of refinement process. The Ramachandran plot is a 2D representation of the ψ and ϕ angles of the protein backbone on each axis. Amino acids can adopt conformations that cluster in specific regions of the Ramachandran plot as a consequence of sterical hindrance. If amino acids exhibits angles not allowed they are considered as outliers and have very low probability to exist, indicating errors in the structural model.

The ultimate measure of the information content in a structural model is the resolution. The resolution is defined as the distance were two close objects can be distinguished as separated objects rather than a single blurry object. It is desirable to attain as good resolution as possible hence it will lead to a more extensive structural model. Although resolution cannot be improved through structure refinement, but is instead an inherent feature of the crystal packing. A more well-ordered crystal will show greater interference and therefore yield better resolution of the electron density map and a more informative model can be constructed.

2.4. From crystals in drops to data collection

2.4.1. Crystal fishing and cryo-protection

Once crystals are formed in a drop they need to be harvested and stored until data collection can be performed. Crystals left in a drop can be dried out over time and hence have reduced quality. Furthermore, the crystals need to be transported from the laboratory to the X-ray source (e.g. synchrotron), mounted on a goniometer head and aligned to the X-ray beam. The crystals are harvested from the drop by so called “fishing”. This is performed by putting a loop (often glass wool or synthetic fiber) in the drop and picking up the crystal. The crystal in the loop is then protected from dehydration by a thin layer of solvent. The size of the loop is important to create a suitable surface tension to keep the crystal within the loop.

When crystals are exposed to X-rays they are likely to experience radiation damage since they are quite fragile biological molecules. Radiation damage cause the formation of free radicals due to electron ejection from the atom when an X-ray photon is absorbed. The free radicals can alter the chemical properties of the molecules and cause damage to the crystal which leads to decreased diffraction quality. One way to decrease radiation damage is to cool down the crystal to cryogenic temperatures (-196°C). It prolongs the crystal lifetime by slowing down the movement

of the free radicals. A cryogenically cooled crystal can survive X-rays up to 30 MGy and enables data collection from one single crystal instead of several crystals that would be necessary for room temperature data collection.[11]

The decrease in temperature should in theory, increase the molecular order and thereby improve diffraction of the crystal. However, the formation of ice crystals and the expansion of water might damage the crystal. To avoid this, one has to perform the freezing in a fast manner and/or use cryoprotectant that hinder ice formation. The cryoprotectant can be glycerol, glucose, polyethylene glycols (PEGs) or other small organic molecules. If the crystal was grown in absence of a cryoprotectant, the crystal is often soaked in reservoir solution containing cryoprotectant when fished. If the crystal does not withstand the new condition directly this can be done stepwise with an increased cryoprotectant concentration. In cryo-crystallography the data collection is performed by placing the crystal loop on a goniometer head where is it maintained at cryogenic temperatures by applying a stream of liquid nitrogen onto it.

2.4.2. Synchrotrons and data collection

Data collection at synchrotrons is well established at beamlines equipped for these types of experiments. They provide the required software and sample exchanging robots to facilitate data collection. Synchrotrons provide X-rays with wavelength around 1 Å by the acceleration of electrons. They consist of big (1000s of meters) rings where electrons are accelerated and bent by magnetic fields (Figure 2.6). Initially the electrons are accelerated in a linear part and boosted to higher speed in a booster ring before they go out to the storage ring. In the storage ring they are circulated at almost the speed of light. The acceleration together with the bending of the electrons pathway makes the electrons emit high energy X-rays waves that are focused into beams and discharged out into beamlines. The use of magnets called undulator and wigglers has enhanced the X-ray flux and enabled the development of so called third generation synchrotrons which provide X-rays with high energy.[17]

During data collection the crystal is mounted on a goniometer head and rotated while being exposed to the X-rays and the diffraction is recorded on a detector. The detector is often a charged-coupled device (CCD) that works as a photon counter. The rotation oscillation of the crystal is often small (0.2-1°) and depending on crystal symmetry and radiation damage endurance, the rotation needed for a complete dataset varies. Often the software available at beamlines enables a fast indexing of the data and can almost immediately estimate the completeness and resolution, as well as cell parameters for the data set collected.

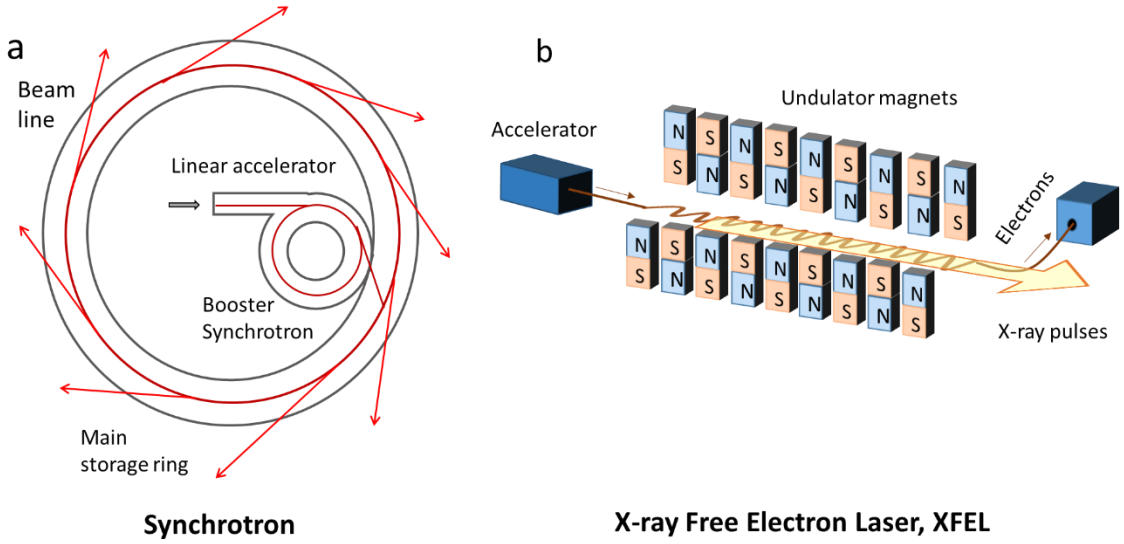

Figure 2.6 Schematic representation of two X-ray sources used for X-ray crystallography a) The representation of a synchrotron X-ray source. The pathway of electrons is shown in red. The electrons are accelerated to high speeds and circulated in a storage ring. The electrons pathway is bent by bending magnets and emits X-rays that are collected in beamlines. b) In an XFEL the electrons are accelerated in a linear tunnel and passed through undulator magnets with switched magnetic fields that make the electrons wiggle. The bending of their pathway make the electrons emit X-rays and the spacing of the magnets makes the X-rays bunch up into coherent pulses with high energy.

2.5. Serial femtosecond crystallography, SFX

Serial femtosecond crystallography is an emerging field and is more or less anticipated to revolutionize structural biology in the near future. It overcomes some of the biggest hurdles in conventional crystallography, in particular the radiation damage aspect.[16,17] However, the rise of a new method requires development and improvement of an extensive number of experimental features such as detectors, sample preparation and delivery methods, data analysis and validation etc. The principle of structure determination by SFX is in general similar to conventional crystallography but the differences are discussed in the following section.

2.5.1. X-ray free electron lasers, XFELs

The limitation of third generation synchrotrons for structural studies is maybe primarily, but not exclusively, the radiation damage. Additional aspects such as intensity limitations and minimal pulse duration also restricts time resolved applications. XFELs have emerged as even more powerful X-ray sources than synchrotrons. The first operating hard XFEL was the Linac Coherent Light Source, LCLS in Stanford, USA which has been running since 2009 and is currently to be updated to LCLS-II. Since 2011, SACLA in Harima, Japan has been operational. In 2017 PAL-XFEL in Pohang, South Korea and European XFEL in Hamburg have commenced operation. SwissFEL in Switzerland is planned to open this year. These sources can produce coherent ultrashort (fs timescale) X-ray pulses with energies up to 25keV and a peak brilliance that are around ten orders of magnitude stronger than a third-generation synchrotron.

Unlike the synchrotrons that are constructed as big rings, the XFELs are very long (km range) linear tunnels. The tunnel possess an electron source and a linear accelerator where the electrons are accelerated up to speeds close to the speed of light (Figure 2.6). Then the electrons interacts with the undulator that consists of magnets spaced to produce X-rays wavelengths (0.01 to 10 nm)[116] When the electron pathway is curved, as a result of the interaction with the undulator they emit the X-rays. As a consequence of the long tunnel length and spacing of the undulator the X-rays are bunched together, thus becoming more coherent. This produce highly coherent X-ray pulses with an ultrashort pulse length (femtosecond scale). It is the coherence of the pulses that produces the high intensity X-rays compared to synchrotrons, where the emission is somewhat cancelled out (incoherent) due to out of phase generated magnetic fields.[117]

An early application for XFEL radiation was the development of serial femtosecond crystallography, SFX, since the wavelength is in the range for atomic resolution of structural determination of biomolecules.[15]

2.5.2. Diffraction before destruction

If the high intensity X-rays produced by synchrotrons can limit structural studies by radiation damage of the biomolecules, how can even higher intensity X-rays be the solution to the very same problem? The answer lies in the ultrashort pulses[118] and the 'diffraction before destruction' principle. The principle of data collection by the short XFEL pulses was proposed to work already in 2000 as demonstrated by the use of molecular dynamic.[119] The conclusion was that diffraction data could be recorded by the ultrashort pulses before the sample exploded by the interaction with the high intensity X-rays. The principle was proved to be correct when the first data collection was performed on crystal of the big protein complex photosystem I in 2011. [15]

2.5.3. SFX experimental set up and data collection

Since the crystals are destroyed by the interaction with the high intensity XFEL pulses, the method requires constant replenish of the sample and is based on the 'one crystal, one shot,' approach. The method thereby require high amounts of nano- to micrometer crystals streaming through the XFEL beam (Figure 2.7). The XFEL- beam is focused by mirrors and randomly oriented fully hydrated microcrystals are injected to interact with it. A crystal hit by a XFEL pulse result in a diffraction pattern recorded by the detector. The detector needs to have a read out that matches the repetition rate of the XFEL and the beamline operates under vacuum to reduce background scattering. At the LCLS a Cornell SLAC Pixel Array detector (CSPAD) is used which can record data with a sufficiently high repetition rate.[120,121]

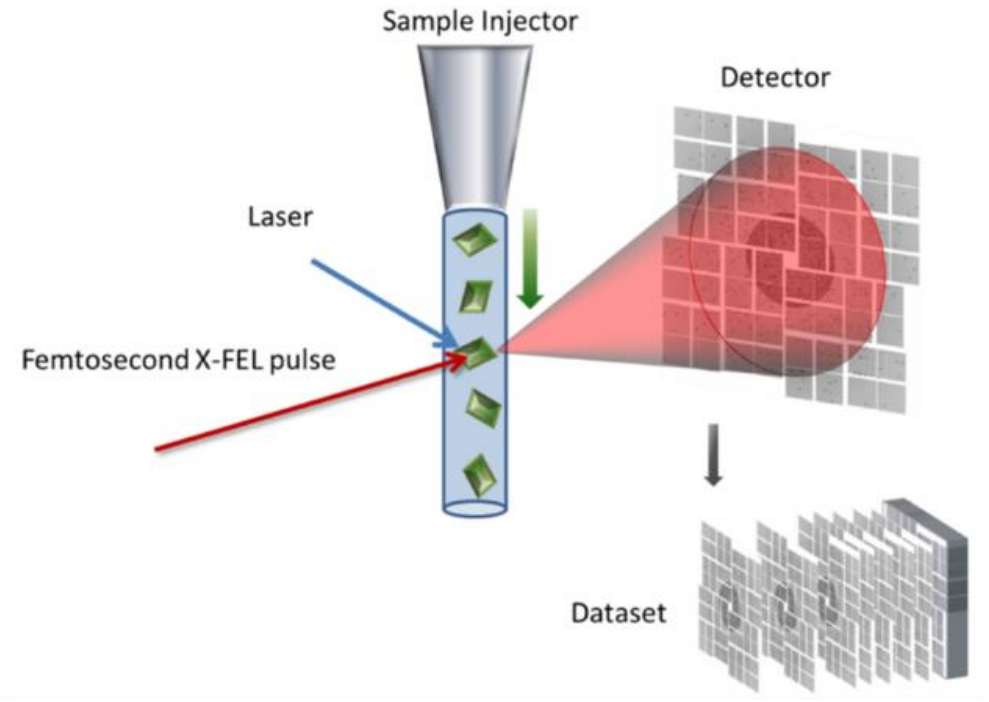

Figure 2.7. Serial femtosecond crystallography experimental set up. The microcrystal sample is injected into the X-FEL beam and yield diffraction patterns from individual random oriented crystals. The diffraction patterns are merged together into a dataset used for structure solving. The interaction point between crystals and XFEL beam can be aligned to a laser to activate light triggered proteins for time resolved structural investigations.

The use of thousands of microcrystals instead of one (or a few) big crystals leads to both advantages and drawbacks compared to conventional crystallography. It can be an advantage, due to the fact that some proteins that are reluctant to form big well diffracting crystals and only forms microcrystals. These can now be structurally determined. It also enables time-resolved experiments to be conducted in a 'pump-probe' manner, which are of particularly high interest for irreversible reactions. The time-resolved experiments can be performed for light triggered reactions by aligning a laser to the interaction point between crystal and XFEL to activate the reaction of interest. The drawbacks with SFX, can be the high sample consumption. Another problem is that the crystals are not identical which can lead to hardships in the data processing.

Figure 2.7 show the general set up described but it all depends on the sample delivery that can differ and is further discussed under the sample delivery section (2.6.4.). Only 10-20% of the recorded images contain diffractions patterns. Furthermore, the diffraction patterns from randomly oriented crystals contain only a few peaks on each image. This leads extremely large data sets and put new demands on data processing and handling which is discussed under section 2.6.7.

Probably the most important aspect of an SFX experiment is the requirement for large amounts of well diffracting microcrystals and the production of these are discussed in section 2.6.5. and in **Paper II**.

2.5.4. SFX sample delivery

SFX experiments are highly dependent on a stable injection system of the sample to the XFEL-beam. The majority of the SFX data has been collected from a stream of microcrystals in their mother liquor (protein and reservoir solution were crystals are grown) that allows constant replenish of fresh sample. It put high demands on controlling the velocity of the injection and the stability of the stream that has to be properly aligned with the XFEL. Several strategies has been developed, but the first and the most common delivery method is by the so called Gas-Dynamic virtual nozzle, GDVN.[122] A smaller capillary (around 50µm) is mounted into a larger capillary and the crystal solution is delivered in the inner capillary, while high pressured gas flows in the larger capillary and focuses the stream of sample into a thin elongated jet (few µm). The minimal flowrate of the jet is about 10 µl/min. One problem with this set up is that the thin jet can be easily clogged, in particular, if the crystals have a large size distribution or tend to aggregate. This is usually avoided by filtering the sample prior to injection and by the use of inline filter in the nozzle.

Since the crystal delivery when using a GDVN is much faster than the XFEL repetition rate the majority of the crystals never interacts with the XFEL. It is not possible to recollect sample since it is injected into vacuum. This, combined with filtering requirements, leads to extremely high sample consumption and might require grams of purified protein. Modifications, to reduce the sample consumption by the GDVN[123] have been investigated but also alternative methods to inject sample have been developed.

The lipidic cubic phase (LCP)/high viscosity injector that can inject crystals grown in viscous fatty media. It can also be used for crystals grown in other media and mixed with grease post growth. The high viscosity injector consists of a plunger that is driven by a HPLC pump to extrude high viscosity liquid that is focused by a co-flowing gas to a stable jet.[124] The injector has been proved to work in both air and vacuum[125] and many types of high viscosity media can be used such as grease matrix, agarose[126] and Vaseline.[127] The high viscosity injector has a substantially reduced flow compared to GDVN.[128] One problem can be that crystals not grown in the fatty media dissolves in it upon mixing. Other sample delivery methods includes fixed targets[129,130], drop on demand[56] and MESH/ coMESH.[131,132]

The hurdle to get the crystallization conditions and the optimal delivery requirements to coincide put additional demands on crystallization that might not always be optimal. The development of improved or new methods for sample delivery is a needed ongoing work.

2.5.5. Microcrystallization

The success of SFX experiments is relaying on the production of well diffracting microcrystals and the area is still unexplored to a great extent. The traditional demand for X-ray crystallography is to produce large, well ordered crystals. Followed by individually screening to find the ‘one’ best suitable for collecting a full data set and

determine the protein structure. In this view, the yield of small microcrystals in crystallization has been judged as something useless that needs optimization. With SFX, the requirements has been altered to instead desire microcrystals homogenous in the nano-micrometer size in high density. This requires new crystallization strategies and ways to evaluate crystal quality.

One advantage of the use of microcrystals is that they exhibit lower mosaicity, since large crystals often suffer from long range disorder and anisotropic resolution. When the SFX method emerged, microcrystallization strategies started to form and the first study of crystals produced especially for this type of experiment was of the photosystem I. The microcrystals used were obtained by the 'ultracentrifugation method'.[15] The protein solution is entering the nucleation zone by slowly concentrating the protein at low ionic strength. But this method is limited to systems that do not include too large molecules (such as PEGs or detergents) that cannot pass the ultrafiltration membranes.

For development of microcrystallization strategies the proteins solubility phase diagram is still of great importance. In conventional crystallography, as discussed previously, large crystals are formed by increasing precipitant (and protein) concentration gradually to touch the nucleation zone where few nuclei are able to form and the protein can add to these and form few large crystals. In microcrystallization it is instead desired to start off in the nucleation zone to form a great number of nucleation sites. This is typically done by increasing either the precipitation agent concentration, protein concentration or both (Figure 2.8.). It leads to a fast decrease in protein concentration when the nuclei forms and grow into many small crystals.

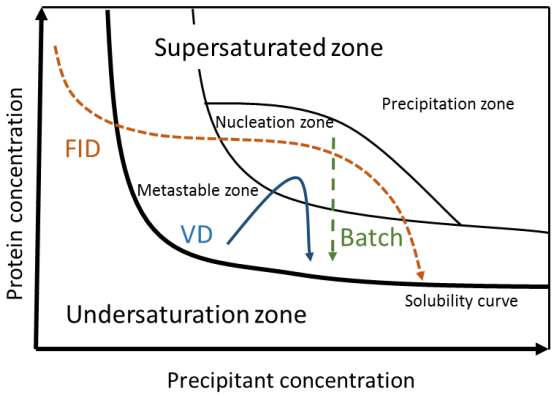

Figure 2.8. Solubility phase diagram with methods applied for microcrystallization purposes. In contrast to vapor diffusion (VD blue) the formation of microcrystals is aimed to start in the nucleation zone directly by alteration of protein and/or precipitant concentration. Batch crystallization (green) is performed in a sealed environment and crystal formation leads to a decrease in protein concentration. Free interface diffusion (FID) (orange) starts off with a low precipitant concentration that is increased to reach the nucleation zone and crystallization growth is ceased with further increased precipitant concentration.

Batch crystallization is a direct method to instantly reach the nucleation zone and is also applied in conventional crystallography. The protein and the reservoir solution is mixed to homogeneity and crystal nuclei are supposed to form instantly. They are then incubated in a sealed environment to mature into crystals. This can be combined with methods of seeding where small microcrystals (that can be large crushed crystals) are added to increase the nucleation rate.[133] The batch method is applicable in microcrystallization since it can easily be scaled up to the great volumes that is needed for SFX. Also crystals produced by vapor diffusion can be used for SFX although it is time and work consuming to collect crystal containing drops in volumes needed for SFX.[133,134]

Another method for producing microcrystals in solution is the free interface diffusion (FID) method. FID is also a method borrowed from conventional crystallography. It is based on that the protein is placed on top of the precipitant solution (which usually has a higher density). This creates a concentration gradient at the mixing area leading to crystal growth of crystals with increasing size along the gradient. In conventional crystallography capillaries are used but this is not suitable for the scaled up volumes of SFX. Since the crystal growth occur only where the two solutions mix the amount of crystals obtained is limited. The method has therefore been modified to better suit SFX applications. It can be either through modifying the mixing of the solutions or by adding centrifugation that makes the nanocrystals sediment into the precipitant when reaching a certain size. This method yields greater amounts of crystals with a narrow size distribution.[135] However the struggle to find the right precipitant, pH, temperature and other crystallization parameters still remains in attempts to obtain microcrystals. If the conditions for obtaining large crystals of the protein of interest is known, it might be a good starting point.

Membrane proteins, are desired to crystallize since they are often responsible for cell signaling and can be potential drug targets. Unfortunately, they are difficult to crystallize due to their mixed hydrophobic and hydrophilic nature. Often they are not stable in aqueous solutions have instead successfully been crystallized in LCP.[136,137] The protein is mixed with monolein and together they spontaneously form the LCP. The LCP acts as a membrane mimicking system where the protein can migrate and form crystals. The drawback is that the crystal solution cannot be concentrated nor be filtered before injection which decreases the maximum hit rate and increases the risk of clogging the jet. Several successful SFX experiments have been conducted with LCP as a crystallization and injection media.[128,137,138] There are also some cases where crystallization has been executed in vivo, and crystals are formed spontaneously in cells (insect cells).[139]

2.5.6. Validation of microcrystal quality

The traditional way to monitor and validate crystal quality is by optical microscopy, followed by X-ray diffraction screening. Although resolution and other information about the crystals are not visible by eye, it often provide some kind of estimation about

packing and diffraction properties. Since the microcrystals are very small (nano- few micrometer range) they are approaching the resolution limit for optical microscopy. Small crystals are difficult to distinguish from precipitate and are problematic to fish. Furthermore small crystals are more sensitive to radiation damage at synchrotrons. These aspects leads to the demand for other validation techniques. Some are dealing with scattering, electron microscopy or nonlinear imaging.[17]

The most common way to estimate crystal diffraction by microcrystals might be powder diffraction which can be performed at both ambient and cryo-temperatures. If large crystals are available one strategy is to perform conventional crystal screening. If the microcrystallization conditions are somewhat similar one can assume that the obtained resolution is somewhat comparable to the resolution of the microcrystals. The development of a proper method of microcrystal screening is needed to avoid problems of bad crystal quality at valuable experimental time.

2.5.7. SFX data processing and validation

As discussed earlier, the experimental set up in SFX differs substantially from conventional X-ray crystallography, and both the data acquisition and the data processing is different. This has required the development of both detectors and computational programs to index, merge and scale the data to a useful dataset for structural determination.

The CSPAD detector used at the LCLS consists of multiple modules that might differ slightly in position between experiments. Hence, the exact location of each module must be known to be able to access the data correctly. The detector geometry is determined before any experiment and saved as a geometry file used in the data processing and indexing of the Bragg peaks.[140,141] The data consist of snapshots collected in a serial manner of crystals with random orientation. This means that the diffraction patterns represents slices through the Ewald sphere in any direction and only partial reflections can be recorded. A high redundancy (above 50) is often needed to cover the full “rotation” of the reciprocal lattice. As already mentioned, the sample delivery is much faster than the repetition rate of the XFEL which leads to that only 10-20% of the detector images contains diffraction spots. The need for high redundancy/multiplicity combined with the low hit rate leads to the requirement of collections of hundreds of thousands images to obtain a full dataset. Not only can images be blank they can also contain reflections from two or more crystals, being so called ‘multi-hits’ that cannot be used in the structure determination. The great amount of data collected is often in the terabyte size and thus put new demands on the data analysis tools.

First the data is sorted, and blanks and multi-hit patterns are eliminated by a special program designed for SFX, called Cheetah.[142] Each pattern is subjected to peak finding. Meaning that Bragg peaks are identified and patterns with too few peaks are rejected. Patterns are sorted and background is corrected by programs under the CrystFEL Suite.[141] Then follows the determination of the cell parameters and indexing of the

patterns by indexamajig (also under CrystFEL). The data is merged and individual intensity measurements from single crystals are combined using a Monte Carlo based approach.[143] The resulting file contains the structure factors and the remaining issue is to regain the phases just like in conventional crystallography. The phases can be obtained with molecular replacement but if it is not possible *de Novo* phasing is required by the use of anomalous scattering methods like SAD.[144]

Problems in SFX data analysis can be inaccuracy in the structure factor amplitudes due to variations in the XFEL-pulse energy, partially recorded reflections and the inhomogeneity of the crystals. Furthermore, presence of larger crystals might need attenuation of the beam to protect the detector and there by the resolution from smaller crystals can be limited and decrease the optimal resolution.

The classical validation of crystallographic data using R_{merge} is not useful for SFX since it combines images from thousands of different crystals. Instead the use of R_{split} has emerged. The data is divided in to two sets that are merged separately and then compared according to the equation:

$$R_{split} = 2^{-1/2} \frac{\sum |I_{even} - I_{odd}|}{\frac{1}{2}\sum (I_{even} + I_{odd})}$$

Where I_{even} and I_{odd} represent intensities in of reflections in the even and odd numbered patterns respectively. The multiplication by $2^{-1/2}$ corrects for the splitting of the data.

2.5.8. Time-resolved SFX

Historically, time resolved crystallographic studies have been confined to Laue crystallography where the use of polychromatic X-rays allows shorter exposing times to collect a dataset compared to monochromatic X-rays.[145] However, the time resolution in Laue crystallography is limited to the millisecond to picosecond timescale which is to long for many fast reactions, radiation damage remains as a problem and the study of irreversible reactions is non-feasible.

The development of TR-SFX opens up for ultrafast time-resolved studies of structural dynamics in proteins and other macromolecules by a pump-probe approach. It makes radiation damage is neglectable and enable the possibility to study irreversible reactions due to the replenish of sample with each X-ray pulse.

Currently, TR-SFX is most applicable for light activated proteins where the protein crystals are activated by a laser pulse of a specific wavelength and probed by an X-ray pulse after a set time to observe the resulting structural changes in the protein. The set up for the experiment is illustrated in Figure 2.7. The laser for photoactivation of the protein is aligned to the interaction point of the sample jet and the XFEL beam. The ultrafast time resolution is limited by the pulse length of the pump laser and the X-ray probe. The major challenges lies in achieving a perfect alignment of the laser and

X-ray beam and activate a sufficient amount of the proteins in the crystals to gain a reliable difference signal. The stable precision in timing between the laser and the X-ray is of highest importance which has led to the development of different so called timing tools.[146]

Successful time resolved experiments has already been conducted with light- activated proteins such as photosystem II, Yellow fluorescent protein and bacterial reaction center.[147-149] Time resolution of TR-SFX has been confirmed to be approaching the sub picosecond regime that could reveal the cis/trans isomerization of the chromophore in PYP at a 1.6 Å resolution.[150]

Although light-triggered proteins are the most convenient to study by TR-SFX most of the proteins in nature are not activated by light. Hence methods to study reactions of other types of proteins are desired. There are for example experiments performed on the dissociation of CO from myoglobin.[151] Moreover, the fact that the SFX utilizes small crystals opens up possibilities that never would be feasible in conventional crystallography. One such experiment that has been performed at the LCLS is the 'mix and inject', where a ligand and a protein were mixed by specific equipment and injected into the beam.[152] Another possibility is to use caged compounds, included in the crystal, that are released by a laser pulse. In these methods the time resolution is limited by the diffusing time within the crystal, and the diffusion is dependent on crystal size. This can be problematic, since the activation must be homogenous in the crystal to be able to observe intermediates at a given time point.

A TR-SFX experiment involves the acquisition of data sets at different time points after reaction initiation. The setup is often to collect every second image as a light activated and the other with no activation. In this way a 'light' and a 'dark' data set are acquired. This reduces the risk of systematic errors that arise if the data sets were collected at separate times. The timing tool and a fast pump laser opens up for a broad range of time points that can be chosen, from the femtosecond to the nanosecond range. From the light and dark datasets difference maps can be created by subtracting the structure factors from the native 'dark' state from the activated 'light' state creating a F_{obs}- F_{obs} map. For this the structure of the native state is crucial and therefore the structures in **Paper II** and **Paper IV** can be seen as the first step towards time resolved studies of phytochromes. The creation of difference maps at various time points allows snapshots of intermediate states of a structural change within a protein and from these a 'molecular movie' can be constructed. This allows insight to ultrafast structural changes and understanding of the reaction mechanisms. In **chapter 5** our primarily results from time resolved investigations of the phytochrome photon absorption and signaling are presented.

Chapter 3

3. Structural Determination of the Chromophore Binding Domian from *D.Radiodurans*. Paper II and Paper III

The project started with the aim to structurally investigate phytochromes by X-ray crystallography, preferably time resolved studies by SFX at an XFEL. The first step was to obtain crystals meeting the requirements from the emerging field. Collaborators had great experience and knowledge of phytochromes and had successfully crystallized the PSM from *D.radiodurans* in both its dark and red light illuminated form.[63] However, the resolution of these crystals were to low (3.8Å) to be able to observe small structural changes such as the D-ring isomerization. Furthermore, Burgie et al. had recently published the near-atomic resolution (1.16Å) crystal structure of the *D.radiodurans* chromophore binding domain, CBD, with a mutation to alter crystallization interactions.[38] This inspired the attempts to crystallize the CBD of *Dr*BphP for SFX purposes and this chapter describes how this led to the structural determination the wild type CBD from *Dr*BphP, presented in **Paper II**. Furthermore, the same crystallization conditions could be used to obtain the crystal structure of the CBD H290T mutant presented in **Paper III.** This mutant was further investigated by vibrational spectroscopy and the results could be supported by the CBD H290T mutant crystal structure.

3.1. Cultivation and purification of the CBD from *D.radiodurans*

Freeze cultures of *E-coli* expressing *Dr*BphP CBD with a His6-affinity tag was kindly provided from collaborators (Ihalainen, University of Jyväskyla). The site directed mutagenesis that altered the histidine at position 290 to a threonine was executed by Heikki Takala. Both proteins were expressed and purified as following.

Protein was expressed as apoprotein in *E.coli* BL21 DE3 with recombinant expression induced by IPTG, and with ampicillin selection. The cells were grown in LB-media at 37° C and 230 rpm rotation until OD_{600} reached 0.6-0.7. Protein expression was induced with I mM IPTG and incubated at 28° C degrees overnight. Cells were harvested by centrifugation and lysed with emulsiflex in presence of protease inhibitor. Biliverdin was incorporated overnight in a >10x molar excess. Large particles was removed by ultracentrifugation and filtration prior to initial purification by affinity chromatography (Histrap, GE healthcare). The protein was washed with 5 mM imidazole to remove impurities and eluted over a linear imidazole gradient. Imidazole was removed by application to a desalting column. (HiPrep 26/10 GE Healthcare). The Protein was concentrated and applied to a size exclusion column (Superdex 200 16/600 or 26/600 GE healthcare). The pure protein was concentrated, using centrifugal concentration tubes, to concentrations around 30mg/ml in 20mM tris,

50mM NaCl pH 8. Protein concentration was measured on a nanodrop and concentration was calculated using the ε=32 430 M^{-1} cm^{-1} and molecular weight =37 188.5 g/mol.

All purification and crystallization was performed in darkness or under green light to avoid photo conversion of the protein. To test the photoactivity of the protein the absorption spectra was recorded in the dark state and after illumination with red light (660) nm.

3.2. Crystallization of the CBD

The first crystallization attempts pursued to mimic conditions previously described for the WT CBD.[49] The method included a recrystallization step were obtained crystals where harvested, dissolved, clarified by centrifugation and recrystallized. This can be considered as an additional purification procedure that removes proteins that does not form crystals. Wagner et al.[49] reported on crystal conditions consisting of 67 mM pH 4.95 sodium acetate, 3.3% PEG 400, 1 mM DTT for both crystallization steps. However, the attempts to reproduce the crystallization procedures led showers of small and thin rhombohedral shaped crystals in the first step and only precipitate or microcrystals in the recrystallization step (Figure 3.1).

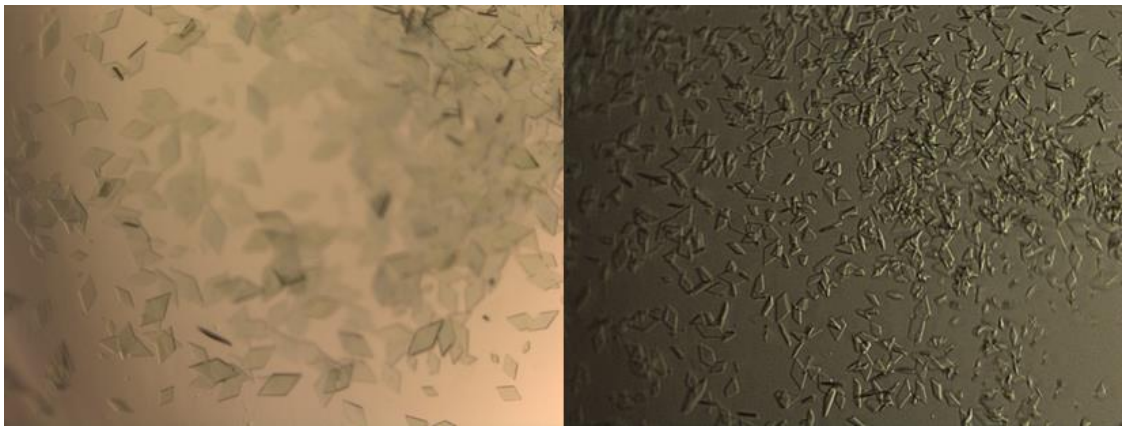

Figure 3. 1. Initial crystals of the chromophore binding domain. Photographs of Small rhombohedral shaped crystals obtained from attempts to reproduce published crystallization of the CBD (left) and even smaller crystals obtained from a recrystallization attempt (right).

Since the crystals were thin flakes they were highly fragile and complicated to fish for X-ray diffraction studies and this is even more difficult when the addition of the cryoprotectant comes into the procedure. The crystals are often fished from the drop and transferred to one, or a series of drops with increased concentration of cryoprotectant and then fished and frozen in liquid nitrogen. To avoid this an attempt to directly add the cryoprotectant (30% 2-methyl-2,4-pentanediol) to the reservoir was tested.

Surprisingly, this led to dramatic changes and well-formed crystals appeared (Figure 3.2.). Hence, the recrystallization step was unnecessary for our protein probably due to higher protein purity compared to published purification methods.[49] The crystals

used for structural determination were grown by vapor diffusion with the hanging drop method. The protein concentration was 20 mg/ml and was mixed with reservoir in a 1:1 ratio with reservoir solution (67 mM Sodium acetate pH 4.95, 3.3% PEG 400, 1 mM DTT and 30% 2-methyl-2, 4-pentanediol. The 2 µl hanging drop was equilibrated against 800 µl reservoir and crystals formed after 24-48 hours at room temperature. The same conditions could be used to yield crystals of the H290T mutant.

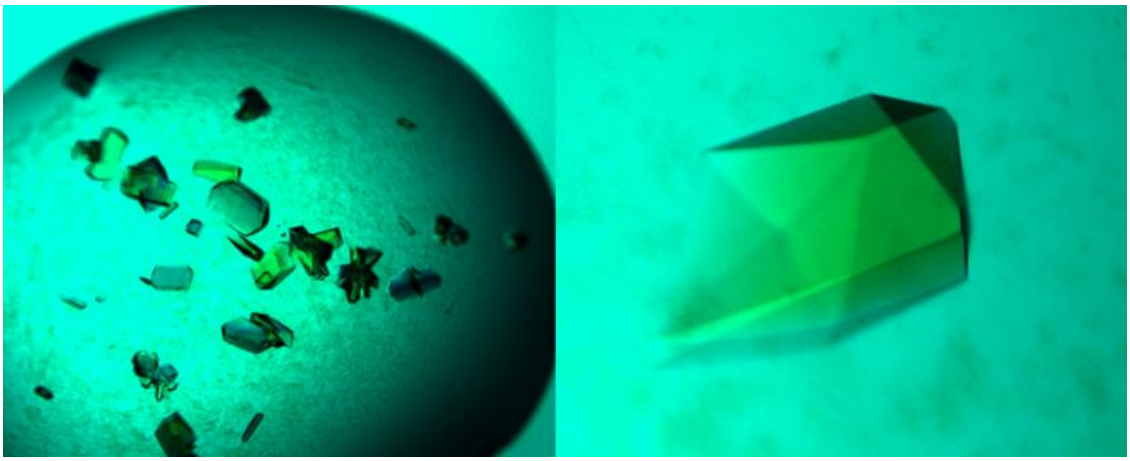

Figure 3.2. Crystals of the chromophore binding domain with altered crystallization condition. Photographs of large well-formed crystals in a drop (left) and a zoomed in crystals with clear 3D shape (right)

3.3. Data collection and structural determination of the CBD

From the obtained crystals a data set could be collected at the European Synchrotron radiation facility (ESRF) in France at beamline 1D23-I. The crystal was placed in a 100K cryostream and the X-ray wavelength was 0.980 Å. The crystal diffracted to 1.35 Å and the obtained diffraction data were processed using the XDS package.[153] The crystal condition made the protein crystallize as a monomer, with one molecule in the asymmetric unit and monoclinic space group C121 in contrast to the previous crystals that crystallized in orthorhombic space group $P2_12_12_1$.[49,51,54] The space group is the same and the cell dimensions are similar to previously a published monomeric CBD crystal structure [154] The CBD WT structure was solved by molecular replacement with Phaser[155] using the high-resolution CBD structure (PDB id 4Q0H)[54] as a search model. The refinement was performed with REFMAC [156] and the model was built in coot. [112] The data set of the CBD H290T mutant was collected at ESRF beamline ID30A-3 and the structure was solved as for the WT but the CBD WT structure (PDB id 5K5B) was used for molecular replacement. All details for data collection, structure solving and refinement can be found in table 1.

Table 1. Data collection and refinement statistics (highest resolution shell in parenthesis)

Data parameter	**CBD WT**	**CBD H290T**
PDB ID	5K5B	6FTD
Space group	C121	C121
Unit cell dimensions a, b, c	93.76, 54.28, 70.15	94.8, 53.0, 135.9
α, β, γ (°)	90.00, 92.00, 90.00	90.0, 93.3, 90.0
Resolution	47.0-1.35 (1.38-1.35)	47.3–1.40 (1.44–1.40)
I/σ	18.37(2.31)	7.48 (0.93)
CC(1/2)	99.9(76.5)	99.7(33.6)
Completeness	98.07(95.69)	99.7 (99.5)
Redundancy	4.36(4.23)	4.25 (4.26)
Refinement resolution	39.37-1.35(1.39-1.35)	47.3–1.40 (1.44-1.40)
Number of reflections	72086 (5211)	125781 (9227)
R_{work}/R_{free}	0.141/0.171 (0.26/0.29)	0.149/0.196 (0.41/0.42)
Number of atoms	2720	5741
Ramachandran, Preferred, allowed, outliers	97.8%, 2.2%, 0%	98.0%,2.0%, 0%
Average B-factor	25.47	24.0
R.m.s deviations Bond lengths(Å)	0.01	0.028
R.m.s deviations Bond angles(°)	1.506	2.55

3.4. The structure of the CBD

3.4.1. Structural features of the CBD WT

The resulting structure of the 321 residues long chromophore binding domain is globular and contains both α-helices and antiparallel β-sheet secondary structure elements. The arrangement of the PAS and the GAF domains is presented in Figure 3.3 (PAS colored pink, GAF colored purple). The PAS domain (residue 38-121) contains an antiparallel five stranded β-sheet surrounded by three α-helices. The GAF domain (residue 140-321) contains an antiparallel six stranded β-sheet, a three helix bundle and two additional helices that sandwich the chromophore together with the β-sheet. The linker region between the PAS and the GAF is a loop that is disordered and is absent in the crystal structure. The interaction between the PAS and the GAF domain is limited but includes the unusual trefoil knot. The 35 residues long N-terminal (colored magenta) passes through a GAF extension (red) that connects back to the PAS domain (pink) and forms a knot structure. One part of the N-terminal forms a miniature β-sheet structure with the GAF extension and one β-strand in the PAS domain. Although the GAF-domain sandwich the biliverdin, it is covalently attached to the N-terminal part of the PAS-domain. From the structure it is clear that the chromophore is tightly embedded within its pocket and that the D-ring has

substantially more space and freedom to move in the pocket compared to the rest of the BV.

Figure 3.3. The three dimensional structure of the WT chromophore binding domain. The N-terminal is shown in magenta, the PAS domain in pink and the GAF domain in purple. The GAF domain has a loop (red) that extrudes back to the PAS domain and forms a three foil knot together with the N terminal. The chromophore is colored black and embedded in the GAF domain.

3.4.2. Structural comparison to high resolution CBD structure

The resulting structure of the chromophore binding domain is highly similar to other published CBD structures from *D. radiodurans.*[49,51-55,154] The alignment between our CBD WT structure (raspberry in Figure 3.4.) and the 1.16Å high resolution (1.16Å) structure[54] (PDB ID: 4Q0H orange in Figure 3.4.) show very similar arrangement (RMSD 0.4 Å). The high resolution structure, contains a mutation Y307S that improved the crystal packing and is hereafter denoted as CBD Y307S.[51] The highly similar structure is to expect from proteins from the same organism.

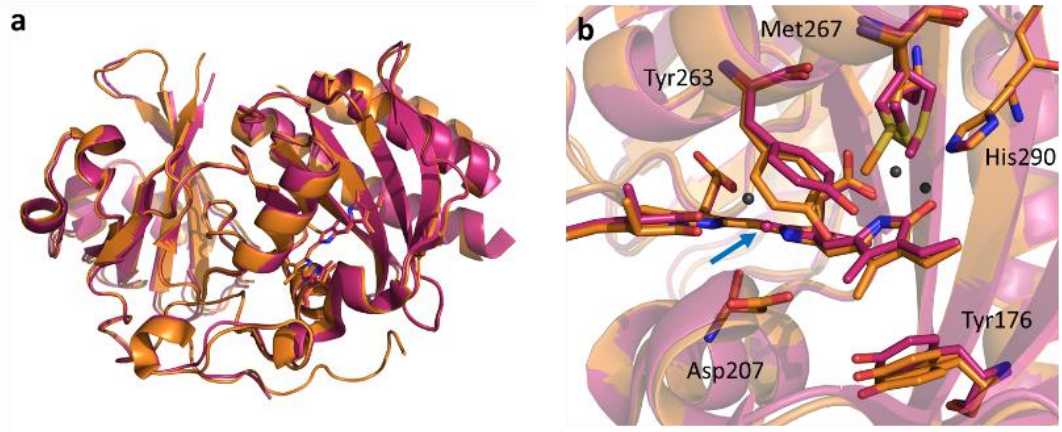

Figure 3.4. Structure comparison between the CBD WT and CBD Y307S. a) The CBD WT structure (raspberry) superimposed on CBD Y307S (4Q0H orange). b) Zoomed in view on the chromophore binding pocket and surrounding amino acids. The displacement of Tyr263 in the CBD WT structure opens up for an additional water indicated by an arrow

However, when zooming in on the chromophore and its interactions with the protein residues small differences are observable (Figure 3.4b). The D-ring of the chromophore in our structure is slightly more planar in relation to the B and C ring compared to CBD Y307S. The positon of Tyr263 is shifted allowing space for an additional water in our structure (pointed out by blue arrow). The small shift of the D-ring allows the Tyr176 to move closer to the chromophore. The Met267 shows different confirmation indicating it is either dynamic or subjected to radiation damage at its sulfur atom. Shown in Figure 3.4b is also His290 (further discussed in the following section) and Asp207 (proposed to be highly important for signal transduction) which are both unaffected.

3.5. The structure of the CBD H290T

The fact that the CBD WT and the CBD H290T crystallized equally well in the same crystallization conditions suggested that the packing and the structures would show no major differences. This was shown to be true and the structure of the CBD H290T mutant is highly similar to the WT CBD (RMSD 0.2 Å). However, the unit cell of CBD H290T was of double length in one direction and the asymmetric unit contained two monomers instead of one as in the CBD WT. The B monomer shows overall less defined electron density than monomer A, but the crystal packing interactions is basically identical. It was indicated that the B monomer is 76% identical to the A monomer but also contains 23% pseudo translational electron density from a non-crystallographic symmetry related monomer.

In Figure 3.5a a superposition of the CBD WT (raspberry) and the CBD H290T (teal) is shown and demonstrates that the structures are highly similar. Figure 3.5b show a zoomed in view of the chromophore binding pocket. The space available due to the missing histidine in CBD H290T contains electron density which we assign two highly disordered waters (teal), which are further discussed in the following sections and in

Paper III. The D-ring is slightly tilted and more planar and closer positioned towards the threonine in the H290T mutant due to altered interactions in the absence of His290. However, this does not lead to other further structural changes in the protein.

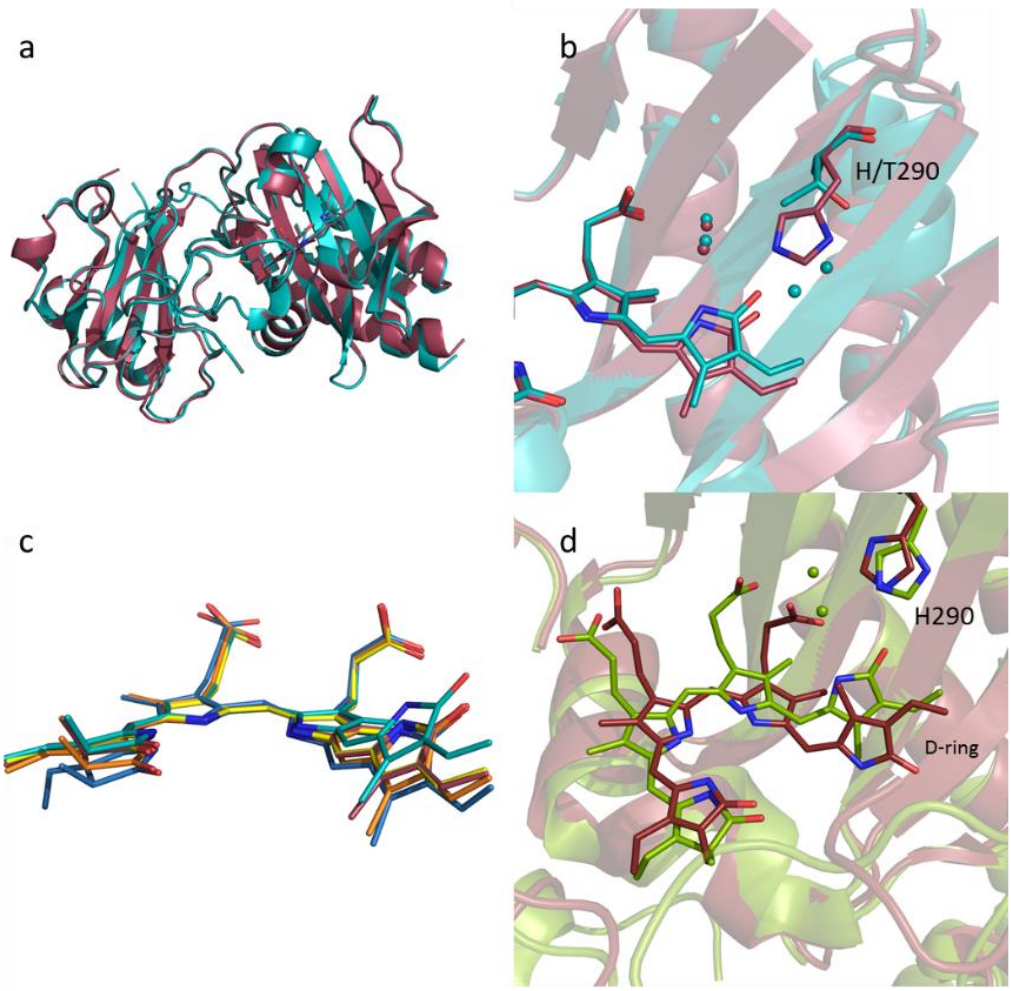

Figure 3.5. The role of the Histidine 290 and chromophore heterogeneity. a) Superimposition of the CBD WT (raspberry) and the CBD H290T structure (teal). b) Zoom in on the chromophore binding pocket. The removal of the His290 leads to a reposition of the chromophore D-ring closer to the 290 residue. In the CBD WT the D-ring carbonyl is coordinated by the histidine and one highly ordered water that forms a hydrogen bonding network with the C-ring propionate via another water. In the mutant the void is filled with two extra waters (teal spheres) that coordinates the D-ring and creates a hydrogen bonding network to the threonine. c) Chromophore configuration in BphPs structures. CBD WT shown in raspberry agrees well with a monomeric structure (yellow PDB 4IJG). The CBD Y307S structure (orange PDB 4Q0H) is positioned closer to the configuration of a PSM structure (blue PDB 4Q0J). d) The isomerization of the chromophore upon light absorption from the Pr (CBD WT structure green) configuration to the Pfr (ruby PDB 5C5K) leads to a repositioning of the chromophore. The highly ordered waters present in the Pr structure (green spheres) must be removed to make room for the reposition of the C-ring propionate and His290 towards each other.

3.5.1. Chromophore heterogeneity in phytochrome structures

In previous section 3.4.2. the CBD WT and the CBD Y307S structures were compared. In Figure 3.5c further chromophore configurations are compared to each other. Slightly different conformations of the D and A ring can be observed. The CBD WT

crystallization conditions makes the protein pack as monomers and the chromophore configuration agrees well with a structure of an engineered monomeric CBD.[157] The CBD Y307S chromophore configuration lies in between the monomeric structures and a *Dr*BphP PSM structure.[54] This suggests that the crystal packing influences orientation of the residues in the chromophore binding pocket and that the D-ring (and A-ring) is adjusting accordingly. Since the CBD WT and the CBD H290T are crystallized in the same conditions and with the same packing interactions the displacement of the D-ring is instead a result of the mutation.

Heterogeneity of the chromophore binding pocket in the Pr state has also been identified for phytochromes in solution. Two Pr-forms were identified in the PSM from Cph1 by NMR.[41]. Furthermore pH dependent heterogeneity has been observed.[158] Although the chromophore heterogeneity in crystal structures probably is an effect of crystal packing interaction, it shows a flexibility that might be highly important for signal transduction in phytochromes.

3.6. Summary crystallization and structure determination of CBD WT and CBD H290T

The developed crystallization conditions were used to crystallize both the CBD WT and the CBD H290T. The resolution of 1.35 Å is the highest resolution for a WT phytochrome fragment obtained so far. The crystallization conditions made the proteins crystallize as monomers instead of dimers that is the normal general interactions for other CBD structures. Although the structures are very similar to other CBD structures they differ in detail around the chromophore. The D-ring of the chromophore exhibits slightly different confirmations indicating a chromophore sensitivity and heterogeneity as an effect of crystal packing and the His290T mutation. The high resolution of the CBD H290T provides valuable structural information on waters that might be important for signal transductions in proteins.

3.7. The role of His290 in Bphps. Paper III

Although the many of the residues in the chromophore pocket are well conserved across phytochrome species, their functional role is not always well established. As mentioned previously, many of the residues have been shown important in photochemistry. One of them is the histidine (290) in H-bonding distance to the chromophore D-ring in the Pr state. It has been proposed to play an role in protonation reactions.[41] Additionally, it has also been seen as a strong candidate of 'locking' the D-ring in place in the Pr state.[96] The H290N and H290Q mutated *Dr*BphP showed incomplete photo conversion to Pfr but their Pr spectra were unaffected.[70] This would indicate the histidine is rather involved in stabilizing the Pfr state.

The incomplete photoconversion as a result of the missing histidine is also observed in **Paper III** for one unusual phytochrome *Sa*BphP1 from the myxobacterium, *Stigmatella aurantiaca.* The bacterium expresses two types of phytochromes of which, *Sa*BphP1 has its histidine naturally exchanged for a threonine. The threonine is

positioned too far away from the D-ring to form any direct interactions, as shown by crystallographic data.[96] Although the UV-VIS spectrum demonstrate an incomplete photo conversion to the Pfr state, the structural changes connected to the signaling are retained as shown by solution X-ray scattering for the PSM of *Sa*BphP1.[87]

In **Paper III** the role of the Histidine 290 is investigated by vibrational spectroscopy. The findings are strengthened by observations of the CBD H290T crystal structure.

3.7.1. The hydrogen bonding network to the D-ring carbonyl

The H-bond network around the D-ring has been shown to be important for the excited state life time and Lumi-R quantum yield in the BphPs, *Rp*BphP2 and *Rp*BphP3 from *Rhodopseudomonas palustris*.[159] The *Rp*BphP3 exhibits a longer excited state life time and lower Lumi-R yield than *Rp*BphP2. This was proposed to originate from the stronger hydrogen bonding of the D-ring carbonyl as determined by FTIR and ultrafast mid-IR.[160] In *Rp*BphP3 the D-ring carbonyl is able to form hydrogen bonds to two or three residues, whereas the *Rp*BphP2 only can form one hydrogen bond (to Histidine 290).[161]

A similar explanation has been proposed when comparing the Lumi-R quantum yield and excited state lifetime of the *Sa*BphP1 WT and T289H mutant. The absence of the histidine in *Sa*BphP1 was suggested to be responsible for its higher Lumi-R quantum yield and its shorter excited state life time compared to the T289H mutant, when the histidine is introduced to resemble prototypical BphPs.[96] This was proposed to be due to the much stronger hydrogen bond between the D-ring carbonyl and the histidine residue (2.7 Å) in the T289H mutant than in the Histidine less WT, where only a weak interaction (4 Å) can be made between the D-ring carbonyl and a serine residue (287). The weaker interaction was proposed to allow the D-ring to rotate more freely and explain the shorter excited lifetime and higher Lumi-R quantum yield.[96]

3.7.2. Vibrational spectroscopy studies of the D-ring H-bonding network

Vibrational spectroscopy can reveal valuable information about chemical groups in the protein and has previously provided insights about protonation events and Pr structural heterogeneity in phytochromes.[70,158,162] Furthermore the changes in secondary structure associated with the tongue refolding upon the Pr to Pfr transition were confirmed with Fourier transform Infrared spectroscopy (FTIR).[163]

Contradictory, IR spectra usually contains so much information that it is difficult to assign spectral features to specific protein sites and thereby the interpretation of spectra is limited. For BphPs, the fact that the peaks from carbonyl groups of the D- and the A-ring has been identified previously in the FTIR Pfr-Pr spectra makes it much easier interpret the carbonyl region of the IR spectrum.[78,164] When carbonyl groups (C=O) form hydrogen bonding interactions with the surroundings the C=O bond is weakened. This results in a downshift in the vibrational frequency of the C=O peak to lower wavenumbers in the spectrum. When exchanging H_2O to D_2O in the protein buffer the shifts are even more pronounced.

In **Paper III** the steady state and the time resolved FTIR spectra of the PSM from WT *Dr*BphP, WT *Sa*BphP1 and *Dr*BphP H290T are analyzed together with the CBD H290T structure solved to 1.4 Å. The IR-data show that D-ring carbonyl peak downshifts when histidine is replaced with a threonine in *Dr*Bphp H290T. This concludes that the C=O bond is weakened in the absence of the histidine residue, due to tighter hydrogen bonds with the surroundings in the Pr state. These results are in contradiction to the findings discussed previously that the missing histidine would lead to a less coordinated D-ring.[96]

3.7.3. Waters are responsible for stronger D-ring coordination

The *Sa*BphP1 structure (PDB 4RPW) has been solved to a 2.7 Å resolution and no waters are modelled in the structure. The CBD H290T structure which resembles the *Sa*BphP1 WT by lacking the highly conserved histidine was solved to a 1.4 Å. The high resolution allowed for modelling of waters as presented in **Paper III.** Two highly ordered waters are located in between the D-ring and the C-ring propionate group can be clearly modelled in both the CBD WT and the CBD H290T structures of *Dr*BphP (Figure 5 in paper III). These are also present in the CBD Y307S structure (PDB 4QOH) and other BphP structures and forms a H-bonding network between the D-ring, His290 and the C-ring propionate.

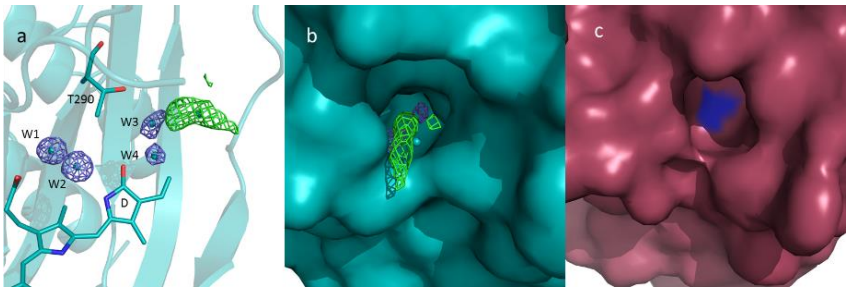

Figure 3.6 The removal of the histidine residues impact on the water network around the D-ring. a) The waters surrounding the D-ring carbonyl in the CBD H290T mutant. W1 and W2 are highly ordered and also present in the CBD WT. W3 and W4 are disordered and extra difference density for a possible water channel in connection with the solvent is shown in green. The density for the waters are shown at 1. 0 RMSD and the difference density is shown at 3.5 RMSD. b) The difference density representing a possible water channel forms a pore in the protein surface and open up for water exchange in the H290T mutant. c) In the WT structure this pore is sealed by the histidine residue (nitrogen shown in blue)

The void created when exchanging the His290 to a threonine is not empty in the CBD H290T structure. The space shows density for at least two additional water molecules (W3 and W4 in Figure 3.6a). The density is relatively weak but the composite omit map confirms that density is present (Figure 5 in paper III). The W3 is positioned so it is able to form hydrogen bonds with both the D-ring carbonyl and the threonine and thereby cover up for the absent histidine (Figure 5 paper III). We suggests that this (or these) water is responsible for the stronger coordination of the D-ring in the H290T mutant compared to the WT, which was determined by IR-spectroscopy.

The disordered waters are positioned slightly too close together (2.5 Å) to occupy their positions simultaneously. Furthermore, W4 might be positioned in an unfavorable angle to form direct interactions with D-ring carbonyl. The close distance between the two waters suggests the presence of a water cluster such as for example a $H_3O_5^+$ ion that would imply strong hydrogen bonding to the D-ring carbonyl. However, this cannot be determined from the data.

Furthermore, within H-bond distance to the waters an additional water is preset and associated with elongated difference electron density. The elongated density suggests the presence of a water channel connecting the D-ring with the solvent (Figure 3.6). This water channel forms a pore in the protein surface that is closed by the histidine residue in the WT structure. The water channel can account for the high disorder of the two additional water molecules present in the CBD H290T structure.

In *Sa*BphP1 the D-ring is thought to form a weak interaction with a serine (287). In *Dr*BphP the corresponding residue is an alanine that cannot form hydrogens bond to the D-ring. However, the serine in *Sa*BphP1 is no hindrance for the possibility of waters also present in the *Sa*BphP1 although they are not observed in the (low resolution) structure.

3.7.4. The possible roles of the conserved His290

Although the introduction of the histidine in *Sa*BphP1 lowers the Lumi-R yield. The *Dr*BphP H290T does not show a higher Lumi-R yield than the WT. This means that the presence of the histidine is not determinant for the Lumi-R quantum yield.

The absence of the histidine and the presence of extra waters leads to a stronger hydrogen network around the D-ring, but the waters are highly disordered. The disorder and the fast exchange of waters due to a water channel around the chromophore might allow the D-ring to rotate more freely. This might facilitate isomerization in both the forward and reverse transition between the Pr and Lumi-R state. The higher spectral heterogeneity previously observed in *Sa*BphP1[96] might also be explained by the high grade of disorder around the D-ring when the histidine is missing. The histidine might play a role in decreasing this disorder.

The Pfr structure of *Dr*BphP reveals that the photoactivation leads to the Z to E isomerization of the D-ring over the C-D double bond. The biliverdin is repositioned within the chromophore pocket and the His290 is moved closer to the C-ring propionate group (Figure3.5d).[42] For this to happen the ordered waters between the D-ring and the C-ring propionate must be moved. The disorder around the D-ring may facilitate the breakage of this hydrogen bonding network and allow easier migration of water molecules in the excited state. Since waters have higher mobility than protein side chains it might be feasible that the initial steps upon light activation would involve displacements of waters around the chromophore.

In conclusion, the role of the histidine is not directly to stabilize the D-ring conformation in the Pr state. Its role can be to close the water pore and reduce the

disorder around the chromophore and thereby be involved in the formation of the excited state. It can also be involved in protonation events in later stages of the photocycle or in stabilization of the Pfr state.

3.8. Summary Paper III

Paper III investigates the role of a highly conserved histidine in the phytochrome superfamily, that is absent in one of the phytochromes from *Stigmatella Aurantiaca.* This study was carried out by performing steady-state and time resolved vibrational spectroscopy providing insights into the strength of the hydrogen binding network around the D-ring in *Sa*Bphp1, WT *Dr*BphP and the mutated *Dr*BphP H290T.

The vibrational spectroscopy reveals that the absence of the histidine leads to a stronger coordination of the D-ring carbonyl. This is explained by the presence of additional waters in the space around the D-ring when histidine is exchanged to a threonine. The presence of the waters was revealed by the high-resolution crystal structure of the *Dr*BphP CBD H290T. The relatively weak electron density and positions indicate that the waters disorder might originate from the presence of a water channel in the CBD H290T structure.

The results elucidate the importance of the H-bonding network around the chromophore in BphPs and that higher resolution structures can provide new information about this topic which has not previously not been accounted for. The disorder around the D-ring might play an important role in the formation of the proteins Lumi-R state.

CHAPTER 4

4. Development of microcrystallization for SFX experiments. Paper II and IV

The large and well-diffracting crystals discussed in **Chapter 3** opened up for possibilities to perform time-resolved studies on the structural events in BphPs upon photo activation. The time-resolution of SFX (down to 100's of femtoseconds) is highly feasible to study the formation of the excited- and the Lumi-R state in BphPs. To achieve this the crystallization parameters had to be altered to obtain microcrystals of comparable diffraction quality. However, they needed to be smaller in size and homogenous enough to successfully be used at an SFX experiment. The high sample consumption put additional requirements on purification and crystallization to yield large amounts of sample. This chapter presents how the development of methods for two types of microcrystals from the same protein were conducted. Furthermore, it is described how these were used for successful SFX data collection at two XFELs with different sample injection systems as presented in **Paper II**. Furthermore I present the of development strategies for microcrystallization of the phytochrome PSM from different species for the same purpose. This led to the successful SFX structure determination of a PSM T289H mutant from *Stigmatella Aurantiaca* which is described in **Paper IV** and summarized in this chapter.

4.1. SFX data collection of the CBD at SACLA

4.1.1. Crushing of macrocrystals

Shortly after obtaining the large well diffracting crystals from the *Dr*BphP CBD WT the possibility of testing the crystals for SFX applications at the XFEL in Japan, SACLA opened up. Since there was no time for microcrystallization protocol development it was decided to mechanically crush the well diffracting CBD WT crystal, into small microcrystals. The method was adapted from successful SFX experiments on photosynthetic reaction center.[133] Crystallization drops were scaled up from 2+2 µl in hanging drop to 20+20 µl in sitting drop plates. Protein concentration had to be optimized to yield many crystals as homogenous in size as possible (Figure 4.1.) The final protein concentration used was 25 mg/ml and the protein was mixed with reservoir solution in a 1:1 ratio. After large crystals with dimensions around 300 μm^3 were formed they were collected into a 1.5 ml microcentrifuge tube by pipetting. One 24x24 well crystallization plate yielded close to one ml of crystal slurry.

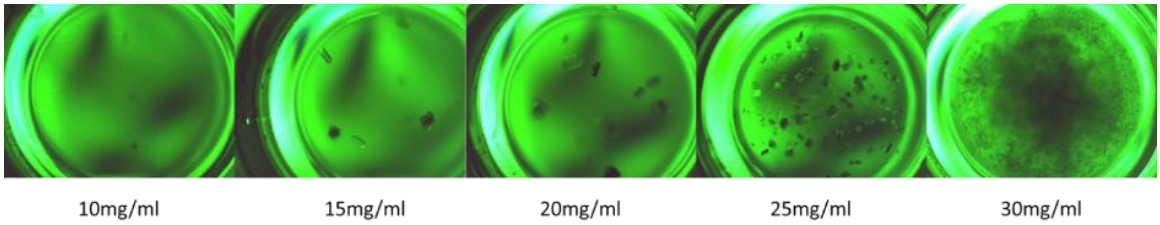

Figure 4.1. Concentration optimization in sitting drops. Different protein concentrations between 10-30 mg/ml were tested to obtain many crystals homogenous in size.

Two seed beads (Molecular dimensions) were added to the tube. The tube was vortexed for approximately 30 s and put on ice to reduce heat formation. The procedure was repeated until the crystals size was estimated to 10-20 µm under an optical microscope (Figure 4.2 left). Oversized crystals were excluded by filtering with a 20 µm cutoff spinning filter (Partec). Crystals were concentrated by a factor of approximately 20x by centrifugation and mixed with grease (Super lube, Syncho chemical corp) (Figure 4.2 right).[126]

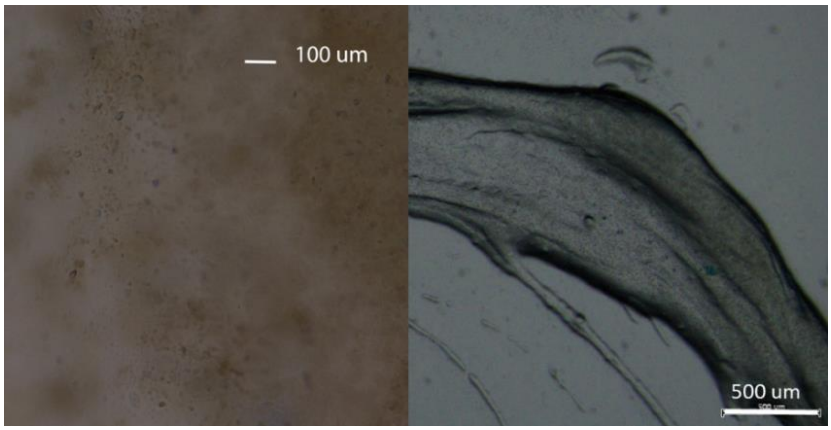

Figure 4.2. Obtained microcrystals by crushing. Small CBD crystals in various sizes between ~5-30 µm obtained by crushing of large crystals (left). The same crystals mixed in grease as a carrier media, used for injection into the XFEL beam (right).

4.1.2. Data collection and processing at SACLA

Diffraction data was collected at the BL3 beamline[165] with X-rays of 7.0 keV, 30Hz repetition rate and a <10 fs pulse duration. The sample dispersed in grease was extruded through a 110 µm diameter nozzle with a syringe-pump injector for high viscosity samples, at a flow rate of 500 nl/min. The injector was installed in a diffraction chamber filled with helium to reduce background scattering and the extruded grease was stabilized by flowing gas. Diffraction patterns were recorded on multiport CCD detector with an eight sensor module.[166] The hit rate varied between 5 15%. The images were filtered and converted by Cheetah (adapted for SACLA) [167] A full data set was acquired in a few hours.

4.2. SFX data collection of the CBD at the LCLS

4.2.1. Development of microcrystallization strategies

When the idea of SFX on phytochromes emerged the area of microcrystallization was even more unexplored than it is today. The new demands on crystal properties and few methods to evaluate microcrystal quality required development of new methods for the emerging field. At the time (2014), fewer than 20 protein structures acquired by SFX had been deposited into the protein data bank.[16] However, due to the intrinsic complex process of protein crystallization none of the crystallization conditions can be directly applied to any protein of interest. Proteins have their own specific requirements for crystallization, which makes it hard to develop common streamlined methods for "microcrystallization." In particular, in combination with the requirement that they should be produced in very high amounts in a reasonable time frame. Preferably on site for the experiment since transport can distort crystal quality.[135]

Although, the crushing of macrocrystals yielded a dataset at SACLA. The method is not always favorable and the LCLS experimental set up put other demands on sample compared to SACLA. The use of the GDVN has a substantially higher sample consumption compared to the high viscosity injector used at SACLA. Furthermore, the total experimental time at SACLA was only 3-4 hours compared to the planned >50 hours of beamtime at the LCLS.

First, the crushing method would have included putting up and harvest 40 µl drops to obtain at least 50-60 ml of sample would have been both time and work consuming. Second, the crushing yields crystals with a broad size distribution which is problematic in several ways. The large crystals exceeding the size limit of the jet must be filtered away which means that a lot of sample is lost if the crushing is insufficient. Furthermore, a sample with a homogenous size distribution is highly desirable since the attenuation of the beam is correlated with the diffraction of large crystals to protect the detector. This reduces the diffraction power of smaller crystals and can affect the final resolution of the dataset. Additionally, the size homogeneity of the crystals is important for laser penetration of light activated proteins for time resolved studies.

These demands required strategies to produce large volumes of small crystals with a narrow size distribution. As described in the methodology chapter, batch crystallization is a way to yield many nucleation sites which can grow into crystals that cease to grow when the protein concentration reaches the solubility curve. This was explored by setting up sitting drops without reservoir solution to equilibrate with. By varying the ratio between protein and precipitant solution the phase diagram was explored. Almost a linear decrease in crystal size was observed with a decreasing protein concentration and increasing precipitant concentration. This resulted in a

decrease of the crystal size from ~500µm to ~3-5 µm (Figure 4.3.) A key factor was also to crystallize at 4°C instead of room temperature.

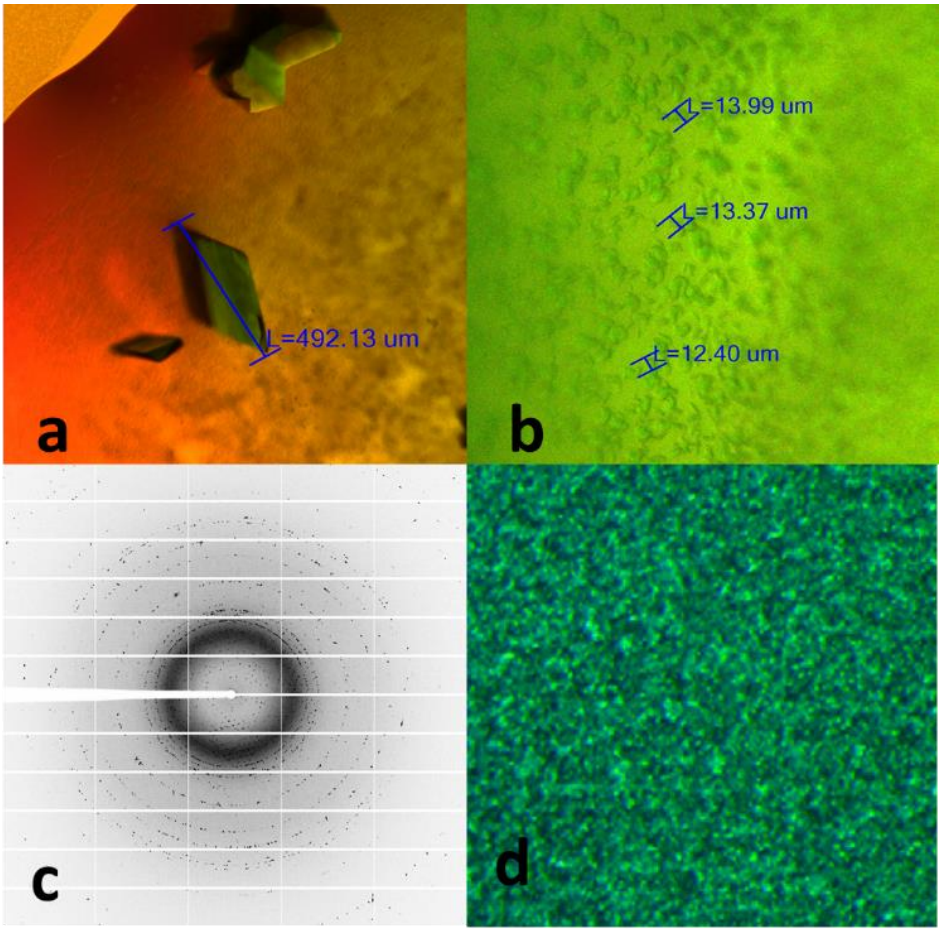

Figure 4.3. Development and validation of CBD microcrystals. a) Photograph of large CBD crystals formed in sitting drops with a 1:1 protein: precipitant solution ratio grown at 4°C. b) Photograph of microcrystals grown in the same conditions as in a) but with a 1:5 ratio between protein: precipitant and grown at 4°C. c) Semi-powder X-ray diffraction pattern from CBD microcrystals. Resolution 1.63Å at edge and 1.34 at corner. d) Photograph of batch crystals produced at the LCLS.

To get an estimation of the crystals diffraction quality the microcrystals were screened at the ESRF. A semi-powder diffraction pattern was obtained, which indicated a resolution beyond 1.4 Å (Figure 3.4c). The X-ray diffraction pattern obtained by many small crystals leads to a diffraction pattern consisting of a mixture between rings and spots. The optimal diffraction cannot be determined since the microcrystals are more sensitive to radiation damage. However an estimation of diffraction power and comparison of different conditions can be performed.

The batch crystallization in sitting drops was further scaled up to 500 µl of sample in 1.5 ml centrifugal tubes and it was found that the optimal conditions for crystals suitable for SFX at the LCLS was to mix 50 µl of protein (25mg/ml) with 450 µl of

reservoir. This mixture was vortexed immediately and incubated under constant mixing on a tipping table at 4°C for ~36 hours. This yielded a homogenous crystal sample with a 10-20 µm size distribution (Figure 4.3d). The crystals where concentrated by centrifugation into a crystal concentration of about 20% (v:v) to increase hit rate and was then filtered in a spin filter with a 20 µm cutoff to avoid clogging of the jet.

4.2.2. Data collection and processing at the LCLS

SFX data was acquired at the coherent X-ray imaging beamline (CXI) at the LCLS with a XFEL pulse of 1.8 mJ/pulse at 9.5 keV X-rays. The pulse duration was ~35 fs and the repetition rate 120 Hz.[168] The sample was injected by a GDVN with a flowrate of 30µl/min using a pressurized reservoir mounted on an antisettling device that prevented the sedimentation of the crystals. The size of the nozzle was 50-100 µm in diameter. The diffraction images were recorded with a CSPAD detector consisting of 64 tiles with a pixel size of 110x110 μm^2. The data was sorted with Cheetah.[142] A hit was defined as an image containing 20 peaks or more and the hit rate varied between 2-5%. The low hit rate was probably due to aggregation and sedimentation of crystals that was occurring over time, which was also observed under an optical microscope. The crystals belonged to the same space group as the low temperature CBD WT structure and the cell parameters were similar. See table 2. for more details on data collection and structure determination.

Table 2. Data collection and refinement statistics for SFX structures.

Data parameter	CBD LCLS	CBD SACLA
PDB ID	5L8M	5LBR
Space group	C121	C121
Unit cell dimensions a, b, c	94.1, 54.8, 69.9	96.2, 55.5, 71.6
α, β, γ (°)	90.0, 92.6, 90.0§	90.0, 92.8, 90.0
Resolution	38.2-2.1(2.15-2.1)	71.55-2.2 (2.28-2.2)
R_{split}	9.19 (85.43)	11.9 (67.7)
I/σ	7.77 (1.8)	5.62 (0.68)
CC(1/2)	98.0 (49)	97.6(60.4)
Completeness	99.97(99.8)	100(100)
Redundancy	819 (526)	81.2(24.1)
Number of reflections	18732(1202)	18389(1326)
No of hits/indexed hits	114409/90428 (79%)	30146/20014(66%)
Rwork/Rfree	0.176/0.218 (0.299/0.297)	0.182/0.203 (0.496/0.483)
Number of atoms	2505	2492
Average B-factor	42.92	50.26
R.m.s deviations Bond lengths(Å)	0.007	0.008
R.m.s deviations Bond angles(°)	1.313	1.287

4.3. SFX structures of the CBD

4.3.1. Structure determination from SFX data and structure refinement.

The structures were solved using molecular replacement and the refinement was performed as described previously for the synchrotron derived CBD WT structure (section 3.3) The same model (4Q0H) was used for regaining the phases. For details see table 2. Refinement was performed with REFMAC version 5.8.0135 [156] and the models were built in coot.

4.3.2. Temperature comparison of CBD structures

Three structures were obtained from the same crystallization conditions and thereby with highly similar crystal properties. This enabled the comparison of the impact from radiation damage and temperature on the structures. Since the crystal packing in the crystals was the same, differences cannot arise from crystal packing properties. The resolution of the data at SACLA was lower, it was due to a limitation that arose from the detector position. Also the pulse energy at SACLA is lower compared to the LCLS. Even though the diffraction power of the crushed crystals might have been affected by the harsh crushing procedure, it was not a decisive factor in this experiment.

The radiation damage in the CBD WT (now called LowT) structure could be elucidated together with temperature effects. Generally, the differences between the LCLS SFX and LowT structures are very small. The structures are close to identical when superimposed as seen in Figure 4.4a (RMSD =0.2 Å). Thus, confirming SFX as a valid method as a complement to datasets acquired at synchrotrons

$Fo_{(SFX)}$-$Fo_{(LowT)}$ difference maps were constructed as described for difference maps of time resolved data. The structure factors of the synchrotron acquired data set were treated as the native state and were subtracted from the structure factors obtained at the LCLS. This elucidate differences in the electron densities between the two data sets. The majority of the difference electron density peaks was located at the outer parts of the protein when comparing the LCLS structure with the LowT structure (Figure 4.4.). This is most likely an effect of the temperature differences since outer parts of the protein are more likely to be affected by thermal motions.

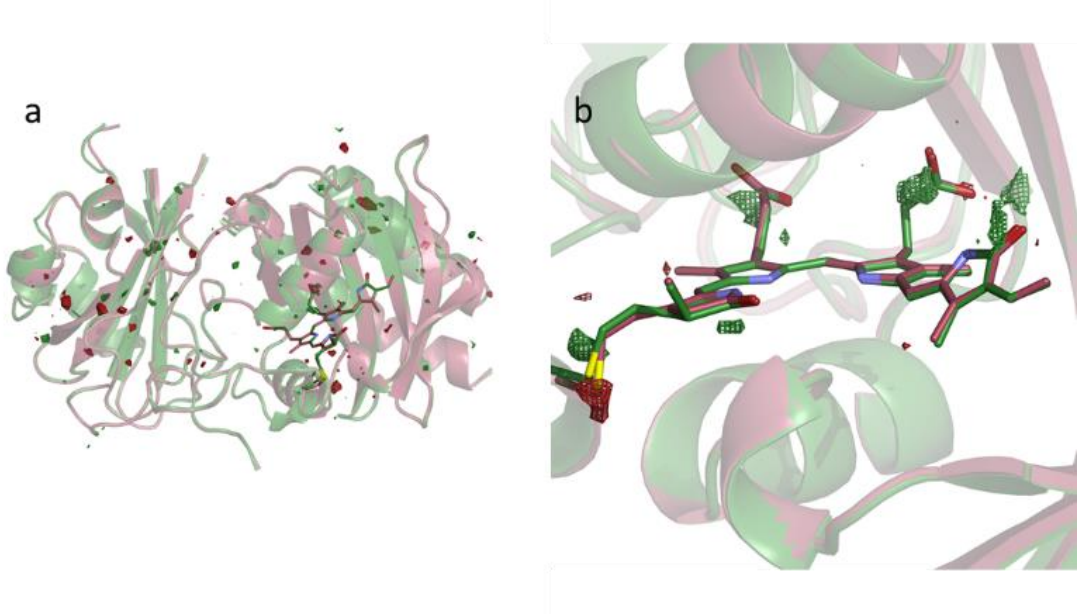

Figure 4.4. Comparison between CBD WT LowT (Synchrotron) and SFX structures of CBD. a) The synchrotron (raspberry) LowT CBD structure aligned with the SFX (LCLS) structure (green). Superimposed is the $Fo_{(LCLS)}$-$Fo_{(LowT)}$ electron difference map at 3.8 RMSD b) The zoom in view of the chromophore region of the same structures and map as in a).

Few difference peaks could be found in the protein and close to the chromophore (Figure 4.4b). In particular, the $Fo_{(SFX)}$-$Fo_{(LowT)}$ map show a negative peak at the bond between the cysteine residue (24) and the chromophore (Figure 4.4b). In the SFX structures this bond could be uniquely resolved but not in the LowT structure. This is due to that the bond is subjected to radiation damage in the Low T but not in the SFX structures. The other cysteines in the Low T structure were not affected by radiation damage.

The two SFX structures also show overall good agreement when superimposed (RMSD 0.4 Å) and the $Fo_{(LCLS)}$-$Fo_{(SACLA)}$ difference map with data cut at 2.5 Å contains quite few difference peaks. (Figure 5 in paper II). However, the same map with data where resolution up to 2.2 Å is included is less tidy and more crowded with difference peaks. This is explained by the low number of high resolution diffraction patterns collected at SACLA due to the smaller data set (see table 2)

As discussed in section 3.4.2. the LowT structure contains an extra water compared to the high resolution structure (4Q0H). This is correlated with repositions of residues, including Tyr263 that moves away from the chromophore. Although the water is not observed in the SFX structures the Tyr263 conformation remains the same as in the

LowT structure, suggesting that the water can still adopt the position but might be exchanged faster in the room temperature structures.

4.4. Summary Paper II

The paper present a novel crystallization condition for the Chromophore binding domain of the phytochrome from *Deinococcus radiodurans*. The obtained structure has the highest resolution (1.35Å) of any WT phytochrome fragment presented so far. The crystals adopted a new crystal packing as monomers.

From the novel crystallization conditions two methods to produce microcrystals that could fulfill the new demands from the emerging SFX field was developed. The first method performed as SACLA involved crushing of macro crystals and the resulting microcrystals where injected with a viscous media that has the advantage of less sample consumption. A data set was collected and the structure was solved to a 2.2 Å resolution.

To meet the sample requirements at LCLS a batch crystallization method for yielding microcrystals was also developed. This overcomes the disadvantages with the great size distribution of the crushed crystals. Furthermore, it saves both time and sample and yield high amounts of crystals with a much more narrow size distribution. The crystals were injected with a gas dynamic virtual nozzle at the LCLS and a data set was collected that yielded a SFX structure with a 2.1 Å resolution.

The three obtained structures enabled comparisons of the different methods used. Overall the structures agreed well which each other. Small differences was assigned to radiation damage and temperature effects. The results shows that the crystals are suitable for different applications and paves the way for future time resolved studies of phytochromes at XFELs. The methods for yielding microcrystals add experience to the field and suggest strategies to develop microcrystals for other proteins.

4.5. SFX structure of the PSM from *S.aurantiaca.* Paper IV

4.5.1. Purification and crystallization of PSM from different species

The final aim of the SFX experiment at the LCLS was to investigate the structural changes upon photoactivation of the protein. In addition to the resting state structure, some preliminary results from TR-SFX was obtained which are presented and discussed in **Chapter 5**. However, the low signal observed raised the question whether that the photoactivity of the CBD crystals is impaired or lowered. It is known that the CBD does not convert completely to the Pfr state[54] and it might even be hindered further by crystallization contacts in the crystals. In response to this, we developed

strategies to crystallize the PSM from different organisms. The aim was to improve resolution of these to visualize the structural changes at detailed level.

4.5.1.1. Mutation, purification and crystallization of the PSM of Agp1.

The structure with the highest resolution of a BphPs PSM published is the Agp1 from *Agrobacterium tumefaciens*.[64] The crystallization contacts in the protein are altered by the so called surface entropy reduction (SER) method. A web tool (http://services.mbi.ucla.edu/SER/) analyses the primary sequence of the protein. Clusters of amino acids of charged or bulky residues at the surface of the protein is identified. These can be mutated (for example to alanine) to minimize regions that might hinder the formation of crystal contacts. The cluster mutations can be combined to yield optimal surfaces to improve crystallization. In the case of Agp1 this was cluster 1 and 3 and the resulting Agp1 SER13 mutant. It had been crystallized in an antiparallel arrangement, which improved the resolution from the WT with almost 1 Å. (1.85 Å compared to 2.7 Å).[64]

The mutations to reproduce the construct were performed by site directed mutagenesis. The protein was expressed as *Dr*BphP CBD, with the alternations that the temperature was decreased to 18 degrees and the cells were cultivated for 36 hours after induction. The purification was performed similar to that of *Dr*BphP. However, NaCl and EDTA was included in the lysis buffer to keep the protein in solution. The protein was further precipitated with ammonium sulfate and loaded onto a Ni-NTA column for affinity chromatography. BV was bound for 30 min in a molar excess of ~10x and the protein was loaded on to a SEC column (Superdex 26/600 200 pg GE healthcare) The protein eluted as a dimer and monomer and were pooled and concentrated separately. The protein was crystallized in 2+2 µl hanging drop yielding needle shaped crystals of the monomer in the time frame of about five days. The crystallization conditions (0.03 M diethylene glycol, 0.03 M triethylene glycol, 0.03 M tetraethylene glycol, 0.03 M pentaethylene glycol, 20% glycerol, 10% PEG8000, 0.05 M MES, 0.05 M imidazole pH 6.5) was directly adapted from Nagano et al.[64] Batch crystallization was performed as for the *Dr*BphP CBD. The protein was mixed with precipitant solution in a 1:10 ratio and incubated for 48 hours on a tipping table.

4.5.1.2. Mutation, purification and crystallization of PSM SaBphP1- T289H

The same strategy as above of introducing SER mutations was applied to PSM-*Sa*BphP1 T289H from *Stigmatella aurantiaca*. In total 7 mutants with different combinations of mutated clusters were constructed. The expression and purification was performed as for Agp1 with exception that BV was incorporated directly after lysis. All constructs except one were soluble and one of the mutants formed crystals in many conditions. The crystals were screened for diffraction but never diffracted better than 4 Å at a synchrotron and are not suitable candidates for TR-SFX experiments.

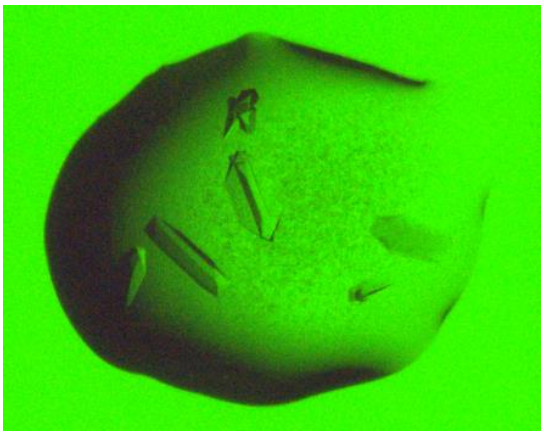

Figure 4.5 . Crystals obtained from the DrBphP PSM with the His-tag truncated.

4.5.1.3. Construct alternation, purification and crystallization of the PSM of DrBphP

The PSM of *Dr*BphP was screened for better crystallization conditions than published.[54,63] Both the his-tagged protein and a variant where the his-tag was removed were screened for obtaining better resolution. The purification of the his-tagged protein was performed as for the CBD constructs. The construct lacking the his-tag was purified with ion-exchange chromatography on a self-packed POROS column and eluted with a linear NaCl gradient. The screening was done using a mosquito crystallization robot (TTP labtech) and commercially available crystallization screens (Hampton research and molecular dimensions). The most promising crystals were formed in 0.1 M PCTP 7.0 25 % w/v PEG 1500 (Figure 4.5.). The crystals have diffracted to 3.4 Å at a synchrotron but has not been screened at an XFEL.

4.5.1.4. Purification and crystallization of the PSM WT and T289H SaBphP1.

Simultaneously, as we produced and crystallized the proteins mentioned above, collaborators managed to crystallize the PSM-WT *Sa*BphP1 and the mutant PSM *Sa*BphP1 H289T as well the CBD of the two types. The proteins were co expressed with heme oxygenase and purification was performed as previously described.[52] Large crystals were produced with the hanging drop method. For SFX experiments microcrystals of the PSM *Sa*BphpP1 T289H mutant was prepared by the batch method in a 1:1 ratio and left to mature for 6 hours at 16° C with continuous stirring.

4.5.2. SFX data collection at SACLA and data processing.

Microcrystals of PSM Agp1 SER 13 and *Sa*BphP1 T289H were tested for optimal diffraction at the beamline BL3 at SACLA. Due to time limitations it is only feasible to acquire data from one construct. The Agp1 SER 13 crystals only diffracted to ~7 Å but the *Sa*BphP1 diffracted to 3.15 Å Hence the *Sa*Bphp1 TH289H was chosen for data collection. The crystals were mixed with grease and injected as described previously (section 4.1.2.). A data set could was collected (see table 2 in paper IV for details). The phases were obtained with molecular replacement using the WT *Sa*BphP1 PSM

structure as a search model. The refinement was performed as described previously. The structure is presented in **Paper IV.**

4.5.3. SFX room temperature structure of the PSM *Sa*BphP1 T289H and comparison with other *Sa*BphP1 structures.

In **Paper IV** we also present cryo-synchrotron structures of the PSM and the CBD *Sa*BphP1 WT and T289H mutant. This enables both comparison of the effect on the CBD structure with the addition of the PHY-domain, and the structural effects of the mutation of the threonine 289 to a histidine. The SFX structure enables comparisons of temperature and radiation damage effects. Overall the structures presented are highly similar. All the structures have their BV attached to a cysteine (16) in the PAS domain and the loop in the GAF domain adopt the knot that is characteristic for all phytochromes. They are all crystallized in their dark adopted Pr state and show a cis-configuration of the chromophore D-ring. The PHY-domains in the PSM structures adopts a bent/kinked structure that pivots around residue 325 in the helical spine and the PHY-tongue exhibits a β-sheet fold and is in close contact with the chromophore environment (Figure 4.6).

4.5.3.1. Structural changes as a consequence of the PHY-domain addition to the CBD

The addition of the PHY-domain on both the WT and T289H mutant of the CBD results in similar changes. The chromophore pocket becomes tighter and the D-ring is buried deeper into the chromophore pockets which makes the residues move slightly. Asp208 (in the DIP motif) and Tyr262 makes contact with the PHY-tongue. These changes leads to an 8 nm red shift of the absorption maximum in the mutant (A_{max}=698nm) compared to WT (A_{max}=706nm). Furthermore, the addition of the PHY-domain alter the CBD-domains relative orientation. The helices involved in the dimerization of the two monomers are tilted in the PSM structure which leads to an increased angle between the dimerization interfaces of 15° (38° compared to 23°, Figure 3, **Paper IV**). This means to that the CBD structures cannot be superimposed on the PSM dimer.

4.5.3.2. Structural changes as a consequence of the T289H mutation

The CBD structures of the WT and the mutant are close to identical and RMSD between them is only ~0.3 Å. The small changes lies the close environment around the T289H mutation. The threonine to histidine substitution induces structural changes within the chromophore binding pocket and also lead to that the D-ring is buried further down in the pocket and is subjected to a more rigid and defined environment.

In the PSM structures the tongue region remains almost identical in the structures independent of the mutation. The WT and the T289H can be superimposed quite well (Figure 4.6). However the T289H mutation leads to a displacement of the PHY-domain in relation to the CBD compared to the WT PSM (Figure 4.6b). In the T289H structure the spine is slightly straightened and both of the PHY-domains are displaced in the same direction. This is in contrast to the Pr to Pfr transition in radiodurans when the PHY-domains moves in opposite directions.[42,63]. However, the proteins also crystallizes with different packings and space groups (C2 and P322 respectively) which indicate

that this is an effect of the crystal contacts and also demonstrates the high plasticity of the PHY-domain in the PSM.

4.5.3.3. Structural differences between synchrotron and SFX structures.

The PSM T289H structure was determined both under cryo conditions and by SFX at room temperature. The asymmetric unit of the PSM T289H structures contains three subunits. In the cryo structure subunit C is poorly resolved and contains large disordered areas. In the room temperature SFX structure the structure of subunit C is improved, although the overall resolution is lower. This indicates that the freezing of the crystals introduces disturbances in the crystal that affects the configuration of the PHY-domain and adds to the theory that its orientation is extremely sensitive to internal and external forces.

Generally the cryo-synchrotron and SFX structure of the PSM T289H are highly similar with a RMSD of 0.5 Å (Figure 4.6a). There is a small difference around the cysteine-biliverdin link that leads to a small tilt of the A-ring. However, this is the similar difference that was observed in **Paper II** and can be result of radiation damage in the synchrotron structure.

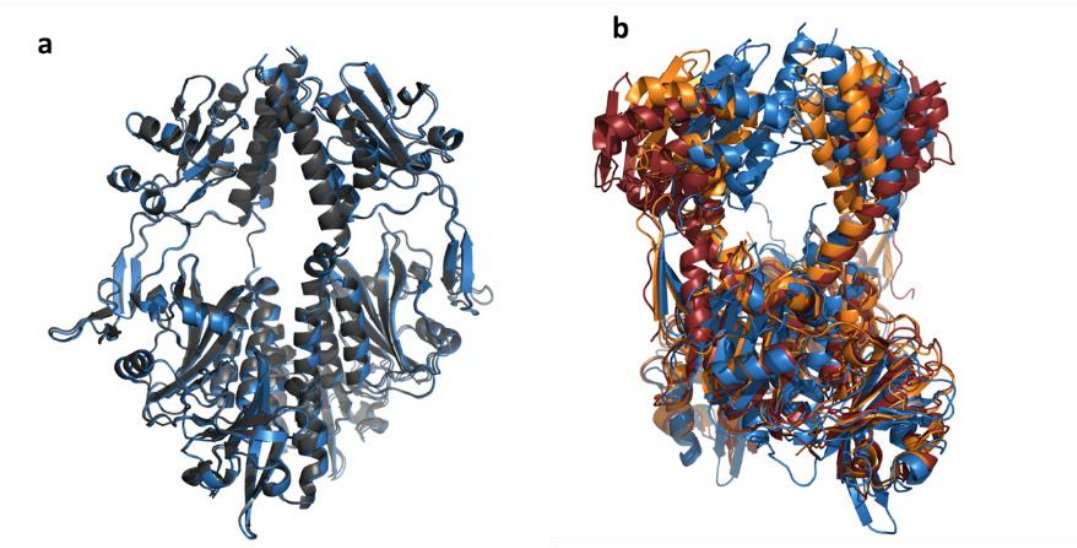

Figure 4.6. Structural comparison of the PSM of SaBphP1 from S.aurantiaca. a) The structural alignment of the SFX structure of *Sa*BphP1 T290H mutant (blue) and the cryo-structure of the same protein (grey). b) Side view of the structural comparison of the *Sa*BphP1 T290H (blue) with the *Dr*BphP PSM in Pr (orange PDB 4OUR) and in Pfr (ruby PDB 5C5K).

4.6. Summary Paper IV

The paper presents structural information about five different crystallographic structures of the *Sa*BphpP1 from *Stigmatella aurantiaca* in their dark adapted (Pr) state. Two WT structures, one of the CBD and one of the PSM. In addition three structures of the protein where the Threonine 289 was mutated to a histidine are provided. One of the CBD and two of the PSM of which, one was a SFX structure collected at room temperature. The mutation makes the protein resemble other bacteriophytochromes, where the histidine at this position is very well conserved. The general arrangement of the protein fragments and the structural features agrees very well with previously published structures. The five structures enabled comparison of temperature, mutation and PHY-domain addition effects of the phytochrome structure.

In general, the structures are highly similar but both the addition of the PHY- domain and the introduced mutation perturbs the structures slightly. Both leads to a tighter chromophore pocket. For the room temperature structure, the same reduction of radiation damage at the cysteine linkage as in **Paper II** was observed. The obtained structures provide information that can be used to discuss the mechanism of signal transduction in phytochromes in general and the physiological response in *S.aurantiaca* in particular.

Furthermore the paper presents evidence that the phytochrome signaling is involved in the cells tendency to form so called fruiting bodies and that it is inhibited by illumination of red light. This connects the phytochrome signaling to a physiological function in a non-photosynthetic prokaryote, something that has not been established before.

It appears like, the PHY-domain orientation is sensitive to both the mutation and freezing, which proposes a high sensitivity to internal and external structural alterations. This is in agreement of earlier observations of the high plasticity of the PHY-domain discussed in **Paper I**. I suggest that this plasticity of the PHY-domain limits the resolution of crystallized PSM constructs.

As described in this chapter, a number of PSM constructs have been screened for better crystallization conditions without much improvement. The screening was my major contribution to this work. The only two PSM structures available, that have a resolution better than 2 Å show antiparallel arrangement of their dimers.[61,64] The antiparallel arrangement might limit the plasticity of the PHY-domain and therefore be responsible for the higher resolution. I suggest that the way forward to obtain high resolution of PSM fragments is to reduce the plasticity of The PHY-domain. This can be achieved via introduction of SER-mutations, as already described for Agp1[64] or by other methods such as locking the PHY-domains together with covalent bonds. However, this have a high risk of limiting the signal transduction in the protein.

CHAPTER 5

5. Time-Resolved SFX of the Chromophore Binding Domain

The relation between structure and function cannot always be resolved by obtaining a static structure. Even if a mechanism can be proposed from an initial and a final structure, often information about structural intermediates are often required to fully understand the coupling between the proteins structural features and its reaction mechanism. By the use of time resolved SFX, structural snapshots of structures can be obtained at determined time points after activation of the protein.

The isomerization over the methine bridge between the C and the D-ring that leads to the rotation of the D-ring is confirmed by both crystal structures[42,60], NMR[79] and spectroscopy[82,159]. But information on how the isomerization reaction occurs and how the signal is further relayed to the rest of the protein is still sparse. IR-spectroscopy and NMR have revealed information about later states of the photocycle[73,83,161,163] but the structural information on the excited state, and the Lumi-R formation remains unclear.

The aim of the project and the data collection at the LCLS and SACLA was to obtain time resolved data of the photoactivation of phytochromes and try to settle the long standing debate on how the cis to trans D-ring isomerization proceeds as well as how the isomerization is relayed to the protein matrix. In phytochromes the life time of the excited states is much longer than in other proteins conducting an isomerization reaction, such as rhodopsin and PYP where it happens on a femtosecond time scale. Hence, it would be interesting to elucidate what the role of this long life time plays in phytochromes.

Paper II and IV report on resting state (dark) structures that can be seen as the first step towards time resolved experiments on phytochromes. In addition to the data of the dark (Pr) structure collected at the LCLS, some preliminary time resolved results from one time point was obtained and will be further discussed in this chapter.

5.1. Red laser activation of the CBD microcrystals as the LCLS

Since phytochromes are light sensitive proteins they are ideal candidates for time-resolved studies by a pump-probe experiment. When acquiring time-resolved data in pump-probe manner the crystals are activated by a laser pulse (pump) and then the structural changes are recorded by the X-ray pulse (probe) at a predetermined time point after the laser activation. The time point is chosen to capture structural events occurring in the protein. The data acquisition is designed to record images in a ‘laser on’ and ‘laser off’ alternating manner to minimize the buildup of a systematic error. In

this way every second diffraction image is assigned to the 'light' data set and every other is assigned to the 'dark'. This runs until several thousands of images have been collected and a 'dark' and a 'light' dataset has been acquired from that specific time point. This is then repeated for other time points of interest. In the case of the CBD *Dr*BphP the time point of 10 ps was chosen with the aim to capture the formation of the excited state. The experiment was conducted at the CXI beamline at the LCLS as described in chapter 4. The CBD microcrystals were illuminated with a 660 nm laser with a 150 fs pulse duration and a with a pulse energy varied between ~0.8-1.3 mJ/mm^2.

5.2. Creation of electron density difference maps

The diffraction data can be sorted into corresponding 'light' and 'dark' data set according to a code given to the recorded pattern in correlation with the timing tool. The data processing of the light data set was performed as for the dark data set described previously. From the two datasets difference maps can be created by subtracting the structure factors from the 'dark' state from the structure factors from the 'light' state, yielding a $Fo_{(light)}$-$Fo_{(dark)}$ map. This is feasible because the phases between the dark and the light activated crystals are assumed to be close to the same. Large changes within the protein structure would disrupt the crystal lattice and destroy the diffraction quality. The $Fo_{(light)}$-$Fo_{(dark)}$ electron density differences between the data sets and can identify where density in the dark state has disappeared and where new density has appeared in the light state. The differences in electron density can then be assigned to conformational changes in the protein between the two states when overlaid on the dark crystal structure model. To access the noise level in the data, a $Fo_{(dark)}$-$Fo_{(dark)}$ map was calculated by dividing the dark data in half and subtract the structure factors of one set from the other, thus representing the inherent noise levels between crystals.

If clear electron density differences are observed in the map the light state can be modelled by applying partial occupancy refinement. In general only a small fraction of the molecules are actually activated by the laser. In principle, alternate conformations are built in where strong electron density peaks are identified and the conformations are assigned occupancy (in percent). Then partial refinement then can be carried out against the structure factors from the light state, whilst the dark state conformations are kept constant. This allows for refinement only of the conformations in the light state.

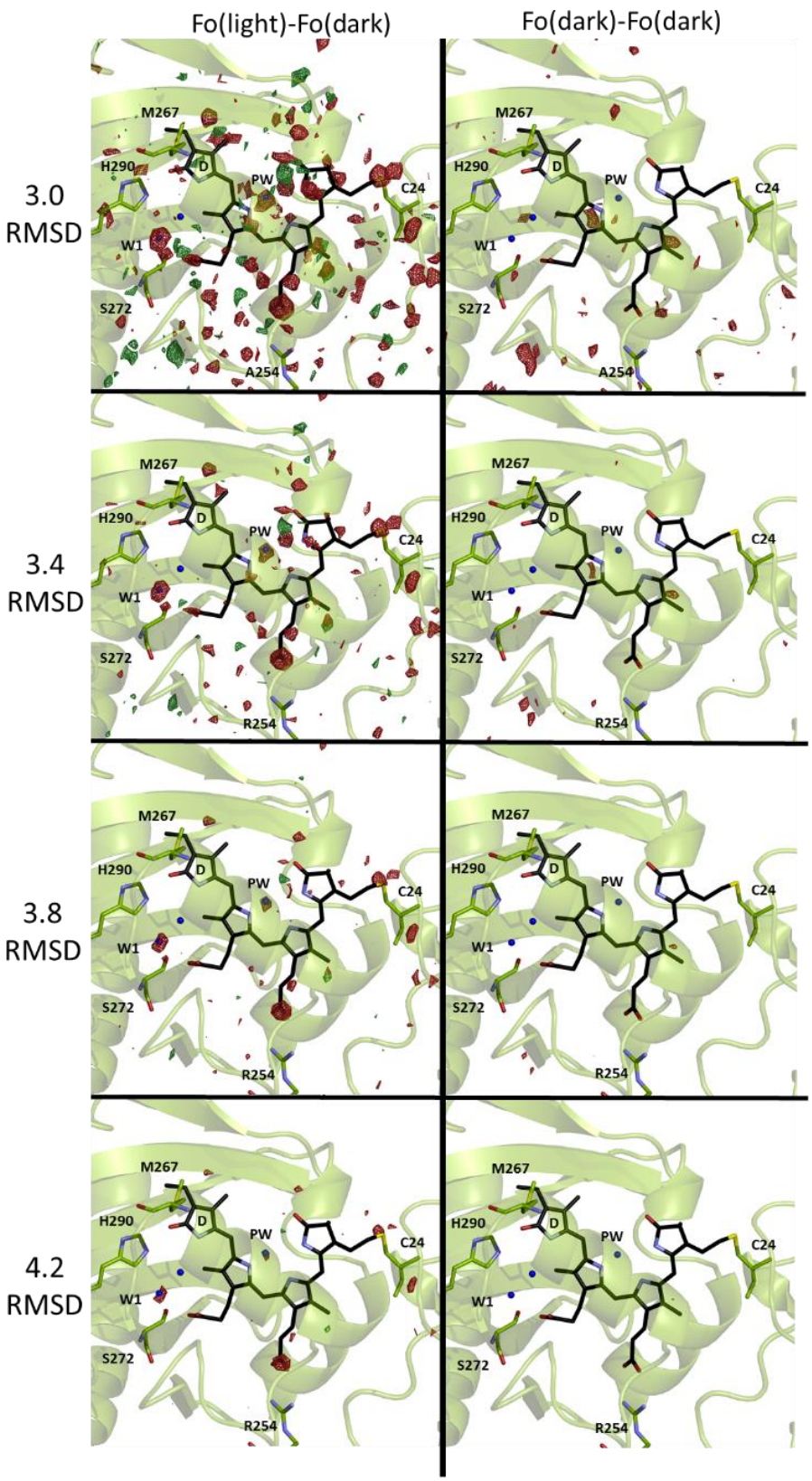

Figure. 5.1. Electron density difference maps calculated from the light and dark data sets overlaid on the dark SFX structure. The left column show the $Fo_{(Light)}$-$Fo_{(Dark)}$ map and the right the $Fo_{(Dark)}$-$Fo_{(Dark)}$ map representing the noise level. Each row showed at different RMSD (Positive density in green and negative in red.

5.3. Preliminary results from TR-SFX after 10 picoseconds

The obtained dark (Pr) structure reported in **Paper II** confirmed that the microcrystals could diffract to a resolution 2.1 at an XFEL, which makes them highly suitable for time resolved studies. We collected data for a time resolved data point of 10 ps for several shifts. However, a clearly identified signal was just obtained for one of the shifts. The reason for the low signal might be the generally low Lumi-R quantum yield (around 20%). If the laser excitation is also insufficient it leads to an even lower yield of the excited state in the photo activated crystals. The attempt to increase the difference signal by altering the laser excitation wavelength and power only led to a decrease in signal.

For the light and dark data sets collected at the LCLS of the CBD, electron density difference maps where created from the most successful shift (Figure 5.1). Generally the maps where noisy containing many difference peaks. In the left column the $Fo_{(Light)}$-$Fo_{(Dark)}$ is shown at increasing RMSD for each row and the corresponding noise level ($Fo_{(Dark)}$-$Fo_{(Dark)}$) is shown in the right column. At low RMSD (3.0) the $Fo_{(Light)}$-$Fo_{(Dark)}$ map is noisy. It contains a large number of positive(green) and negative (red) peaks that cannot be assigned to structural elements. When increasing the RMSD level so that the noise level is decreased more significant peaks in $Fo_{(Light)}$-$Fo_{(Dark)}$ can be read from the map. At 3.4 RMSD the noise level is low, showing only a few weak peaks and at even higher RMSD the (3.8 and especially 4.2) the noise is no longer detectable. The most pronounced difference peaks in the $Fo_{(Light)}$-$Fo_{(Dark)}$ are still detectable at the higher RMSD levels.

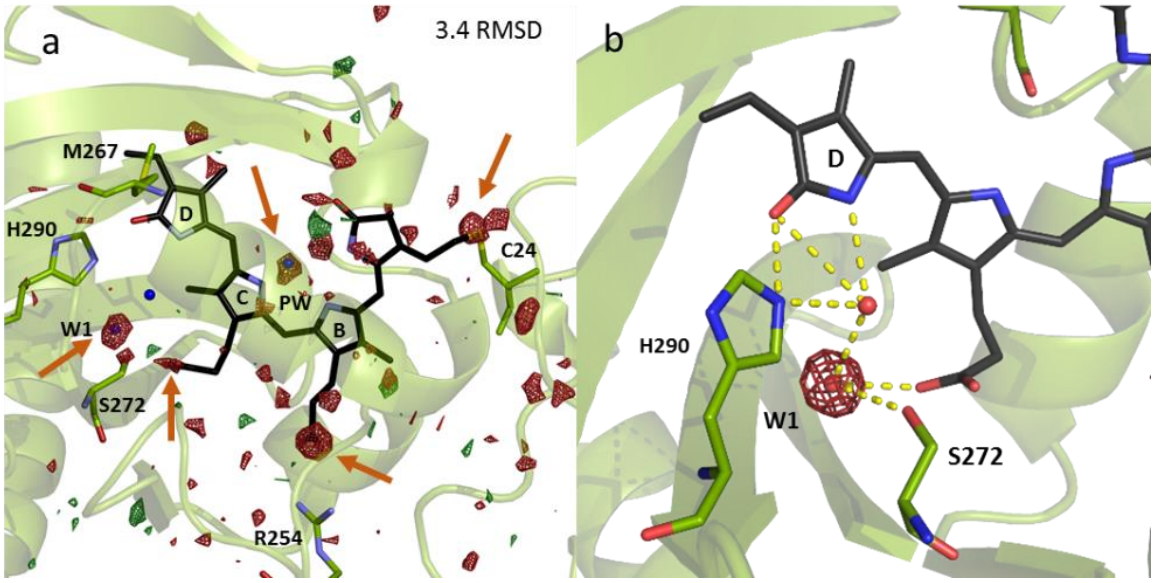

Figure. 5.2. Interpretation of the electron difference maps. *a)* $Fo_{(Light)}$-$Fo_{(Dark)}$ map shown at 3.4 RMSD. The most pronounced difference peaks are pointed out by arrows. b) *The hydrogen binding network between the D-ring, Histidine 290, Serine 272 and The C-ring propionate. The density for* Water 1 (W1) that is observed to move away in the resolved data is shown in red.

In Figure 5.2a the most pronounced difference signals at 3.4 RMSD are pointed out by arrows. They are the negative density peaks around propionate on the B-ring, on the cysteine linkage and around two waters (W1 and Pyrrole-water PW). All detectable also at an RMSD of 4.2 (Figure 5.1). There is additional negative density on the C-ring propionate and Ser272 but it is not detectable at higher RMSD levels. Water 1 (W1) in the Pr-state is highly ordered as pointed out in **Paper III** and forms a link in the hydrogen binding network between the D-ring carbonyl, His290, Ser272 and the C-ring propionate (Figure 5.2b). The negative difference electron density at W1 indicates that this water moves away in correlation with a repositioning of the C-propionate the Ser272 and the B-ring propionate breaks it salt bridge to Arg254.

Surprisingly, there is no difference density indicating that the D-ring has rotated at this time point. This suggests that the isomerization reaction happens later than 10ps. This is in support on IR-spectroscopy results that the D-ring isomerases on a 10ps timescale rather than within femtoseconds and is proceeded by structural changes in the protein.[169] But it could also be that the rotation of the D-ring is hindered in the crystal and the Lumi-R yield is very low. This would mean that the protein is locked in the excited state.

Regardless of the reason for no detectable signs of D-ring rotation it is feasible that what we observe is partly the excited state. The initial steps lies in loosening the hydrogen network around the chromophore by the repositioning B-ring propionate (and possible the C-ring propionate). The elimination of W1 would break the H-bonding network and thereby facilitate the subsequent rotation of the D-ring in the formation of the Lumi-R state. A corresponding detachment of the B-ring propionate has been observed previously for the Pfr to Pr transition by cryo-trapping of intermediate states in bathy *Pa*BphP.[60] The removal of water 1 and 2 (Figure 5.2b) is also essential for the allowing space for the biliverdin repositioning and the move of His290 closer to the C-propionate that is observed in the Pfr state.[42]

The results are in agreement with the results in **Paper III** that highlights importance of the hydrogen bonding network around the D-ring for formation of the excited state. The removal of the W1 would break the hydrogen bonding network that is more essential for the coordination of the D-ring than the direct interactions with the His290. However, to confirm the observed signal, further data is required and the signal must be improved by raising the laser power. The excited state could be recorded at an earlier time point of 1ps and the formation of the Lumi-r state should be observable on 1 ns timescale. This would reveal if the crystals are halted in the excited state or if time for the relaxation to the Lumi-R state is longer than 1 ps.

6. Concluding remarks and future perspectives

The papers comprising this thesis explore the structural features of bacteriophytochromes and the mechanisms involved in the transition from the dark adapted (Pr) to the light activated meta-stable (Pfr) state.

Paper I is a review which compiles the current structural information on the different states in the photocycle and proposes a structural mechanism for the signal transduction in BphPs. This is done by combining X-ray crystallographic structures on shorter BphP fragments with findings acquired from other methods such as solution X-ray scattering and FTIR. The future challenges lies in structure determination of a BphPs full-length structure, with a histidine kinase attached, in both its dark adapted and light activated state to confirm the structural mechanism. Furthermore, the information of the phytochromes interactions with their signaling partner, the response regulator remains sparse. It is also desirable to crystallize plant phytochromes to obtain more knowledge about their special features. **Paper I** also points out the lack of structural information available for the BphPs intermediate states formed before the light activated meta-stable Pfr state. To obtain structural information on these short (pico-ms timescales) other methods than conventional crystallography are required. One suitable method is the Time-Resolved Serial Femtosecond Crystallography presented in **Chapter 5**.

Paper II presents novel crystallization conditions that yielded crystals of a wildtype phytochrome fragment with the highest resolution so far. The near atomic 1.35 Å resolution is promising for TR-SFX studies of BphPs visualize the detailed structural events of biliverdin isomerization upon photoactivation. A first SFX structure of the CBD-domain was achieved by crushing the big microcrystals and injecting them into the XFEL beam at SACLA. A batch method for generating large amounts of crystals homogenous in size were developed. This is important for sample conserving purposes and more homogenous laser penetration in future TR-SFX experiments. These crystals was used for obtaining another SFX structure at the LCLS presented in **Paper II**. The room temperature structures were similar to the cryo-conditions but some effects of possible radiation damage could be identified on the cysteine linkage between the protein and the biliverdin chromophore in the synchrotron structure.

The same crystallization conditions developed for the WT CBD could be used for crystallization of the CBD H290T mutant which is presented in **Paper III**. The structures are close to identical to the CBD WT although chromophore structural heterogeneity can be identified when comparing the CBD structures in **Paper II** and **Paper III** with previously published phytochrome structures.

Both **Paper III** and **IV** investigates the role of the Histidine residue in H-bonding distance of the carbonyl of the D-ring. In *Sa*BphP1 from *S.aurantiaca* this residue is naturally exchanged for a threonine and in **Paper IV** this residue is mutated to a His to

resemble other phytochromes and in the opposite manner the histidine in *Dr*BphP is mutated to a threonine in **Paper III** to mimic *Sa*BphP1.

Paper III reveals information about the role of the histidine by analyzing the hydrogen bonding network around the D-ring by infrared spectroscopic methods. The conclusion is that the D-ring is more strongly coordinated in the absence of the Histidine residue. Although the threonine is situated too far away to directly coordinate the D-ring it allows more space for water molecules in the area. This leads to a stronger H-bonding network as confirmed by additional electron density around the D-ring in the H290T crystal structure together with peak shifts in the IR-data. The additional waters are highly disordered, and this disorder might be important for the initial steps upon photoactivation

Paper IV presents a SFX structure of the full photosensory module, PSM from *S.aurantiaca*. This structure was collected at SACLA after crystallization attempts of several PSM constructs from different species were the *Sa*BphP1 T289H gave the best resolution. However, even better resolution might be desirable to obtain detailed information about the structural changes associated with photoactivation.

Additionally, I report TR- SFX data of a phytochrome fragment of at 10 ps time point after laser activation collected at the LCLS. The data is preliminary but indicate that the D-ring isomerization is not occurring at this time scale. A loosening of the hydrogen bonding network around the D-ring by the removal of a water was observed. Both this and the results from **Paper III** emphasize the importance of the water hydrogen bonding network around the chromophore for photoactivation.

Altogether, the crystallization of BphPs and the SFX structures presented in **Paper II** and **IV** lay the foundation for future TR-SFX of phytochromes. Although the CBD might not photoconvert completely it is possible that it can be used to provide structural information of the formation of the excited state. In future SFX experiments the intensity of the red light laser should be increased to activate a larger proportion of the proteins. An increased excited state population would raise the signal above the noise level. Several time points would be interesting to consider for future TR-SFX experiments. Further data collection is needed to determine the excited state. It can either be done at 10 ps to complete the data presented in chapter 5. Alternatively, collection at 1ps might be more suitable to avoid that the excited state partly relaxes back to the resting dark (Pr) state. Collection at 10 ns after photoactivation would capture the rotation of the D-ring and confirm if this happens in the CBD construct.

If the rotation of the D-ring is hindered in the CBD crystals it would be more feasible to collect data on the PSM to study the formation of the Lumi-R state. However, further crystal optimization is needed to improve resolution for this to be achievable. The plasticity of the PHY-domain probably limits the resolution of the PSM constructs.

Hence, methods to reduce this without affecting the formation of the Lumi-R state are needed.

Until now, TR-SFX has mostly been performed on light activated proteins but is evolving to include other types of proteins, thanks to applications such as ‘mix and inject’ for ligand activated enzymes. The growing field put new demands on crystallization strategies and the microcrystallization development presented in **Paper II and IV** can be used as inspiration for microcrystallization of other proteins.

The field of SFX is growing and is likely to have a bright future for studies of all kind of reactions and to make ‘molecular movies’ of the structural events. The bright flux also starts to blur the boundaries between crystallography and single particle structure determination since it requires smaller and smaller crystals.

In conclusion this work has led to both structural and functional insights about bacteriophytochromes as well as development of the currently emerging SFX field. Especially, the advancement of crystallization methods to yield suitable samples for this expanding method that enables the visualization of protein structural dynamics with an ultrafast time resolution.

Chapter 7

7. Acknowledgements

This work is of course not a one woman's contribution. Without all support and help from both professionals and people outside the scientific bubble it would never have been possible. I started working the Lundberg lab more than 8-years ago when I got the opportunity to do a bachelors diploma work. Since then I have hanged around, learned a lot and met so many inspiring people. I am deeply grateful for all the experiences it has provided but also very content that my time to leave for the next adventure has finally arrived.

My first and greatest thanks goes to my supervisor **Sebastian**, who finally decided it was a good thing to get a biochemist as a PhD student after some doubts. I hope that you still think it was a good decision. I am thankful for the great independence you have given me and I truly admires your ability to always have the balance between being extremely focused on problem solving, enthusiastic and 'somewhat' realistic in your ideas. Like when I called you in the middle of the night from japan with crystal issues, and you in some way managed to get a tube of green slurry wrapped in foil from the lab, via a detour to France to Japan, in your hand luggage.

When I had to leave the lab for some 7 months just before the upcoming LCLS beamtime **Elin C** got onboard on the SFX train and managed everything like a boss. Since then we have worked close together, solving all kind of problems and issues. Although we are different I appreciate your hard work and I could never have found a better partner to both cry, laugh and travel with. Thank you for always listening and saying exactly your thoughts like a true friend. I have learned a lot and I am convinced you will do excellent.

Thanks to **Leo** that also helped out with the prepping and crystallization. We had not been able to do anything without those amounts of protein.

I am grateful that you, **Emil** accepted to be my student although my not really 'correct' recruitment methods and took on the struggle with those impossible mutants. Thanks to **Oskar** for always helping out with *everything* and being the calm (when sober) opposite of me when I freak out (like in Argentina). If it were not for the rest of the Westenhoff crew it would have not been the same; **Alex B, Maria, Stephan, Ash, Linnea, Joachim, Matt.** Thanks!

None of this work had been possible without the collaboration with the Finnish people. Thanks to **Heikki** and **Janne** who has been invaluable as phytochromes experts.

I am also very grateful for having **Richard** as my examiner. You are such an inspiring person and thanks for keeping the free scientific atmosphere in the corridor and recruiting fabulous people that make this corridor what it is.

I miss all the 'oldies' and our BCs. **Linda** who got me into the corridor from the beginning, being my supervisor and drinking too much beer with me on all kind of

occasions together with other goldies such as the ginger men, **David** (miss your dances) and **Rob** (thanks for sharing office and Japan with me) as well as the one and only, **Kakan**. And of course all the other oldies, **Gerhard, Anette, Erik, Mikael E, Ida, Mike, Jennie, Madde, Anna.**

Thanks to the other seniors; **Gergerly, Rosie, Gisela, Kristina, Björn** and **Johanna** (Thanks for letting me onboard on the gender balance), who always answer all kind of questions and recruit a lot of diverse people that make this environment what is.

All the present people; **Petra B** who helped me out on all kind of biochemistry related issues and especially the waste. **Elin D**. you will do great! **Rebecka**, miss you and all the laughs and weird subjects during lunch and **Cecilia W**, soon is your turn; Good luck! **Rajiv**, you made all the difference for me in now in the end being in the same situation. **Rhawnie**, good luck now! Thanks to **Weixiao**, you always have an answer for everything! **Majo**, I will always be grateful for your kindness and help with the data. **Cecilia S** you will see that your work will pay off. **Florian** the sandwich and present aquaporing master. **Daniel**, keep up working on your pymol skills. **Greger**, **Viktor, Rob B, Alex J, Stefan, Davide, Dimitra, Darius and Tinna** keep up the good work**.** Welcome to all the newbies; **Maja, Giorgia, Emelie, Swagatha Stanislav, Laura, Andreas** and **Per** (welcome to the office.)

Thanks to **Lars** and **Bruno** who keep the lab from falling apart completely.

Tack till min familj och mina fina vänner som alltid får stå ut med att jag “måste jobba”. Ni är bäst!

Tack till **John** för att du alltid ställer upp och förlåt för allt du fått/får stå ut med. Utan dig hade det inte blivit något alls. Nu ska det bli lättare hoppas jag. Älskar dig mest.

8. Bibliography

1 Pierik, R. & de Wit, M. Shade avoidance: phytochrome signalling and other aboveground neighbour detection cues. *Journal of Experimental Botany* **65**, 2815-2824, doi:10.1093/jxb/ert389 (2014).

2 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

3 Ansari, A. *et al.* Protein States and Protein Quakes. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 5000-5004, doi:DOI 10.1073/pnas.82.15.5000 (1985).

4 Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **254**, 1598-1603, doi:DOI 10.1126/science.1749933 (1991).

5 Voet, D. & Voet, J. G. *Biochemistry*. 4th edn, (John Wiley & Sons, 2011).

6 Williams, R. J. P. Conformational Properties of Proteins in Solution. *Biol Rev* **54**, 389-437, doi:DOI 10.1111/j.1469-185X.1979.tb00843.x (1979).

7 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972, doi:10.1038/nature06522 (2007).

8 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).

9 Blundell, T. L. & Johnson, L. N. *Protein crystallography*. (Academic Press, 1976).

10 Bilderback, D. H., Elleaume, P. & Weckert, E. Review of third and next generation synchrotron light sources. *J Phys B-at Mol Opt* **38**, S773-S797, doi:10.1088/0953-4075/38/9/022 (2005).

11 Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* **66**, 339-351, doi:10.1107/S0907444910008656 (2010).

12 Garman, E. F. & Weik, M. Radiation Damage in Macromolecular Crystallography. *Methods Mol Biol* **1607**, 467-489, doi:10.1007/978-1-4939-7000-1_20 (2017).

13 Smith, J. L., Fischetti, R. F. & Yamamoto, M. Micro-crystallography comes of age. *Curr Opin Struc Biol* **22**, 602-612, doi:10.1016/j.sbi.2012.09.001 (2012).

14 Levantino, M., Yorke, B. A., Monteiro, D. C. F., Cammarata, M. & Pearson, A. R. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr Opin Struc Biol* **35**, 41-48, doi:10.1016/j.sbi.2015.07.017 (2015).

15 Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77, doi:10.1038/nature09750 (2011).

16 Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends in Biochemical Sciences* **42**, 749-762, doi:10.1016.j.tibs.2017.06.007 (2017).

17 Martin-Garcia, J. M., Conrad, C. E., Coe, J., Roy-Chowdhury, S. & Fromme, P. Serial femtosecond crystallography: A revolution in structural biology. *Arch Biochem Biophys* **602**, 32-47, doi:10.1016/j.abb.2016.03.036 (2016).

18 Garner, W. W., Allard, H.A. . Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J Agric Res 18: 553–606* **18**, 553-606 (1920).

19 Borthwick, H. A., Hendricks, S. B., Parker, M. W., Toole, E. H. & Toole, V. K. A Reversible Photoreaction Controlling Seed Germination. *Proc. Natl. Acad. Sci. U. S. A.* **38**, 662-666 (1952).

20 Butler, W. L., Norris, K. H., Siegelman, H. W. & Hendricks, S. B. Detection, Assay, and Preliminary Purification of the Pigment Controlling Photoresponsive Development of Plants. *Proc. Natl. Acad. Sci. U. S. A.* **45**, 1703-1708 (1959).

21 Vierstra, R. D. & Quail, P. H. Photochemistry of 124 kilodalton Avena phytochrome in vitro. *Plant Physiol* **72**, 264-267 (1983).
22 Hershey, H. P., Barker, R. F., Idler, K. B., Lissemore, J. L. & Quail, P. H. Analysis of cloned cDNA and genomic sequences for phytochrome: complete amino acid sequences for two gene products expressed in etiolated Avena. *Nucleic Acids Res* **13**, 8543-8559 (1985).
23 Blumenstein, A. *et al.* The Aspergillus nidulans phytochrome FphA represses sexual development in red light. *Curr. Biol.* **15**, 1833-1838, doi:10.1016/j.cub.2005.08.061 (2005).
24 Kehoe, D. M. & Grossman, A. R. Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* **273**, 1409-1412 (1996).
25 Hughes, J. *et al.* A prokaryotic phytochrome. *Nature* **386**, 663, doi:10.1038/386663a0 (1997).
26 Davis, S. J., Vener, A. V. & Vierstra, R. D. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. *Science* **286**, 2517-2520 (1999).
27 Deng, X. W. & Quail, P. H. Signalling in light-controlled development. *Semin Cell Dev Biol* **10**, 121-129, doi:10.1006/scdb.1999.0287 (1999).
28 Wang, H. & Deng, X. W. Dissecting the phytochrome A-dependent signaling network in higher plants. *Trends Plant Sci* **8**, 172-178, doi:10.1016/S1360-1385(03)00049-9 (2003).
29 Jiao, Y., Lau, O. S. & Deng, X. W. Light-regulated transcriptional networks in higher plants. *Nat Rev Genet* **8**, 217-230, doi:10.1038/nrg2049 (2007).
30 Robson, P. R., McCormac, A. C., Irvine, A. S. & Smith, H. Genetic engineering of harvest index in tobacco through overexpression of a phytochrome gene. *Nat Biotechnol* **14**, 995-998, doi:10.1038/nbt0896-995 (1996).
31 Moglich, A. & Moffat, K. Engineered photoreceptors as novel optogenetic tools. *Photoch Photobio Sci* **9**, 1286-1300, doi:10.1039/c0pp00167h (2010).
32 Chernov, K. G., Redchuk, T. A., Omelina, E. S. & Verkhushaa, V. V. Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. *Chem Rev* **117**, 6423-6446, doi:10.1021/acs.chemrev.6b00700 (2017).
33 Kaberniuk, A. A., Shemetov, A. A. & Verkhusha, V. V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nature Methods* **13**, 591-+, doi:10.1038/Nmeth.3864 (2016).
34 Shcherbakova, D. M. *et al.* Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat Commun* **7**, 12405, doi:10.1038/ncomms12405 (2016).
35 Shcherbakova, D. M., Baloban, M. & Verkhusha, V. V. Near-infrared fluorescent proteins engineered from bacterial phytochromes. *Curr Opin Chem Biol* **27**, 52-63, doi:10.1016/j.cbpa.2015.06.005 (2015).
36 Burgie, E. S. & Vierstra, R. D. Phytochromes: an atomic perspective on photoactivation and signaling. *Plant Cell* **26**, 4568-4583, doi:10.1105/tpc.114.131623 (2014).
37 Auldridge, M. E. & Forest, K. T. Bacterial phytochromes: More than meets the light. *Crit Rev Biochem Mol* **46**, 67-88, doi:10.3109/10409238.2010.546389 (2011).
38 Burgie, E. S., Bussell, A. N., Walker, J. M., Dubiel, K. & Vierstra, R. D. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10179-10184, doi:10.1073/pnas.1403096111 (2014).
39 Vierstra, R. D. & Zhang, J. Phytochrome signaling: solving the Gordian knot with microbial relatives. *Trends Plant Sci* **16**, 417-426, doi:10.1016/j.tplants.2011.05.011 (2011).

40 Vierstra, R. D. & Davis, S. J. Bacteriophytochromes: new tools for understanding phytochrome signal transduction. *Semin Cell Dev Biol* **11**, 511-521, doi:10.1006/scdb.2000.0206 (2000).
41 Song, C. *et al.* Two ground state isoforms and a chromophore D-ring photoflip triggering extensive intramolecular changes in a canonical phytochrome. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3842-3847, doi:10.1073/pnas.1013377108 (2011).
42 Burgie, E. S., Zhang, J. & Vierstra, R. D. Crystal Structure of Deinococcus Phytochrome in the Photoactivated State Reveals a Cascade of Structural Rearrangements during Photoconversion. *Structure* **24**, 448-457, doi:10.1016/j.str.2016.01.001 (2016).
43 Mitrophanov, A. Y. & Groisman, E. A. Signal integration in bacterial two-component regulatory systems. *Gene Dev* **22**, 2601-2611, doi:10.1101/gad.1700308 (2008).
44 Yang, X. *et al.* Light Signaling Mechanism of Two Tandem Bacteriophytochromes. *Structure* **23**, 1179-1189, doi:10.1016/j.str.2015.04.022 (2015).
45 Moglich, A., Ayers, R. A. & Moffat, K. Addition at the molecular level: signal integration in designed Per-ARNT-Sim receptor proteins. *J Mol Biol* **400**, 477-486, doi:10.1016/j.jmb.2010.05.019 (2010).
46 Rockwell, N. C. & Lagarias, J. C. A brief history of phytochromes. *Chemphyschem* **11**, 1172-1180, doi:10.1002/cphc.200900894 (2010).
47 Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu Rev Plant Biol* **57**, 837-858, doi:10.1146/annurev.arplant.56.032604.144208 (2006).
48 Li, H., Zhang, J., Vierstra, R. D. & Li, H. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10872-10877, doi:10.1073/pnas.1001908107 (2010).
49 Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**, 325-331, doi:10.1038/nature04118 (2005).
50 Karniol, B., Wagner, J. R., Walker, J. M. & Vierstra, R. D. Phylogenetic analysis of the phytochrome superfamily reveals distinct microbial subfamilies of photoreceptors. *Biochem J* **392**, 103-116, doi:10.1042/BJ20050826 (2005).
51 Wagner, J. R., Zhang, J., Brunzelle, J. S., Vierstra, R. D. & Forest, K. T. High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem* **282**, 12298-12309, doi:10.1074/jbc.M611824200 (2007).
52 Yang, X., Stojkovic, E. A., Kuk, J. & Moffat, K. Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12571-12576, doi:10.1073/pnas.0701737104 (2007).
53 Auldridge, M. E., Satyshur, K. A., Anstrom, D. M. & Forest, K. T. Structure-guided Engineering Enhances a Phytochrome-based Infrared Fluorescent Protein. *Journal of Biological Chemistry* **287**, 7000-7009, doi:10.1074/jbc.M111.295121 (2012).
54 Burgie, E. S. *et al.* Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion. *J Biol Chem* **289**, 24573-24587, doi:10.1074/jbc.M114.571661 (2014).
55 Li, F. *et al.* X-ray radiation induces deprotonation of the bilin chromophore in crystalline D. radiodurans phytochrome. *J Am Chem Soc* **137**, 2792-2795, doi:10.1021/ja510923m (2015).

56 Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat Methods* **14**, 443-449, doi:10.1038/nmeth.4195 (2017).
57 Yang, X., Kuk, J. & Moffat, K. Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: photoconversion and signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14715-14720, doi:10.1073/pnas.0806718105 (2008).
58 Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14709-14714, doi:10.1073/pnas.0806477105 (2008).
59 Yang, X., Kuk, J. & Moffat, K. Conformational differences between the Pfr and Pr states in Pseudomonas aeruginosa bacteriophytochrome. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15639-15644, doi:10.1073/pnas.0902178106 (2009).
60 Yang, X., Ren, Z., Kuk, J. & Moffat, K. Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome. *Nature* **479**, 428-432, doi:10.1038/nature10506 (2011).
61 Mailliet, J. *et al.* Spectroscopy and a High-Resolution Crystal Structure of Tyr263 Mutants of Cyanobacterial Phytochrome Cph1. *Journal of Molecular Biology* **413**, 115-127, doi:10.1016/j.jmb.2011.08.023 (2011).
62 Anders, K., Daminelli-Widany, G., Mroginski, M. A., von Stetten, D. & Essen, L. O. Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. *J Biol Chem* **288**, 35714-35725, doi:10.1074/jbc.M113.510461 (2013).
63 Takala, H. *et al.* Signal amplification and transduction in phytochrome photosensors. *Nature* **509**, 245-248, doi:10.1038/nature13310 (2014).
64 Nagano, S. *et al.* The Crystal Structures of the N-terminal Photosensory Core Module of Agrobacterium Phytochrome Agp1 as Parallel and Anti-parallel Dimers. *J Biol Chem* **291**, 20674-20691, doi:10.1074/jbc.M116.739136 (2016).
65 Bellini, D. & Papiz, M. Z. Structure of a bacteriophytochrome and light-stimulated protomer swapping with a gene repressor. *Structure* **20**, 1436-1446, doi:10.1016/j.str.2012.06.002 (2012).
66 Otero, L. H. *et al.* Structure of the Full-Length Bacteriophytochrome from the Plant Pathogen Xanthomonas campestris Provides Clues to its Long-Range Signaling Mechanism. *J Mol Biol* **428**, 3702-3720, doi:10.1016/j.jmb.2016.04.012 (2016).
67 Gourinchas, G. *et al.* Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Sci Adv* **3**, e1602498, doi:10.1126/sciadv.1602498 (2017).
68 Edlund, P. *et al.* The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography. *Scientific Reports* **6**, 35279, doi:10.1038/srep35279 (2016).
69 Nicole C. Woitowich , A. S. H., Patricia Waltz, Christopher Kupitz, Joseph Varela, Gregory Tracy, Kevin D. Gallagher, Elin Claesson, Takanori Nakane, Suraj Pandey, Garrett Nelson, Rie Tanaka, Eriko Nango, Eiichi Mizohata, Shigeki Owada, Kensure Tono, Yasumasa Joti, Angela C. Nugent, Hardik Patel, Ayesha Mapara, James Hopkins, Phu Duong, Dorina Bizhga, Svetlana E. Kovaleva, Rachael St. Peter, Cynthia N. Hernandez, & Wesley B. Ozarowski, S. R.-C., Jay-How Yang, Petra Edlund, HeikkiTakala, Janne Ihalainen, Jennifer Scales, Tyler Norwood, Ishwor Poudyal, Petra Fromme., John Spence, Keith Moffat, Sebastian Westenhoff, Marius Schmidt, & Emina A. Stojković,*. Structural basis for light control of cell development revealed by crystal structures of a Myxobacterial phytochrome. (Submitted 2017).
70 Wagner, J. R. *et al.* Mutational analysis of Deinococcus radiodurans bacteriophytochrome reveals key amino acids necessary for the photochromicity and

proton exchange cycle of phytochromes. *Journal of Biological Chemistry* **283**, 12212-12226, doi:10.1074/jbc.M709355200 (2008).

71 Björling, A. *et al.* Structural photoactivation of a full-length bacterial phytochrome. *Science Advances* **2** (2016).

72 Kendrick, R. E. & Spruit, C. J. Phototransformations of phytochrome. *Photochem Photobiol* **26**, 201-214 (1977).

73 Borucki, B. *et al.* Light-induced proton release of phytochrome is coupled to the transient deprotonation of the tetrapyrrole chromophore. *Journal of Biological Chemistry* **280**, 34358-34364, doi:10.1074/jbc.M505493200 (2005).

74 Aramendia, P. F., Ruzsicska, B. P., Braslavsky, S. E. & Schaffner, K. Laser Flash-Photolysis of 124-Kilodalton Oat Phytochrome in H2o and D2o Solutions - Formation and Decay of the I700 Intermediates. *Biochemistry* **26**, 1418-1422, doi:DOI 10.1021/bi00379a031 (1987).

75 Matysik, J., Hildebrandt, P., Schlamann, W., Braslavsky, S. E. & Schaffner, K. Fourier-transform resonance Raman spectroscopy of intermediates of the phytochrome photocycle. *Biochemistry* **34**, 10497-10507 (1995).

76 Andel, F., Hasson, K. C., Gai, F., Anfinrud, P. A. & Mathies, R. A. Femtosecond time-resolved spectroscopy of the primary photochemistry of phytochrome. *Biospectroscopy* **3**, 421-433, doi:Doi 10.1002/(Sici)1520-6343(1997)3:6<421::Aid-Bspy1>3.0.Co;2-3 (1997).

77 Remberg, A. *et al.* Raman spectroscopic and light-induced kinetic characterization of a recombinant phytochrome of the cyanobacterium Synechocystis. *Biochemistry* **36**, 13389-13395, doi:10.1021/bi971563z (1997).

78 Foerstendorf, H. *et al.* FTIR studies of phytochrome photoreactions reveal the C=O bands of the chromophore: consequences for its protonation states, conformation, and protein interaction. *Biochemistry* **40**, 14952-14959 (2001).

79 Rudiger, W., Thummler, F., Cmiel, E. & Schneider, S. Chromophore Structure of the Physiologically Active Form (Pfr) of Phytochrome. *P Natl Acad Sci-Biol* **80**, 6244-6248, doi:DOI 10.1073/pnas.80.20.6244 (1983).

80 Andel, F., 3rd, Lagarias, J. C. & Mathies, R. A. Resonance raman analysis of chromophore structure in the lumi-R photoproduct of phytochrome. *Biochemistry* **35**, 15997-16008, doi:10.1021/bi962175k (1996).

81 van Thor, J. J., Ronayne, K. L. & Towrie, M. Formation of the early photoproduct Lumi-R of cyanobacterial phytochrome Cph1 observed by ultrafast mid-infrared spectroscopy. *J Am Chem Soc* **129**, 126-132, doi:10.1021/ja0660709 (2007).

82 Dasgupta, J., Frontiera, R. R., Taylor, K. C., Lagarias, J. C. & Mathies, R. A. Ultrafast excited-state isomerization in phytochrome revealed by femtosecond stimulated Raman spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1784-1789, doi:10.1073/pnas.0812056106 (2009).

83 van Thor, J. J. *et al.* Light-induced proton release and proton uptake reactions in the cyanobacterial phytochrome Cph1. *Biochemistry* **40**, 11460-11471 (2001).

84 Sineshchekov, V. A. Polymorphism of phytochrome a and its functional implications. *Light Sensing in Plants*, 95-102, doi:Doi 10.1007/4-431-27092-2_10 (2005).

85 Braslavsky, S. E., Gartner, W. & Schaffner, K. Phytochrome photoconversion. *Plant Cell Environ* **20**, 700-706, doi:DOI 10.1046/j.1365-3040.1997.d01-101.x (1997).

86 Hughes, J. Phytochrome three-dimensional structures and functions. *Biochem Soc Trans* **38**, 710-716, doi:10.1042/BST0380710 (2010).

87 Bjorling, A. *et al.* Ubiquitous Structural Signaling in Bacterial Phytochromes. *J Phys Chem Lett* **6**, 3379-3383, doi:10.1021/acs.jpclett.5b01629 (2015).

88 Berntsson, O. *et al.* Sequential conformational transitions and alpha-helical supercoiling regulate a sensor histidine kinase. *Nat Commun* **8**, 284, doi:10.1038/s41467-017-00300-5 (2017).
89 Gushchin, I. *et al.* Mechanism of transmembrane signaling by sensor histidine kinases. *Science* **356**, doi:10.1126/science.aah6345 (2017).
90 Casino, P., Rubio, V. & Marina, A. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**, 325-336, doi:10.1016/j.cell.2009.08.032 (2009).
91 Cox, M. M. & Battista, J. R. Deinococcus radiodurans - The consummate survivor. *Nature Reviews Microbiology* **3**, 882-892, doi:10.1038/nrmicro1264 (2005).
92 Battista, J. R. Against all odds: The survival strategies of Deinococcus radiodurans. *Annual Review of Microbiology* **51**, 203-224, doi:DOI 10.1146/annurev.micro.51.1.203 (1997).
93 Bhoo, S. H., Davis, S. J., Walker, J., Karniol, B. & Vierstra, R. D. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**, 776-779, doi:10.1038/414776a (2001).
94 Huntley, S. *et al.* Comparative Genomic Analysis of Fruiting Body Formation in Myxococcales. *Mol Biol Evol* **28**, 1083-1097, doi:10.1093/molbev/msq292 (2011).
95 Qualls, G. T., Stephens, K. & White, D. Morphogenetic Movements and Multicellular Development in Fruiting Myxobacterium, Stigmatella-Aurantiaca. *Dev Biol* **66**, 270-274, doi:Doi 10.1016/0012-1606(78)90291-9 (1978).
96 Mathes, T. *et al.* Femto- to Microsecond Photodynamics of an Unusual Bacteriophytochrome. *J Phys Chem Lett* **6**, 239-243, doi:10.1021/jz502408n (2015).
97 Stevens, R. C. Design of high-throughput methods of protein production for structural biology. *Structure* **8**, R177-185 (2000).
98 Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol Rev* **60**, 512-538 (1996).
99 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in microbial systems. *Front Microbiol* **5**, 341, doi:10.3389/fmicb.2014.00341 (2014).
100 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).
101 Amann, E. & Brosius, J. "ATG vectors' for regulated high-level expression of cloned genes in Escherichia coli. *Gene* **40**, 183-190 (1985).
102 Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry*. 7th edn, (W.H. Freeman, 2012).
103 Becker, W. M. *The world of the cell*. 7th edn, (Pearson/Benjamin Cummings, 2009).
104 Rontgen, W. C. On a New Kind of Rays. *Radiology* **45**, 428-435 (1945).
105 Rhodes, G. *Crystallography made crystal clear : a guide for users of macromolecular models*. 3rd edn, (Elsevier/Academic Press, 2006).
106 Russo Krauss, I., Merlino, A., Vergara, A. & Sica, F. An overview of biological macromolecule crystallization. *Int J Mol Sci* **14**, 11643-11691, doi:10.3390/ijms140611643 (2013).
107 McPherson, A. *et al.* The effects of microgravity on protein crystallization: evidence for concentration gradients around growing crystals. *J Cryst Growth* **196**, 572-586, doi:Doi 10.1016/S0022-0248(98)00853-7 (1999).
108 Leslie, A. G. W. & Powell, H. R. Processing diffraction data with MOSFLM. *Nato Sci Ser Ii-Math* **245**, 41-+ (2007).

109 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
110 Mccoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
111 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
112 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
113 Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195, doi:10.1107/S0907444904023510 (2004).
114 Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475 (1992).
115 Brunger, A. T. Free R value: cross-validation in crystallography. *Methods Enzymol* **277**, 366-396 (1997).
116 Madey, J. M. J. Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field. *J Appl Phys* **42**, 1906-&, doi:Doi 10.1063/1.1660466 (1971).
117 Pellegrini, C. & Reiche, S. The development of X-ray free-electron lasers. *Ieee J Sel Top Quant* **10**, 1393-1404, doi:10.1109/Jstqe.2004.838037 (2004).
118 Solem, J. C. Imaging Biological Specimens with High-Intensity Soft X-Rays. *J Opt Soc Am B* **3**, 1551-1565, doi:Doi 10.1364/Josab.3.001551 (1986).
119 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757, doi:10.1038/35021099 (2000).
120 Philipp, H. T., Hromalik, M., Tate, M., Koerner, L. & Gruner, S. M. Pixel array detector for X-ray free electron laser experiments. *Nucl Instrum Meth A* **649**, 67-69, doi:10.1016/j.nima.2010.11.189 (2011).
121 Philipp, H. T., Koerner, L. J., Hromalik, M., Tate, M. W. & Gruner, S. M. Pixel array detector for the capture of femtosecond duration x-ray images - art. no. 67030O. *P Soc Photo-Opt Ins* **6703**, O7030-O7030, doi:Doi 10.1117/12.733866 (2007).
122 DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J Phys D Appl Phys* **41**, doi:10.1088/0022-3727/41/19/195505 (2008).
123 Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Scientific Reports* **7**, doi:10.1038/srep44628 (2017).
124 Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Commun* **5**, 3309, doi:10.1038/ncomms4309 (2014).
125 Nogly, P. *et al.* Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *Iucrj* **2**, 168-176, doi:10.1107/S2052252514026487 (2015).
126 Sugahara, M. *et al.* Grease matrix as a versatile carrier of proteins for serial crystallography. *Nature Methods* **12**, 61-63, doi:10.1038/Nmeth.3172 (2015).
127 Botha, S. *et al.* Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr D* **71**, 387-397, doi:10.1107/S1399004714026327 (2015).
128 Liu, W., Ishchenko, A. & Cherezov, V. Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat Protoc* **9**, 2123-2134, doi:10.1038/nprot.2014.141 (2014).

129 Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Scientific Reports* **4**, doi:10.1038/srep06026 (2014).
130 Mueller, C. *et al.* Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct Dynam-Us* **2**, doi:10.1063/1.4928706 (2015).
131 Sierra, R. G. *et al.* Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nature Methods* **13**, 59-+, doi:10.1038/Nmeth.3667 (2016).
132 Mafune, F. *et al.* Microcrystal delivery by pulsed liquid droplet for serial femtosecond crystallography. *Acta Crystallogr D* **72**, 520-523, doi:10.1107/S2059798316001480 (2016).
133 Dods, R. *et al.* From Macrocrystals to Microcrystals: A Strategy for Membrane Protein Serial Crystallography. *Structure* **25**, 1461-1468 e1462, doi:10.1016/j.str.2017.07.002 (2017).
134 Lawrence, R. M. *et al.* Serial femtosecond X-ray diffraction of enveloped virus microcrystals. *Struct Dynam-Us* **2**, doi:10.1063/1.4929410 (2015).
135 Kupitz, C. *et al.* Microcrystallization techniques for serial femtosecond crystallography using photosystem II from Thermosynechococcus elongatus as a model system. *Philos T R Soc B* **369**, doi:10.1098/rstb.2013.0316 (2014).
136 Cherezov, V. Lipidic cubic phase technologies for membrane protein structural studies. *Curr Opin Struc Biol* **21**, 559-566, doi:10.1016/j.sbi.2011.06.007 (2011).
137 Johansson, L. C. *et al.* Lipidic phase membrane protein serial femtosecond crystallography. *Nat Methods* **9**, 263-265, doi:10.1038/nmeth.1867 (2012).
138 Andersson, R. *et al.* Serial femtosecond crystallography structure of cytochrome c oxidase at room temperature. *Sci Rep* **7**, 4518, doi:10.1038/s41598-017-04817-z (2017).
139 Koopmann, R. *et al.* In vivo protein crystallization opens new routes in structural biology. *Nature Methods* **9**, 259-U254, doi:10.1038/Nmeth.1859 (2012).
140 Koerner, L. J., Philipp, H. T., Hromalik, M. S., Tate, M. W. & Gruner, S. M. X-ray tests of a Pixel Array Detector for coherent x-ray imaging at the Linac Coherent Light Source. *J Instrum* **4**, doi:10.1088/1748-0221/4/03/P03001 (2009).
141 White, T. A. *et al.* Crystallographic data processing for free-electron laser sources. *Acta Crystallographica Section D-Biological Crystallography* **69**, 1231-1240, doi:10.1107/S0907444913013620 (2013).
142 Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J Appl Crystallogr* **47**, 1118-1131, doi:10.1107/S1600576714007626 (2014).
143 Kirian, R. A. *et al.* Femtosecond protein nanocrystallography-data analysis methods. *Opt Express* **18**, 5713-5723, doi:10.1364/Oe.18.005713 (2010).
144 Barends, T. R. M. *et al.* De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **505**, 244-+, doi:10.1038/nature12773 (2014).
145 Hajdu, J. & Johnson, L. N. Progress with Laue Diffraction Studies on Protein and Virus Crystals. *Biochemistry* **29**, 1669-1678, doi:DOI 10.1021/bi00459a001 (1990).
146 Harmand, M. *et al.* Achieving few-femtosecond time-sorting at hard X-ray free-electron lasers. *Nat Photonics* **7**, 215-218, doi:10.1038/Nphoton.2013.11 (2013).
147 Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242-1246, doi:10.1126/science.1259357 (2014).
148 Kupitz, C. *et al.* Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261-+, doi:10.1038/nature13453 (2014).

149 Arnlund, D. *et al.* Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat Methods* **11**, 923-926, doi:10.1038/nmeth.3067 (2014).
150 Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725-729, doi:10.1126/science.aad5081 (2016).
151 Barends, T. R. M. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445-450, doi:10.1126/science.aac5492 (2015).
152 Kupitz, C. *et al.* Structural enzymology using X-ray free electron lasers. *Struct Dyn* **4**, 044003, doi:10.1063/1.4972069 (2017).
153 Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* **26**, 795-800, doi:Doi 10.1107/S0021889893005588 (1993).
154 Bhattacharya, S., Auldridge, M. E., Lehtivuori, H., Ihalainen, J. A. & Forest, K. T. Origins of fluorescence in evolved bacteriophytochromes. *J Biol Chem* **289**, 32144-32152, doi:10.1074/jbc.M114.589739 (2014).
155 Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D-Biological Crystallography* **60**, 432-438, doi:10.1107/S0907444903028956 (2004).
156 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).
157 Bhate, M. P., Molnar, K. S., Goulian, M. & DeGrado, W. F. Signal transduction in histidine kinases: insights from new structures. *Structure* **23**, 981-994, doi:10.1016/j.str.2015.04.002 (2015).
158 Escobar, F. V. *et al.* Protonation-Dependent Structural Heterogeneity in the Chromophore Binding Site of Cyanobacterial Phytochrome Cph1. *Journal of Physical Chemistry B* **121**, 47-57, doi:10.1021/acs.jpcb.6b09600 (2017).
159 Toh, K. C. *et al.* Primary reactions of bacteriophytochrome observed with ultrafast mid-infrared spectroscopy. *J Phys Chem A* **115**, 3778-3786, doi:10.1021/jp106891x (2011).
160 Toh, K. C., Stojkovic, E., van Stokkum, I. H. M., Moffat, K. & Kennis, J. T. Proton Transfer and Hydrogen-Bond Interactions Determine the Fluorescence Quantum Yield of Bacteriophytochrome, a Novel Deep-Tissue Fluorescent Probe. *Biophys J* **98**, 413a-413a, doi:DOI 10.1016/j.bpj.2009.12.2227 (2010).
161 Toh, K. C., Stojkovic, E. A., van Stokkum, I. H., Moffat, K. & Kennis, J. T. Proton-transfer and hydrogen-bond interactions determine fluorescence quantum yield and photochemical efficiency of bacteriophytochrome. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9170-9175, doi:10.1073/pnas.0911535107 (2010).
162 Escobar, F. V., von Stetten. D., Gunther-Lutkens. M., Keidel A., Michael N., Lamparter., T. Essen, L-O., Hughes J., Gärtner W., Yang Y., Heyne K., Mroginski M.A., and Hilderbrandt P Conformational heterogeneity of the Pfr chromphore in plant and cyanobacterial phytochromes. *Frontier in Molecular Biosciences*, doi:10.3389/fmolb.2015.00037 (2015).
163 Stojkovic, E. A. *et al.* FTIR Spectroscopy Revealing Light-Dependent Refolding of the Conserved Tongue Region of Bacteriophytochrome. *J Phys Chem Lett* **5**, 2512-2515, doi:10.1021/jz501189t (2014).
164 van Thor, J. J., Fisher, N. & Rich, P. R. Assignments of the Pfr-Pr FTIR difference spectrum of cyanobacterial phytochrome Cph1 using (15)N and (13)C isotopically labeled phycocyanobilin chromophore. *Journal of Physical Chemistry B* **109**, 20597-20604, doi:10.1021/jp052323t (2005).

165 Tono, K. *et al.* Beamline, experimental stations and photon beam diagnostics for the hard x-ray free electron laser of SACLA. *New J Phys* **15**, doi:10.1088/1367-2630/15/8/083035 (2013).
166 Kameshima, T. *et al.* Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev Sci Instrum* **85**, doi:10.1063/1.4867668 (2014).
167 Nakane, T. *et al.* Data processing pipeline for serial femtosecond crystallography at SACLA. *J Appl Crystallogr* **49**, 1035-1041, doi:10.1107/S1600576716005720 (2016).
168 Boutet, S. & Williams, G. J. The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New J Phys* **12**, doi:10.1088/1367-2630/12/3/035024 (2010).
169 Yang, Y. *et al.* Real-Time Tracking of Phytochrome's Orientational Changes During Pr Photoisomerization. *J Am Chem Soc* **134**, 1408-1411, doi:10.1021/ja209413d (2012).