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

Protein Structural Dynamics Revealed by Time-Resolved X-ray Solution Scattering

Léocadie Henry

UNIVERSITY OF GOTHENBURG
Department of Chemistry and Molecular Biology
Gothenburg, 2020

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

Protein Structural Dynamics Revealed by Time-Resolved X-ray Solution Scattering

Léocadie Henry

Cover: Dynamic clock of apomyoglobin unfolding unveiled by X-ray solution scattering.

ISBN 978-91-7833-816-0 (Print)
ISBN 978-91-7833-817-7 (PDF)
Available online at http://hdl.handle.net/2077/63085

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, 2020

Abstract

Proteins fold in very specific three-dimensional structures within a few seconds. Highly complex and organized, their three-dimensional structure is closely related and often essential for their function. They perform a vast array of vital functions within organisms. Myoglobin, for example, stores dioxygen within muscles, while photoreceptors, such as phototropins or cryptochromes, respond to their light-environment to regulate growth, metabolism or the circadian clock. Protein aggregation, misfolding and unfolding have been associated with many pandemic pathologies. Therefore, a crucial need for biophysical techniques investigating protein structure and dynamics in real-time, has emerged over the past decades. Time-resolved X-ray scattering in solution has the advantage of being directly sensitive to structural changes happening on broad time scales from the atomic level to the general shape of a protein.

In this thesis, time resolved X-ray solution scattering has been applied together with molecular dynamics simulations to address various biochemical questions. These techniques were used to bring new insights in the folding process of small globular proteins at the atomic level. By following the conformational transitions of apomyoglobin unfolding, we showed that globular proteins fold in defined pathways starting with the hydrophobic collapse of non-conserved residues. Time-resolved X-ray scattering has also been applied to establish a comprehensive mechanism of signal transduction in the light-oxygen-voltage domain of two blue-light photoreceptors: Ytva and phototropin. The data showed that the mechanism of signal transduction is conserved in photosensory domains. On a greater length-scale, time-resolved solution scattering gave a detailed picture of the signal relay in cryptochromes. The results provided a structural link between the conformational changes in cryptochrome, and the interaction and regulation with its downstream partners.

This work demonstrated the potential and various applications of time-resolved X-ray solution scattering. This technique allows the investigation of protein structure and dynamics on different length-scales, with high-temporal resolution.

Sammanfattning på Svenska

Proteiner veckar sig i mycket specifika tredimensionella strukturer inom några få sekunder. Dessa strukturer är mycket komplexa, organiserade och ofta väsentliga för proteiners funktioner. Proteiner utför ett stort antal viktiga funktioner inom organismer. Myoglobin, till exempel, lagrar dioxygen i muskler, medan fotoreceptorer, såsom fototropiner eller kryptokromer, svarar på deras ljusmiljö för att reglera tillväxt, ämnesomsättning eller dygnsrytmen. Proteinaggregering, felveckning och att proteiner vecklar ut sig har kopplats ihop med många pandemiska patologier. Därför har ett avgörande behov av biofysiska tekniker, som undersöker proteinstruktur och dynamik i realtid, utvecklats under de senaste decennierna. Tidsupplöst röntgendiffraktion i lösning har fördelen av att vara direkt känslig för strukturella förändringar som sker på breda tidsskalor från atomnivån till den generella strukturen för ett protein. I denna avhandling har tidsupplöst röntgendiffraktion tillämpats tillsammans med molekylära dynamikssimuleringar för att adressera olika biokemiska frågor. Dessa tekniker användes för att få nya insikter i strukturveckningsprocessen för små globulära proteiner på atomnivå. Genom att följa övergångarna mellan olika konformationer av apomyoglobin under dess utveckling, visade vi att globulära proteiner veckas i definierade vägar som börjar med den hydrofoba kollapsen av icke-konserverade aminosyror. Tidsupplöst röntgendiffraktion har också använts för att upprätta en omfattande mekanism för signaltransduktion i ljus-syre-spänningsdomänen för två blåljuskänsliga fotoreceptorer: Ytva och fototropin. Resultaten visade att mekanismen för signaltransduktion bevaras i fotosensoriska domäner. Sett i ett större längdskalor gav tidsupplöst röntgendiffraktion en detaljerad bild av signalreläet i kryptokromer. Resultaten gav en strukturell koppling mellan de konformationella förändringarna i kryptokromer, och interaktionen och regleringen med dess nedströmspartners. Med detta arbete demonstrerade vi de potentiella och olika tillämpningarna av tidsupplöst röntgendiffraktion. Denna teknik möjliggör undersökning av proteinstruktur och dynamik på olika längdskalor, med hög upplösning.

Publications

This thesis consists of the following research papers:

PAPER I: **Léocadie Henry**, Matthijs Panman, Linnéa Isaksson, Elin Claesson, Irena Kosheleva, Robert Henning, Sebastian Westenhoff, and Oskar Berntsson.
"Tracking the unfolding of apomyoglobin with time-resolved X-ray solution scattering", Submitted (2019)

PAPER II: **Léocadie Henry**, Oskar Berntsson, Matthijs Panman, Ashley Hughes, Stephan Niebling, Irena Kosheleva, Robert Henning, Andreas Menzel, Sebastian Westenhoff.
"New light on the mechanism of phototransduction in the LOV1 domain from Chlamydomonas reinhardtii phototropin", Manuscript (2020)

PAPER III: Oskar Berntsson, Ralph Diensthuber, Matthijs Panman, Alexander Björling, Ashley Hughes, **Léocadie Henry**, Stephan Niebling, Gemma Newby, Marianne Liebi, Andreas Menzel, Robert Henning, Irena Kosheleva, Andreas Möglich and Sebastian Westenhoff.
"Time-resolved X-ray solution scattering reveals the structural photoactivation of a Light-Oxygen-Voltage photoreceptor" Structure. (2017), doi: 10.1016/j.str.2017.04.006

PAPER IV: Oskar Berntsson*, Ryan Rodriguez*, **Léocadie Henry**, Matthijs Panman, Ashley Hughes, Christopher Einholz, Stefan Weber, Janne Ihalainen, Robert Henning, Irena Kosheleva, Erik Schleicher, Sebastian Westenhoff.
"Photoactivation of Drosophila melanogaster cryptochrome through sequential conformational transitions" Sci Adv. (2019), doi: 10.1126/sciadv.aaw1531

Related papers that I have co-authored but that are not included in this thesis:

PAPER VI: Oskar Berntsson*, Ralph Diensthuber*, Matthijs Panman, Alexander Björling, Emil Gustavsson, Maria Hoernke, Ashley Hughes, **Léocadie Henry**, Stephan Niebling, Heikki Takala, Janne Ihalainen, Gemma Newby, Silke Kerruth, Joachim Heberle, Marianne Liebi, Andreas Menzel, Robert Henning, Irina Kosheleva, Andreas Möglich, and Sebastian Westenhoff.
"Sequential conformational transitions and α-helical supercoiling regulate a sensor histidine kinase." Nat Commun. (2017) doi: 10.1038/s41467-017-00300-5

PAPER VII: 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 Bth, Vaibhav Modi, Shatabdi Roy-Chowdhury, James Zook, Peter Berntsen, Suraj Pandey, Ishwor Poudyal, Jason Tenboer, Christopher Kupitz, Anton Barty, Petra Fromme, Jake Koralek, Tomoyuki Tanaka, John Spence, Mengning Liang, Mark Hunter, Sebastien Boutet, Eriko Nango, Keith Moffat, Gerrit Groenhof, Janne Ihalainen, Emina Stojković, Marius Schmidt, and Sebastian Westenhoff.
"The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography." Sci Rep. (2016) doi: 10.1038/srep35279

PAPER VIII: Alexander Björling*, Oskar Berntsson*, Heli Lehtivuori, Heikki Takala, Ashley Hughes, Matthijs Panman, Maria Hoernke, Stephan Niebling, **Léocadie Henry**, Robert Henning, Irina Kosheleva, Vladimir Chukharev, Nikolai Tkachenko, Andreas Menzel, Gemma Newby, Dmitry Khakhulin, Michael Wulff, Janne Ihalainen, and Sebastian Westenhoff.
"Structural photoactivation of a full-length bacterial phytochrome." Sci Adv. (2016) doi: 10.1126/sciadv.1600920

Contribution report

PAPER I: I took part in the planning and execution of the X-ray solution scattering experiments. I performed the structural analysis of the X-ray scattering data and wrote the paper.

PAPER II: I designed the research together with my supervisor. I developed, and carried out, expression and purification protocols of the protein. I took part in the planning and execution of the X-ray solution scattering experiments. I performed the structural analysis of the X-ray scattering data and wrote the manuscript.

PAPER III: I participated in the planning and execution of the X-ray solution scattering experiments. I prepared the samples and collected data.

PAPER IV: I was involved in the planning and execution of the X-ray solution scattering experiments. I also prepared the samples and collected data.

Abbreviations

Here follows a list and short explanation of the different abbreviations used in this thesis.

apoMb	**apo**Myoglo**b**in (myoglobin without the heme group)
APS	**A**dvanced **P**hoton **S**ource (synchrotron radiation facility outside Chicago)
BioCARS	**Bio** **C**enter for **A**dvanced **R**adiation **S**ources (beamline at the APS)
***Bs*YtvA**	***B**acillus **S**ubtilis* **Y**tv**A** (photoreceptor protein)
CCD	**C**harge **C**oupled **D**evice (type of detector)
***Cr*Phot**	***C**hlamydomonas **r**einhardtii* **P**hototropin (photoreceptor protein)
cSAXS	**c**oherent **S**mall **A**ngle **X**-ray **S**cattering (beamline at the SLS)
CTT	**C**arboxy **T**erminal **T**ail (small part of *Dm*Cry)
DASH	***D**rosophila **A**rabidopsis **S**ynechocystis Homo* (Subfamily of cryptochromes)
D_{max}	**max**imal **D**imension
***Dm*Cry**	***D**rosophila **m**elanogaster* **Cry**ptochrome (photoreceptor protein)
EPR	**E**lectron **P**aramagnetic **R**esonance (experimental technique)
ESRF	**E**uropean **S**ynchrotron **R**aditation **F**acility (synchrotron radiation facility outside Grenoble)
FAD	**F**lavin **A**denine **D**inucleotide (chromophore)
FMN	**F**lavin **M**ono**N**ucleotide (chromophore)
ID09b	**I**nsertion **D**evice **09b** (beamline at the ESRF)
IR	**I**nfra-**R**ed
LOV	**L**ight **O**xygen **V**oltage (Photoreceptor protein domain)
MD	**M**olecular **D**ynamics (Computer simulations of protein dynamic)
MW	**M**olecular **W**eight
NMR	**N**uclear **M**agnetic **R**esonance

PLUMED	**PLU**gin for **M**ol**E**cular **D**ynamics (Computer tools for protein dynamic study)
Rg	**R**adius of **G**yration
SAXS	**S**mall **A**ngle **X**-ray **S**olution **S**cattering
T-Jump	**T**emperature-**J**ump
TRXSS	**T**ime-**R**essolved **X**-ray **S**olution **S**cattering
UV	**U**ltra**V**iolet
WAXS	**W**ide **A**ngle **X**-ray **S**olution **S**cattering

Contents

Chapter 1

Introduction

At the core of any biological process lie proteins. Composed of amino-acids, proteins are essential biological macro-molecules. Proteins perform a large variety of vital functions such as sensory perception, evolutionary selection, immune defense or energy transformations. To ensure their functionality, many proteins spontaneously fold into a well-defined three-dimensional structure, dictated by their primary sequence. However, the question arises as to how do protein fold into an organized structure on a biologically relevant time scale, when billions of possible conformations are available. As of today, solving this question still remains one the biggest challenges in structural biology and is known as "the protein folding problem". Over the past 20 years, cancer and neuro-degenerative diseases, such as Alzheimer or Parkinson, have been associated with protein misfolding and aggregation. These discoveries highlight the vital importance of understanding protein structure and their folding mechanism. The determination of the three-dimensional structure of a protein can now be resolved at very high resolution by various biophysical techniques. However, the folding and conformational rearrangements of proteins happen on a time-scale hardly reachable by those methods. New approaches need to be developed in order to study the relationship between the structure and dynamics of a protein.

1.1 How do proteins fold?

A protein structure is defined at four levels. The primary structure (Figure 1.1 A) is the unique sequence of the twenty naturally occurring

amino-acids, composing a protein. This polypeptide chain associates in secondary structure motifs, sheets or helices for example (Figure 1.1 B), which pack together in a highly organized three-dimensional structure. The overall folding and shape of the protein characterizes the tertiary structure (Figure 1.1 C), essential for its specific function. The secondary structure is primarily stabilized by hydrogen bonds between the amide groups of amino acids that are close in sequence [1,2], whereas the tertiary structure is maintained by hydrophobic interactions among side-chains of more distant parts of the chain [3]. One or several protein tertiary sub-units can assemble in a multi-domain structure forming a quaternary structure (Figure 1.1 D).

Proteins spontaneously fold into their specific three-dimensional native structure after they emerge from the ribosome. Cyrus Levinthal demonstrated in 1968 [4] that proteins can fold in an astronomical number of possible conformations due to the very large degree of freedom in the unfolded state. Additionally, the number of conformations increases exponentially with the chain-length. Anfinsen also pointed out in the early 60's that a protein refolds to its native state when the optimal environmental conditions (such as pH, solvent or ionic strength) are restored [5]. The question arises as to how a protein spontaneously folds into its defined and highly organized three-dimensional structure. If a protein were in fact to randomly explore every different conformation to attain its correctly folded state, it would require more than the age of the universe. However, the folding process of proteins naturally occurs in less than a few seconds [6]. Furthermore, folding converges to a unique and conformationally stable state, at least for small globular proteins, in which the Gibbs free energy of the entire system is the lowest [7]. This is defined as the "Levinthal paradox" and remains, to this day, a big challenge in structural biology. In the last 50 years, tremendous efforts have been carried out to understand the mechanism of protein folding and different models have been proposed.

The "classical view" suggests that, since proteins cannot sample the millions of possible conformations within observed folding times, they must fold in predetermined pathways [4, 7]. Therefore, there must exist a high conformational bias in the protein primary sequence, to accelerate

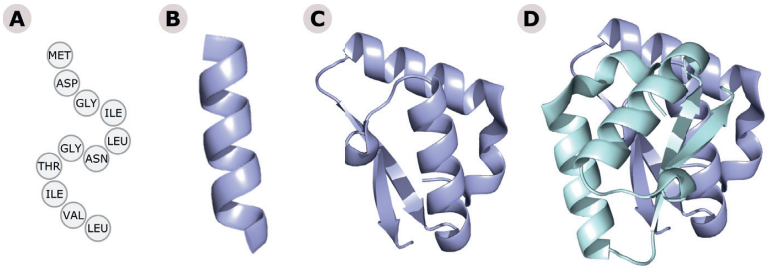

Figure 1.1: Overview of the four levels of organization of a protein structure. The primary sequence, a unique combination of amino-acids, is presented in A. The polypeptide backbone of a protein forms secondary structural motifs (B), that pack together in a highly organized three-dimensional structure (C). Several subunits of a protein can assemble in a quaternary structure (D).

the folding process (Figure 1.2 A). This theory describes the folding pathway as a predefined succession of events, from an unfolded to a native state passing by a few obligatory intermediate structures.

Along the technical advances, a "new view" of protein folding arose in the 1990's. Different partially folded forms were observed and it was suggested that proteins fold into their native states by following multiple unpredictable routes instead. This theory is known as the funnel shape energy landscape [8, 9]. As opposed to a well-defined sequence of events, the folding funnel hypothesis assumes that proteins essentially fold downhill in an unpredictable succession of trajectories towards a global minimum. Conformational sub-ensembles can thus coexist (Figure 1.2 B).

More recently, it has been proposed that proteins are built from separately cooperative foldon units [10–13]. Foldons are defined as a strip of polypeptide that folds cooperatively [10] and contains about 25 residues. Foldons would collapse together in the initial stage of folding, forming a native-like structural nucleus lowering the folding energy barrier (Figure 1.2 C). The protein would then fold in its native state, as soon as enough unfolded segments simultaneously adopt the native structure. This model is very similar to the "folding upon binding" theory of disordered proteins [14]. This current view partially solves the Levinthal Paradox since

the initial random search would not be needed for the entire protein but only for the formation of native-like foldons instead. This guided approach would therefore prominently accelerate the search for the native conformation.

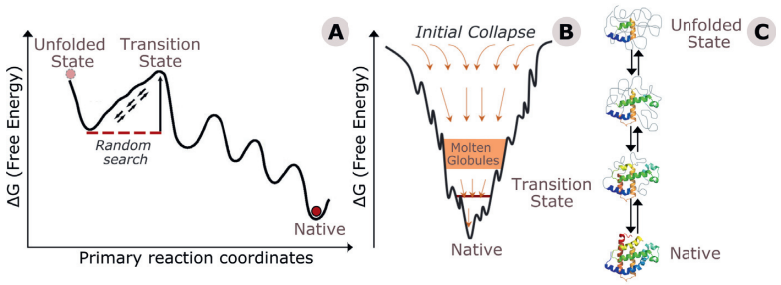

Figure 1.2: The classical (A) view of a defined folding pathway and the funnel energy landscape (B) approach through multiple routes. The new view of protein folding by foldon association is depicted in (C).

1.2 Protein Dynamics

The overall protein structure is defined by its primary, secondary, tertiary and quaternary structure. As trivial as this may sounds, virtually nothing was known of the three-dimensional structure of proteins until a few decades ago. The first structure to be solved was for sperm whale myoglobin in 1958 [15]. Since then,over 150,000 protein structures have been described at the atomic level, largely by X-ray crystallography [16].

X-ray crystallography records the diffraction of electromagnetic radiation passing through repeated objects, or proteins in this case [17]. For this purpose, proteins need to be periodically aligned and oriented in a crystal. The obtained diffraction pattern, consisting of constructively interfering scattered waves from the proteins, is a finger print of the protein structure [18]. However, it is important to keep in mind that proteins are not static objects but rather dynamic entities with internal motions. Proteins can adopt different conformational states, both at

equilibrium and under perturbation, which play an essential role for their functions [19, 20]. A full understanding of the protein structure-function relationship demands, therefore, an analysis of their dynamic behavior in addition to their static picture.

Going beyond the static structure of a single protein has proven to be difficult. Tremendous efforts have been made over the years to develop techniques to study those motions. Among these, nuclear magnetic resonance (NMR) is a powerful technique where NMR relaxation, as a consequence of local fluctuating magnetic fields within a molecule, provides information on the dynamics of a protein [21, 22]. However, as larger molecules tumble slower, traditional NMR is currently limited to protein under ca. 35 kDa [23]. Another powerful technique to study dynamics, is absorption spectroscopy (infrared and visible), which gives information about electronic and vibration transitions [24, 25]. As crucial as they are to understand the function of proteins, these spectroscopic techniques generally do not directly relate to the protein structure. Time-resolved X-ray crystallography has developed within the past decades providing information about internal flexibility [26, 27], but still requires the formation of crystals and very powerful X-ray sources (X-ray Free-Electron Lasers or XFEL). Consequently, computational solutions for the prediction of protein dynamics have emerged. Molecular dynamics (MD) simulations currently provide dynamical information with high spatial resolution [28]. MD simulations also produce details of protein motions on time scales that are difficult to access with experimental techniques. Nevertheless, MD simulations still remain computationally intensive and need to be combined with other techniques to enhance conformational sampling of large molecules.

X-ray solution scattering measures the intensity of the scattered electrons of a sample, in solution [29]. As opposed to X-ray crystallography, it does not require the formation of crystals and can be performed under physiological conditions [30]. Furthermore, the scattering signal provides direct information on the three-dimensional shape and conformation of the protein [31, 32]. X-ray solution scattering can be used in real time,

on time scales hardly accessible by other techniques, from nano- to seconds. Time-resolved X-ray solution scattering (TRXSS) can, therefore, detect global conformational rearrangements, secondary-structure movements and changes in the hydration layer of a protein over time [33, 34]. Unfortunately, TRXSS retrieves only low-resolution information. Combined with MD simulations, these methods are particularly suited to study the kinetics of proteins in solution (Chapter 3), ultimately providing further understanding on how a protein structure relates to its function [35–37].

1.3 An example of structure-function relationship: Blue-light photoreceptors

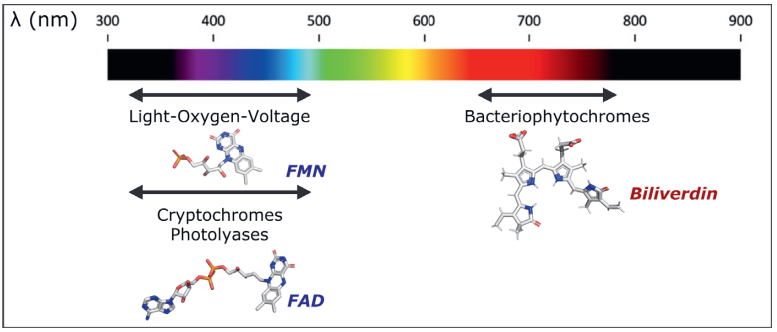

Figure 1.3: Examples of photoreceptor spectral ranges. Depending on the enclosed chromophore moiety, photoreceptor proteins absorb light ranging from ultra-violet (UV) to infra-red (IR) wavelengths.

The dynamic changes in a protein structure are closely related to its function. Photoreceptors, or photosensors, are a great example of such structure-dynamic relationship. Photoreceptors are essential proteins, which can sense and respond to environmental light cues [38, 39]. Crucial in many organisms, they are for instance responsible of shade avoidance and phototropism in plants [40, 41], or regulation of the circadian-clock and vision in animals [38]. To be able to react to light, photoreceptors contain, in addition to the protein moiety, a photopigment called chromophore. These non-proteinic entities can absorb visible light,

ranging from near-UV to IR wavelengths (Figure 1.3). Examples of chromophores are flavins, absorbing blue-light, and bilins that absorb far-red-light. Upon the absorption of a photon, delocalized electrons along a conjugating π system, drive the system into an excited state [40] and provide the molecule with absorption in the visible region. The change in chromophore configuration, alters its vicinal environment, by bond-isomerization or -reduction with the protein. Light-absorption by the chromophore induces changes in the protein conformation enabling the interaction with downstream partners.

By their nature, photoreceptors are particularly suited to study the protein structure-function relationship. Biological signaling starts with the association of a ligand with an activated receptor protein. Triggered by a laser, of relevant wavelength, the photochemistry and mechanism of such an interaction in photoreceptors, can thus easily be explored by a wide range of biophysical methods. An example is the study of the signal relay in blue-light photoreceptors by TRXSS [42–44]. As many photoreceptors are modular and conserved in nature, they can be great tools and have several promising applications. Photoreceptor engineering can be an asset in optogenetics for example, the manipulation of cellular events by light [45–47].

1.4 Scope of this thesis

This thesis focuses on the various applications of time-resolved X-ray solution scattering to understand the true essence of proteins. It is organized into two parts. The first sections (Chapters 2 and 3) describe the principle of this method and the second parts (Chapters 4 to 6) present TRXSS applications to gain mechanistic insight into the protein structure-function relationship.

Chapter 2 introduces the basics of X-ray scattering in solution and explains how it can be applied to protein structure prediction.

Chapter 3 describes how a time-resolved X-ray solution scattering is conducted. It also provides a guide on how scattering data can be processed and analyzed to retrieve high resolution protein structural information.

Chapter 4 shows how time-resolved X-ray solution scattering can help solving the mystery of protein folding. **Paper I** provides new insights into the folding mechanism of the well-studied model: apomyoglobin.

Chapter 5 and 6 focus on the structural and dynamics information that time-resolved X-ray solution scattering can provide for different blue-light sensitive proteins. **Paper II** and **Paper III** describe the photochemistry and signal transduction in Light-Oxygen-Voltage photoreceptors while **Paper IV** aims to Cryptochromes.

Finally, **Chapter 7** summarises, concludes and gives future outlook of the work presented in this thesis.

Chapter 2

X-ray Scattering

With a regular confocal microscope, we can only resolve objects larger than a micrometer. However, the wavelength of the light we use, must be no larger than the molecule we want to analyze [48]. To observe atoms and chemical bonds of ca. 1 Å, we must therefore use X-rays, which are electromagnetic waves possessing a wavelength ranging from 0.1 Å to 100 Å. In section 2.1 of this Chapter, I will introduce the principle and fundamentals of X-ray scattering. I will then discuss, in section 2.2, how small-angle X-ray scattering can be used in solution to study proteins.

2.1 The principle of X-ray scattering

Elastic scattering

When a beam of X-ray photons strikes a molecule, its electrons start to oscillate, emitting (diffracting) a wave with an amplitude given by fundamental constants [48]. This phenomenon is known as elastic scattering whereby the diffracted wave recorded by a detector has the same energy as the incident photon.

Let's consider two points A and B (Figure 2.1). The amplitude measured at a point of observation P, is the sum of the waves emitted from A and B. Each wave has distinct phases as they travel a different distance from the plane front to the point of observation P. The scattered waves emitted from A and B add and interfere.

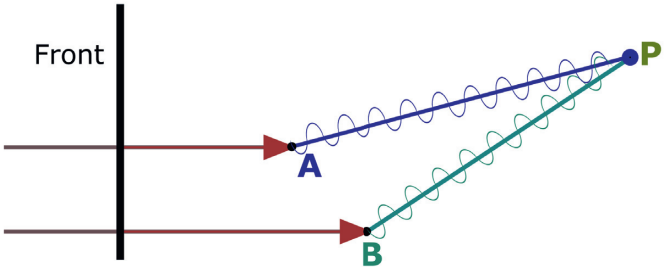

Figure 2.1: Scattering from two points.

Scattering of a group of electrons.

If we consider a group of electrons positioned at O (Figure 2.2), the distance between the sample and the detector is much greater than the distances within the sample. The elastic scattering can then be described as an incident plane wave $\vec{k_0}$, with $|\vec{k_0}| = \frac{2\pi}{\lambda}$, scattered in the direction of $\vec{k}$. The difference of $\vec{k} - \vec{k_0}$ defines $\vec{q}$, the momentum transfer. The modulus of $\vec{q}$ is equal to $\frac{4\pi}{\lambda} sin(\theta)$, where λ is the incident X-ray wavelength and 2θ is the scattering angle. With all the phases taken into account, the sum of the scattered waves for N electrons placed at a distance r_n from an atom centered at O (Figure 2.2), is described by equation 2.1.

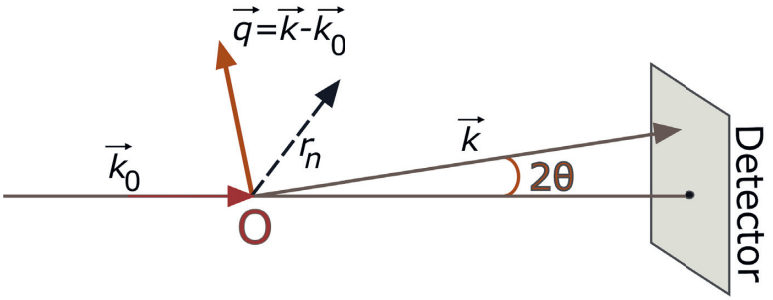

Figure 2.2: Scattering by an assembly of electrons, or a single molecule.

$$F(\mathbf{q}) = \sum_{n=1}^{N} f_e e^{i\mathbf{q}\cdot\mathbf{r_n}} \tag{2.1}$$

Scattering in solution.

During X-ray solution scattering experiments, one measures the scattering of a large number of identical molecules (Figure 2.3). Since the distance between two molecules is not fixed in solution, the intensities are measured instead of amplitudes. The atomic scattering factor $f(q)$ is a measure of the scattering amplitude of a wave by an isolated atom. For an object with a spatial density distribution $\rho(r)$, $f(q)$ is given by Equation 2.2.

$$f(q) = 4\pi \int \rho(r) r^2 \frac{sin(qr)}{qr} dr \tag{2.2}$$

The scattering factor per electron f_e, from Equation 2.1, can then be exchanged by the atomic scattering factor f_q, from Equation 2.2, to calculate the Fourier transform, $F(\vec{q})$. The intensities are obtained by multiplying $F(\vec{q})$, of the electron density $\rho(r)$, describing the scattering object, by its complex conjugate (Equation 2.3).

$$I(\vec{q}) = F(\vec{q})F(\vec{q})^* \tag{2.3}$$

As the molecules tumble freely in solution, the scattering of all their possible orientations is simultaneously recorded (Figure 2.3). Consequently, the detector image in a solution X-ray scattering experiment will show a spherical, uniform and essentially one-dimensional pattern; as opposed to a clear single-crystal X-ray diffraction pattern, when all the molecules are well ordered in a crystal.

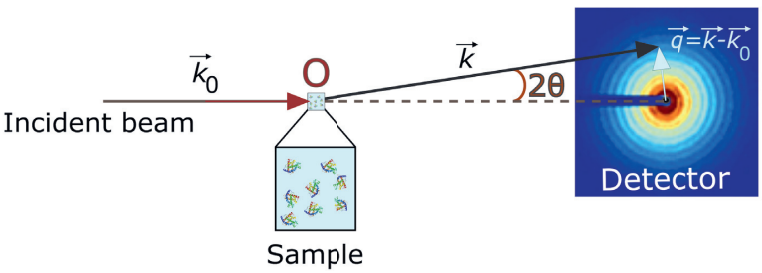

Figure 2.3: Scattering from an ensemble of molecules randomly oriented in solution. The detector image shows a spherical averaging of the intensities.

In solution, particles are more disordered than in solids. Therefore, all possible positions and orientations are simultaneously recorded by the detector. A spherical averaging can be performed to obtain the one-dimensional intensity, leading to the Debye formula (Equation 2.4).

$$I(q) = \sum_m \sum_n f_m(q) f_n(q) \frac{sin(qr_{mn})}{qr_{mn}} \tag{2.4}$$

where f_m and f_n are the atomic scattering factors of groups of atoms located at r_m and r_n, respectively, and $r_{mn} = |r_m - r_n|$.

Equation 2.5 defines the relationship between real (r) and reciprocal (q) space distances.

$$r = \frac{2\pi}{q} \tag{2.5}$$

As a matter of fact, the Debye formula (Equation 2.4) gives a positive contribution for $q < 2\pi/r$ (Figure 2.4). The $sin(qr)/(qr)$ term decays for higher q values, to finally oscillate around 0. This means that long distances in real space, will contribute more to the scattering at lower angles in reciprocal space while, short distances in real space will contribute to high-angle scattering. Two experimental ranges of distances are then defined in solution scattering experiments: small-angles (SAXS), for typically q<0.2 Å^{-1} and wide angles (WAXS), for q>0.2 Å^{-1}.

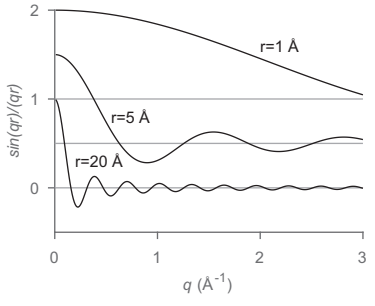

Figure 2.4: The q dependence of the Debye term for three different distances. The three curves are offset with respect to each other for better visibility.

One drawback of X-ray scattering, is the ambiguity of the resulting data. As we have seen, only the scattering intensities are collected and not the phases of the photons. In turn, this prevents the full retrieval of the original structure and is commonly known as the "phase problem" [17, 49]. Despite this drawback, SAXS is a powerful method to probe protein structures on multiple length scales [50]. SAXS can provide information on interactions between protein molecules, on the global shape of the protein and on the tertiary and secondary structure of the protein [51–54]. To gain further atomic details, SAXS can be combined with other techniques such as X-ray crystallography or MD simulations [30].

2.2 Small-Angle X-ray solution scattering

Running an experiment

When running a SAXS experiment, the sample of interest must be monodisperse, meaning that all the molecules must be of the same shape and size. In addition, the sample is considered ideal when inter-particle interactions are not predominant. As shown in Section 2.1, the intensities of the scattered waves, as a function of q, can be collected from the detector image. In order to obtain the scattering of the protein of interest, the contribution of the solvent must be subtracted first. To do so, the scattering of the solvent alone before and after each sample measurement is recorded. The solvent scattering comprises the scattering contribution of the bulk solvent, but also the hydration layer, which is the solvent layer close to the solute. The hydration layer possesses a different density compared to that of the bulk solvent and must therefore be treated differently. The scattering of the protein only can be derived from Equation 2.6.

$$\begin{aligned} I_{Protein} &= I_{Sample} - I_{Solvent} \\ &= \langle (F_{vac.} - F_{excl.} + F_{hyd.})(F_{vac.} - F_{excl.} + F_{hyd.})^* \end{aligned} \tag{2.6}$$

The terms in parenthesis represent the average over all orientation of the scattering amplitudes from the solute scattering in vacuum, $F_{vac.}$, from the bulk solvent, $F_{excl.}$ and from hydration layer, $F_{hyd.}$.

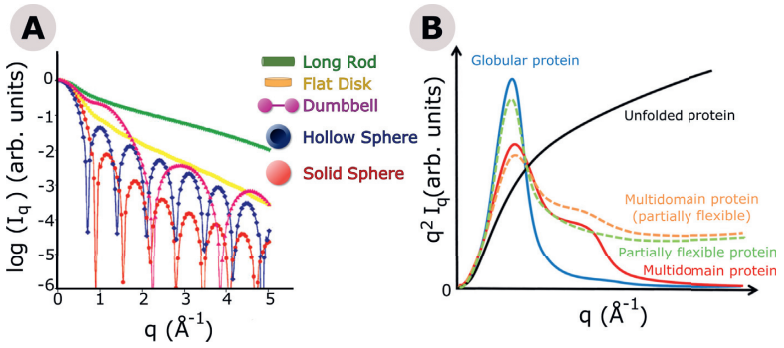

Figure 2.5: The scattering intensities (logarithmic scale), in panel A, provide information on the particle shape without previous knowledge. The Kratky plot (B) can identify flexibility of a protein and especially unfolding. Figure adapted from [55].

Extraction of structural parameters

Without any prior knowledge of the protein of interest, shape information can be directly obtained from the intensities (Figure 2.5 A). In addition, the flexibility and folded state of a protein can be assessed by analyzing the Kratky plot (Figure 2.5 B) given by plotting $q^2 * I(q)$ as a function of q.

The scattering intensity of a protein is described by a Gaussian curve in the vicinity of the origin. The radius of gyration (R_g), defined as the distribution of atoms of a particle around its axis [56], can be directly extracted from the experimental data by the Guinier law given by Equation 2.7.

$$ln[I(q)] \cong ln[I(0)] - \frac{R_g^2}{3} q^2 \tag{2.7}$$

where $I(0)$ is the scattering intensity at zero angle (q=0). The Guinier approximation is valid for $q \leq 1.3\ R_g$.

Similarly, the molecular weight (MW) of a particle can be estimated by different procedures. Equation 2.8 is an example of MW estimation of a molecule, independent of its concentration [57], where Q_R, is the ratio of the square of the Volume of correlation (V_c) to R_g in Å^3, e is the Euler

number and the parameters k and c are empirically determined.

$$MW = \left(\frac{Q_R}{e^c} \right)^{\frac{1}{k}} \tag{2.8}$$

The one-dimensional detector image, only gives information in the reciprocal space. To obtain direct structural details of the real distances within a protein, an indirect Fourier transform can be done and is given by the pair distribution function ($P(r)$) in Equation 2.9.

$$P(r) = \frac{r^2}{2\pi^2} \int q^2 I(q) \frac{sin(qr)}{qr} dq \tag{2.9}$$

In addition to the Guinier approximation, the R_g can also be estimated from the pair distribution function ($P(r)$) without previous knowledge of $I(0)$ (Equation 2.10).

$$R_g = \sqrt{\frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr}} \tag{2.10}$$

Modelling

The distances, calculated with the pair distribution function ($\Delta P(r)$), can be used to create three-dimensional models (Figure 2.6), select ensembles of structures or even validate MD simulations models.

***Ab-initio* modelling.**

From the low-information content scattering curves, *ab-initio* modelling can be used to reconstruct the overall shapes of a protein. This can be done without any prior knowledge of the protein of interest. Conventional SAXS analysis pipeline comprises the DAMMIN and GASBOR programs for *ab-initio* modelling [31, 54]. Both programs, create *ab-initio* low-resolution shapes of randomly oriented particles in solution. The software searches a volume that encloses the protein. This volume is filled with N densely packed spheres of radius r, referred to as dummy atoms or residues, for DAMMIN and GASBOR respectively. The programs use simulated annealing (SA) to perform a global minimization of the discrepancy between the experimental data and the calculated scattering of

the dummy-atom model. Several consecutive runs need to be iterated to refine models. Multiple runs of the modeling process significantly decrease the risk of over interpretation of underdetermined, low resolution and ambiguous models.

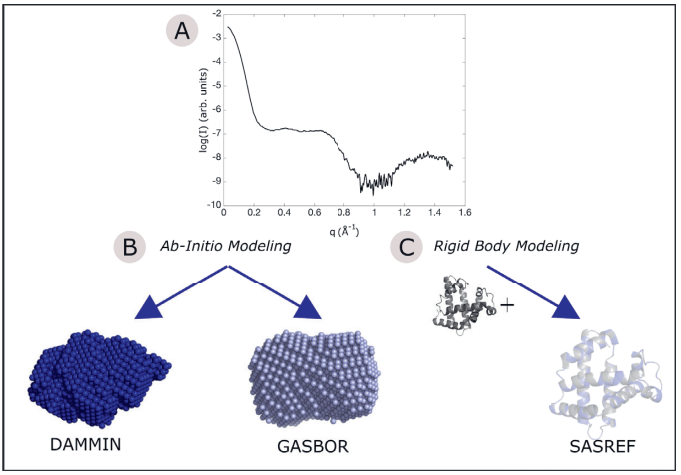

Figure 2.6: Three-dimensional models, characterizing the shape and size of apoMb, can be obtained from the one-dimensional scattering curve (A). Two different methods can be used: *ab-initio* modelling (DAMMIN, GASBOR) without any prior knowledge of the protein (B) or rigid body modelling (SASREF) when a high resolution structure is available (C). The crystal structure of myoglobin (PDB code: 1BVC) was used for the rigid body modelling, after biliverdin ligand removal.

Rigid body modelling.

Another method can be applied, when the high-resolution structure of the protein is known. In rigid body modelling, the high-resolution structure is used as template to create rigid bodies that are then reoriented and fitted against the experimental scattering data. Even though this method, computed for example with SASREF [32], provides higher resolution models, it has several major drawbacks. Since the reconstruction uses an existing structural model, the entire conformational space is not sampled. Furthermore, it cannot be verified whether the models are physically and biologically sensible.

2.3 Summary

X-rays are high-energy electromagnetic radiations that can be used to study the structure of proteins. At small angles, X-ray solution scattering (SAXS) characterizes the overall molecular shape and assembly of molecules. The lower resolution results can be used to describe flexibility and answer biological questions that enforce a single conformation. However, if higher resolution is needed for structural interpretation, SAXS can be combined with other methods such as MD (Chapter 3) or X-ray crystallography ([58]).

Chapter 3

Time-Resolved X-ray Solution Scattering

As discussed in Chapter 1, the function of a protein is closely related to its structure. Over the years, many biophysical techniques have been developed to explore the three-dimensional structure of proteins. However, many of these techniques, such as static NMR and X-ray crystallography, provide only a snapshot of a steady-state structure at equilibrium, or inform indirectly on the structure and kinetics of the protein such as circular dichroism (CD) or spectroscopy [59–65]. Time-resolved X-ray solution scattering (TRXSS) was recently introduced to study light-induced conformational changes in proteins [33, 34]. X-ray solution scattering measures distances between scattered electrons in a sample in solution (Chapter 2). Therefore, TRXSS measures the change in distances between atoms occurring after an induced perturbation. Examples of such a trigger can be a temperature jump or a UV-vis laser pulse. This technique is sensitive and more capable to detect subtle changes than conventional X-ray solution scattering [33, 52, 66–68] as it focuses on the observable differences between two states. TRXSS has many applications and has been used in this work to study the unfolding process of apomyoglobin and the signal-transduction pathway of blue-light sensitive proteins. In this chapter, I will first describe and explain how to conduct a TRXSS experiment in section 3.1. I will subsequently discuss the different steps of data processing in section 3.2. Finally, section 3.3 describes the tools and methods that have been used in this thesis to structurally interpret time-resolved data.

3.1 Conducting a TRXSS experiment

Generating X-rays

To carry on TRXSS experiments, and investigate protein conformational changes with a pico- to millisecond time resolution, a very high X-ray flux is necessary. As such, TRXSS experiments are performed in synchrotron X-ray laboratories. A synchrotron is a particle accelerator ring in which accelerated electrons produce a very intense light. Electrons are generated by an electron gun and gather speed in a linear accelerator, where they reach the near-speed of light. The electrons travel billions of kilometers per hour around the ring to produce intense synchrotron light. The ring possesses end-stations (or beamlines), where a particular light wavelength is selected to suit experimental purposes, from infrared to X-ray. TRXSS experiments in this thesis were realized in BioCARS beamline at APS and in cSAXS at PSI.

Triggering and detecting a conformational change

TRXSS investigates conformational changes in protein, triggered by a sudden perturbation. The experiments presented in this thesis, consider two different types of trigger: a temperature-jump (T-Jump) and a laser flash of appropriate wavelength. In **Paper I**, the unfolding mechanism of apomyoglobin (apoMb) was investigated. ApoMb is stable and folded up to 50°C after which the equilibrium gradually shifts towards the unfolded state. Accordingly, conformational change in apoMb has been triggered by a nanosecond T-Jump generated by an IR laser pulse (Figure 3.1). The laser pulse rapidly (<10 ns) increased the sample temperature from 50°C to 60°C. The proteins studied in **Paper II-IV** are blue-light photoreceptors. Their function is to sense and respond to ambient light, with a wavelength ranging from 400 to 495 nm. Consequently, conformational changes in the signaling pathway of blue-light photoreceptor proteins were initiated by a blue-light laser flash. The energy of the laser needs to be carefully selected to excite enough protein to observe a significant scattering signal. Yet, a too high laser power leads to a very high thermal solvent response (see Section 3.1) and potentially harms the protein. Laser-energy titration is therefore often required for each experiment.

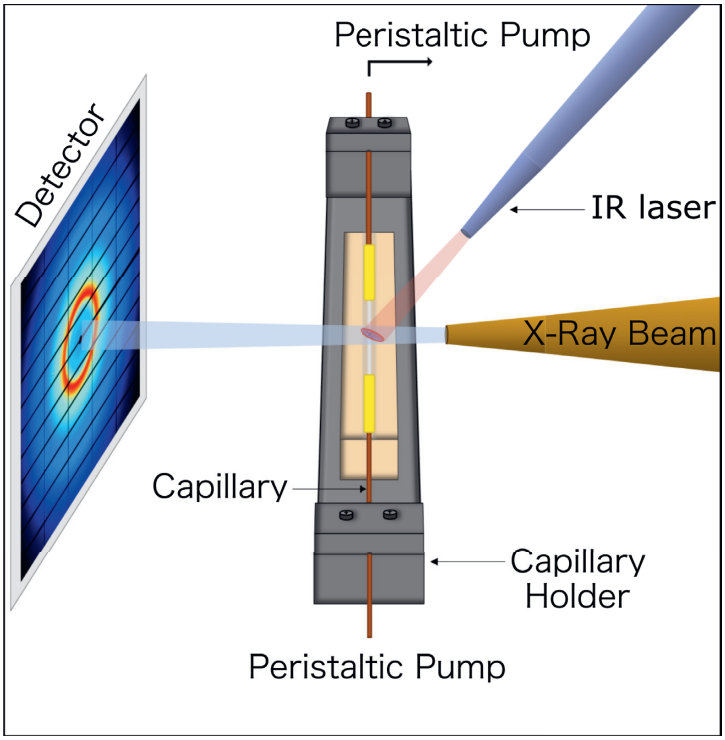

Figure 3.1: Representation of a TRXSS/T-Jump experiment set-up. The capillary is connected to a peristaltic pump (See Figure 3.2). Reproduced from Paper I.

Sample delivery and recovery

For the data collection, we circulated the sample in a round capillary made of quartz (Figure 3.1). The circular capillary had an inner diameter of 1 mm and a wall thickness of 10 µm. It is important to choose a thin capillary wall to limit background scattering emanating from the capillary itself. This was especially true in the T-Jump experiment (**Paper I**). The sample had to be aligned with the laser and X-rays, and stay in position during each measurement. We therefore mounted the capillary on a movable goniometer. Additionally, the sample environment temperature, at the cross-section of laser and X-rays, had to be constantly regulated and monitored in a T-Jump experiment. To do so, the sample capillary was

mounted on a temperature-controlled capillary-holder ensuring a stable sample temperature in-between the laser pulses (Figure 3.1).

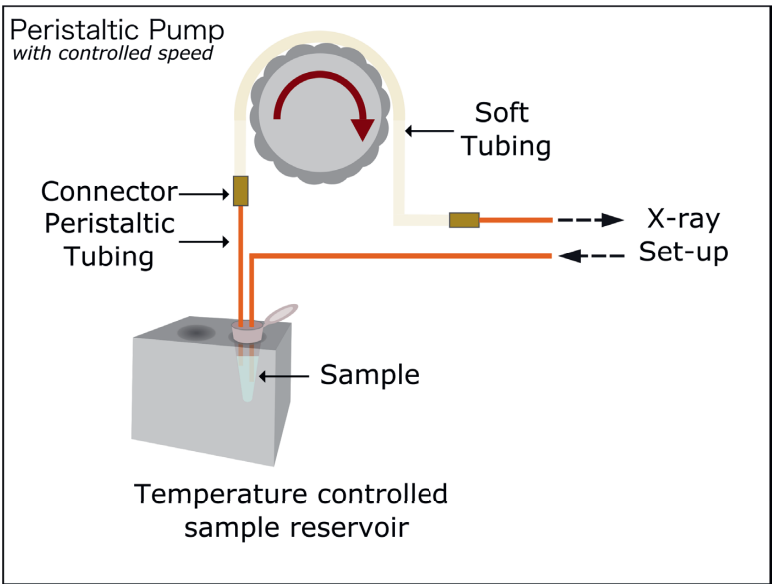

Figure 3.2: Representation of the sample environment and delivery. The sample reservoir is connected to a chiller to control and adjust the temperature. The speed and activity of the peristaltic pump can be locally or remotely adjusted.

To avoid pressure changes in the system, we delivered the sample by the means of a peristaltic pump (Figure 3.2), as opposed to syringe pumps in other experiments [36, 66, 69, 70]. The sample was circulating in a closed-circuit, to avoid producing and preparing large quantities of proteins. The speed of the pump was set such that the laser-activated sample would be expelled from the probing volume after each collection series but not before the triggered response could be measured at the desired time-delay.

Furthermore, since the sample was circulating in a closed loop, it had to be reverted back to its resting-state after each data collection event. Different strategies were employed depending of the sample of interest. Conveniently, the light-oxygen-voltage (LOV) domain of *Chlamydomonas reinhardtii* phototropin spontaneously reverts back to its dark-adapted

state in few seconds after its photoconversion. For the other samples studied in this thesis, the resting state could not be recovered so easily. For the T-Jump experiment, apoMb was reverted back to the initial folded state by circulating the sample in a "cooling" tube, set to 50°C. Temperature control was also used to revert the LOV domain from *Bacillus subtilis* YtvA. The photoproduct state of LOV *Bs*YtvA is highly dependent on the temperature and is stable at 20°C up to thousands of seconds. However, the photoproduct state reverts to the resting state at temperatures above 20°C. The sample reservoir (Figure 3.2) was then maintained at 35°C. An additional problem arises with the reversion of the cryptochromes that were not particularly sensitive to temperature changes. Instead, potassium ferricyanide was added to the buffer as oxidizing agent to recover the dark-adapted state of cryptochromes.

Sample concentration and scattering signal are linearly correlated. As such, a high concentration is preferred to obtain high-amplitude signals. However, a high concentration comes with side effects. As seen in Chapter 2, section 2.2, the sample in a X-ray solution scattering experiment must be monodisperse and ideal. At high concentrations, aggregation and inter-particle interactions may appear, altering the scattering at low angles (q<0.1 Å^{-1}). To account for these effects, X-ray scattering data were collected for a dilution series and extrapolated to infinite dilution [70].

Solvent thermal contribution

When a laser-pulse is used to trigger a conformational change in a sample, the energy of the absorbed photons will dissipate into the solvent. This energy will cause an increase in temperature and a conformational/structural change of the solvent that needs to be accounted for in the processing of the data ([34, 71]). The scattering that we directly observe in an experiment will thus comprise both the scattering of the solute and the scattering of the heated solvent (Figure 3.3). The contribution of the solvent is particularly important in a T-jump experiment and needs to be quantified. To do so, it is necessary to record the "pure heat" scattering signal. The solvent contribution can then be subtracted (Section 3.2), from the convoluted data, to obtain only the scattering of the particles in solution.

To record the scattering of the heated solvent, we used an IR laser

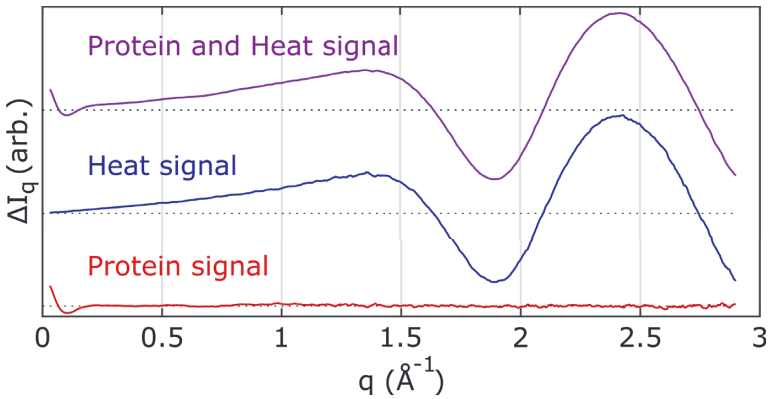

Figure 3.3: Thermal solvent response to an IR laser pulse. After triggering the sample with a laser energy, in the form of heat, is absorbed by the solvent. The signal measured (purple) contains the contribution of the scattering of the protein (red) and the scattering of the heated sample (blue).

pulse directly on the solvent [71]. This can also be done by heating a solvent dye (**Paper IV**) instead of the protein buffer [72]. For light-sensitive protein experiments, measuring the heat contribution of the solvent also confirms that the protein conformational changes measured are solely light-induced rather than heat-induced.

Recording the signal

During a TRXSS experiment, a sample is exposed to an intense X-ray beam (Figure 3.1). The laser and X-rays spots are aligned on an intersection of several μm^2 on the sample. The scattered waves are readout by a detector placed at a set distance from the sample. Two different methods were used in this work depending on the time scales investigated: a rapid readout approach (Figure 3.4 A) or a pump-probe acquisition (Figure 3.4 B). In the latter, the laser-trigger is synchronized with the X-ray pulse in order to record images at desired time-delays ($\Delta(t)$). This way, all the measuring power of the beamline can be dedicated to one delay-time. This method allows a higher time resolution in practice. Very short time scale such as pico- to nano-seconds can be reached, opening the field to the observation of very fast biological processes [73, 74]. In the rapid

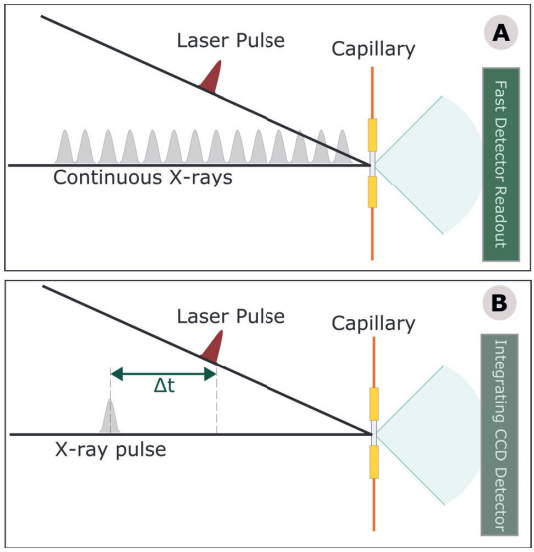

Figure 3.4: Representation of the different acquisition mode in a TRXSS experiment. Panel A represents a rapid readout acquisition while panel B represents the pump-probe acquisition. The time delays (Δt) of each data series, are precisely adjusted and synchronized with the X-ray pulse

readout mode on the other hand, continuous X-rays are delivered. The time limit of this experiment is defined by the detector read-out in itself and the time required to count the photons in each pixel. Therefore, the measurable time points will depend on the time spent for integrating the images. Currently, the readout of such detectors is on the microsecond time scale [75, 76]. Consequently, this rapid readout mode can be used to measure kinetics on a time range of several microseconds to seconds. Different beamlines offer different detectors. The pump-probe experiments were carried out at BioCARS (APS) while fast readout mode was done at cSAXS (PSI).

3.2 Processing the data

After an experiment, the images read by the detector need to be processed.

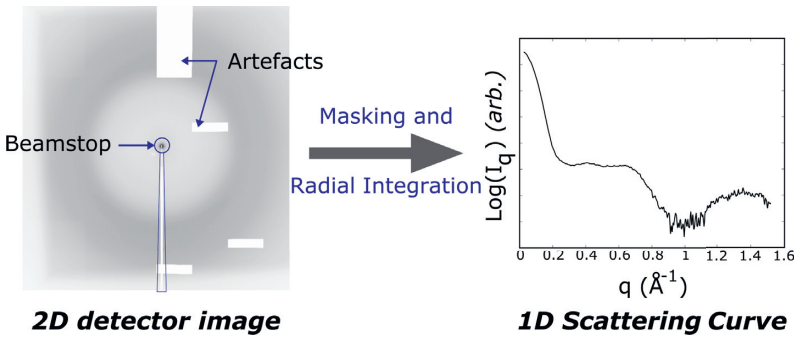

Figure 3.5: Radial integration of 2D detector images. The beam stop, that can be seen in the center of the detector image, stops the flux of excess X-rays not interacting with the sample. The beam center, together with the artefacts, need to be masked before obtaining the 1D scattering curve by radial integration.

Radial integration of the detector images

As seen in Chapter 2, the detector produces a two-dimensional spherical image of the average scattering. Even though the scattering pattern observed is an average over all directions, the scattering should be isotropic at a defined angle. As such, the one-dimensional profile of the scattering pattern is obtained by radial integration of the two-dimensional detector image (Figure 3.5). We reviewed in Section 3.1, the different acquisition modes: pump-probe or rapid readout approach. In a pump-probe experiment, a polychromatic beam is used as the entire X-ray flux is going through. Since the X-rays contain a number of different wavelengths, a smearing of the data will be observed and should be treated during the analysis. This is particularly true when comparing experimental and predicted scattering curve [33]. A monochromatic scattering curve can be transformed to a polychromatic beam when the wavelength distribution is known [77]. Furthermore, manipulations of the images need to be performed to mask some brighter pixels, along with the pixels in the vicinity of the beam stop (Figure 3.5). As those artefacts

are specific to each detector used at a specific beamline, software to mask them are provided by the beamline scientists.

Producing difference data

Once the one-dimensional scattering curves are obtained, difference scattering curves can be computed. In a conventional SAXS experiment, the buffer scattering is scaled and subtracted from the sample scattering. Similarly, the finger print of a conformational change, occurring after a perturbation (light or temperature), can be obtained over-time, by subtracting the scattering of the sample before laser excitation, to the scattering of the sample at each time-delay. The scattering of all curves also needs to be scaled and normalized before subtraction. It is common to normalize the data for $2<q<2.2$ Å^{-1}. In this q-range, the scattering of the solvent does not change with respect to the heating, and the protein difference WAXS signal is usually negligible. Unlike a regular SAXS experiment, the buffer scattering does not need to be subtracted as it will automatically be removed when difference scattering curves are generated (Equation 3.1)

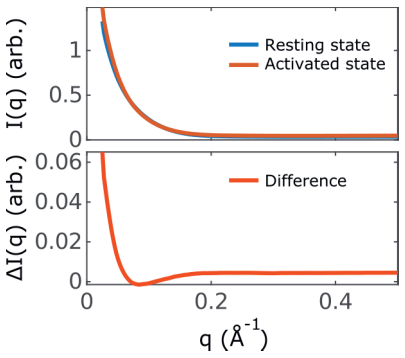

Figure 3.6: Even though very little deviation is observed in the scattering of the resting and activated states, a clear difference scattering signal is obtained highlighting structural changes.

$$\Delta I = (I_{RestingState} + I_{Buffer}) - (I_{ActivatedState} + I_{Buffer}) \tag{3.1}$$

X-ray solution scattering is a very sensitive technique. Even though the absolute scattering of the protein in the resting and the activated state are very similar, due to the overall agreement of their global shape, changes can be highlighted in the difference scattering curves (Figure 3.6).

Outlier rejections

Like in any experiments, anomalies exist in the data and have to be accounted for. Anomalies can be caused by an air bubble in the capillary or by protein aggregation on the capillary walls for example. To remove outlier frames in each data-set, we analyzed the q-range ($2<q<2.5$ Å^{-1}) where the scattering of water molecules predominates. For the study of photoreceptors, the conformational changes are triggered by a laser of a particular wavelength. In an ideal experiment, the water scattering should be the same for each time delays. Outliers were removed for scattering curves shifting more than a set percentage or standard deviation from the median. The data has to be treated differently in a T-Jump experiment, as the water scattering evolved during the course of the experiment. Typically, the solvent scattering is very high after the T-Jump and decays after ca. 1 ms when the heat starts to dissipate (Figure 3.7 A). The outlier rejection was then performed, based on the same criteria mentioned above, on the difference scattering data binned for each time delay instead of the entire data-set.

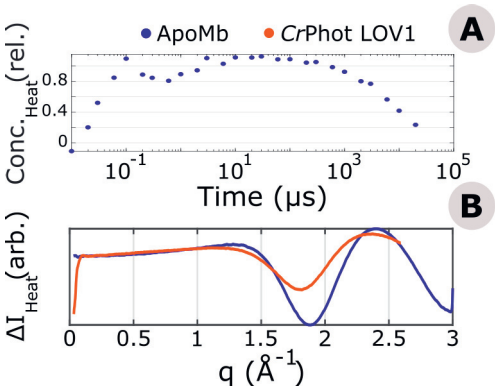

Figure 3.7: The relative slow heat contribution/concentration over time in a T-Jump experiment is shown in A. The slow heat difference signal produced in a regular TRXSS experiment (LOV1 in orange) and in a T-Jump experiment (ApoMb in purple) are depicted in B.

Subtracting the heat contribution

After the outliers have been removed, the heat contribution had to be measured and subtracted. As discussed in section 3.2, the heat contribution is particularly substantial in a T-Jump experiment (Figure 3.7 B) compared to a regular light-induced experiment. The fingerprint of the water scattering is conveniently located at $1.5<q<3$, where the protein scattering is hardly visible. The difference signal of the "pure" heat (Section 3.1) can thus be scaled and subtracted from the regular experiment difference data, in this particular q-range, to observe the conformational changes of the "pure" protein.

3.3 Analyzing and Interpreting the data

Aside from the radius of gyration (R_g), maximum dimension (D_{max}), and molecular weight (MW), described in Chapter 2 for conventional SAXS, a lot of information can be derived from a one-dimensional difference scattering curve. Different analysis strategies can be adopted depending on the information that is sought and the available knowledge of the system. Several strategies that have been adopted in this research are presented in this section.

Kinetic decomposition

Photoreceptors respond to ambient light. The chromophore moiety, embedded in their matrix, absorbs photons and undergoes conformational changes. For the majority of photoreceptor proteins, the photo-induced structural conversion occurs within seconds [39]. The reversion to their dark-adapted state, on the other hand, can take anywhere from a few minutes to a few hours. Due to this difference in time scales between photo-conversion and reversion, the photo-activation of a photoreceptor can be treated as an apparent forward reaction within the time frame of our experiments (nanoseconds to milliseconds). This simplifies greatly the search for a kinetic model. The difference scattering in a photoinduced experiment can conveniently be described with a sequential combination of basis spectra. These spectra can be obtained by singular value decomposition (SVD), principal component analysis (PCA) or kinetic

analysis, for instance. A mix of these methods is usually employed to find a kinetic model representing the experimental data and were used in **Paper II, III and IV**.

Information Content

Only a certain defined number of independent datapoints can be used to reliably interpret the experimental scattering data. The number of independent points (n_s) generally equals the number of independent Shannon channels [78] and can be estimated by Equation 3.2.

$$ns = (q_{max} - q_{min})\frac{D_{max}}{\pi} \tag{3.2}$$

Where D_{max} is the maximum dimension of the particle, and q_{min} and q_{max} represent the lowest and largest angle respectively that can be used. This means that a scattering curve has typically 10 to 50 independent data points after spherical averaging. Their information content is usually insufficient to determine all degrees of freedom and infer a three-dimensional molecular model without prior physical knowledge.

However, like in conventional SAXS analysis, an inverse Fourier transform can be performed on the difference scattering (Equation 3.3) to obtain the "difference pair-distribution function" ($\Delta P(r)$). Since, the pair-distribution function measures distances in real space between scattered electrons in regular SAXS, $\Delta P(r)$ will consequently measure the changes in distances between electrons in different conformational states.

$$\Delta\ P(r) = \frac{r}{2\pi^2}\int_{q_{min}}^{q_{max}} q\Delta I(q) sin(qr).\exp(q^{-2}\alpha^2)dq \tag{3.3}$$

where q_{min} has to be chosen such that the condition $q_{min} \leqslant \pi D_{max}$ is fulfilled, with D_{max} being the maximum dimension of the protein [79].

Molecular Dynamics Simulations

It has become common usage to integrate molecular dynamics (MD) simulations in the TRXSS analysis pipeline. TRXSS generates only low-resolution scattering curves providing shape, size and kinetics information. However, TRXSS data can be used to filter relevant models generated by

MD simulations in order to supply insights of dynamic processes at the atomic level [35–37].

MD simulations were initially developed in the 1970's and have thrived with the increase of computational power [80].

In this thesis, *Classical* MD simulations have been run to simulate the unfolding of apomyoglobin and the photoactivation of blue-light photoreceptors. The applied *Classical* MD simulations resolve Newton's equation of motion, numerically and stepwise for all particles simultaneously [81], meaning that quantum mechanics are not involved. The CHARMM forcefield [82] has been used to ensure that the generated models are sensible, by specifying the form of bond angles, bond lengths, dihedral angles etc. The forces are calculated from the potential energy, inevitably governed by the chosen forcefield and the atomic coordinates. The pKa of polar residues are estimated at the TRXSS experimental pH and adjusted accordingly. Finally, the salt concentration of the system is fixed to neutralize the net charge.

A homology model of the protein of interest has to be built prior to any MD simulations, if a refined deposited three-dimensional structure is not available. This is especially true for photoreceptors, where atomic models of intermediates in the light state are not always accessible. Furthermore, different forcefields have to be adapted manually depending on the chromophore conformational state. The flavin mononucleotide (FMN) (**Paper II-III**) and the flavin adenine dinucleotide (FAD) (**Paper IV**) chromophores and their environment, were modeled in their dark and light adapted conformations to simulate the resting and excited states.

Restricted MD Simulations

Although MD simulations are generally accepted to accurately model protein dynamics [83], *in-silico* simulations are often limited to a few microseconds [84, 85]. However, many biological processes associated with large conformational changes like in unfolding for example, happen in a few milliseconds to seconds. Furthermore, the high energetic barriers and cost to brake bonds make it very difficult to simulate the unfolding of large systems such as apomyoglobin with the computational power currently available. The system remains trapped in a local minimum of the energy landscape when using regular MD simulations at high

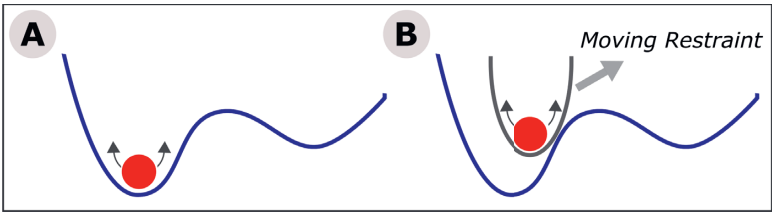

Figure 3.8: If a system is trapped in a local energy minimum (A), moving restraints (B) can be applied to force the system to cross higher energetic barriers.

temperatures (Figure 3.8 A). Over the years, different solutions have been developed [86] to overcome this limitation. In **Paper I**, we applied moving restraints on the MD simulations. PLUMED [87, 88] was used to add a time-dependent, harmonic restraint on the radius of gyration (R_g) which evolved from the folded to the experimentally determined unfolded values. This form of bias permits to cross the high energy barriers and eventually forces the protein to unfold at a lower computational cost (Figure 3.8 B).

X-ray scattering filtered MD

In the course of MD simulations many physically relevant models are generated. To select the conformations that best represent the experimental data, it is also possible to bias the energy landscape imposed by MD simulations with the X-ray scattering [89]. As an educated guess, this sampling method will guide the simulations towards the experimental data and will more likely create correct models. Used by many other biophysical techniques, such as NMR [90], this guiding method has unfortunately the major drawback of being very computationally intensive for large molecules. Generally, if further developed, this method could become a major strategy for the analysis of TRXSS data as it is sensitive to concerted conformational changes, even very small, in large parts of the protein.

We used a different method to overcome the lack of computational power (**Paper I-IV**). Instead, the experimental difference X-ray scattering data were used to filter the structural models generated by *classical*

and unbiased MD simulations. Different fitting strategies have been used. For photoreceptors (**Paper II-III**), the theoretical scattering of the dark and light models (see Section 3.3) were compared pair-wise to the experimental difference scattering. On the other hand, for the unfolding studies (**Paper I**) the average scattering of an ensemble of models has been fitted to the experimental scattering data to select relevant structures (see Section 3.3).

Scattering predictions from atomic coordinates

To gain structural and mechanistic understanding of a protein, the difference X-ray solution scattering can be compared to theoretical predicted curves derived from three-dimensional models. Several software suites exist to compute the scattering from atomic models such as CRYSOL, SASTBX, WAXSiS [54, 91, 92]. They differ principally on the spherical-averaging method, the hydration-layer representation and the excluded-solvent treatment. CRYSOL [54] was used in **Paper I-III** to calculate the scattering from models generated with MD simulations. This program is fast and is probably the most widely-used software for theoretical scattering computations. It uses multipole expansion of the scattering amplitudes to calculate the spherically-averaged scattering pattern. One drawback of CRYSOL, is the assumption that the hydration shell has a homogeneous and uniform density, which, in reality, is not the case. This consideration will *in fine* affects the scattering at q<0.1 Å^{-1}. Furthermore, due to the introduction of an excluded-solvent correction factor for the calculation of the theoretical X-ray scattering, artificial scattering features can appear at higher q. For this reason, CRYSOL was essentially used to calculate the X-ray scattering for 0.1<q<0.8 Å^{-1} of the generated models.

In **Paper II** and **Paper III**, the theoretical scattering of the dark and light adapted state models of *Cr*Phot and *Bs*Ytva LOV domains were calculated. Theoretical difference-scattering data was generated by considering the scattering of every possible pair of models. The difference data of each pair was then compared to the experimental scattering data. The best pairs were selected based on the minimum discrepancy, defined by the lowest R-factor, between experimental and theoretical difference data. This method, although very useful, is restricted to homogeneous states since only one structure representing a defined conformation is

selected.

Ensemble fitting

Thermal unfolding of a protein occurs in a succession of discrete equilibria. As such, a mixture of conformational states is observed for each delay-time rather than one single state, as opposed to light-driven TRXSS experiments.

A different approach to the selection of a single model was used to fit the experimental scattering data in unfolding studies (**Paper I**). Instead, an ensemble of structures was fitted to the experimental difference scattering by the means of an adapted genetic algorithm. Analogous to biological evolution, a genetic algorithm is an optimization method based on natural selection. In this genetic algorithm, similar to that of the ensemble optimization method (EOM) in the ATSAS suite of analysis tools [93], ensembles of structures evolve towards the optimal solution (Figure 3.9). The scattering of each MD-simulated structure is calculated with CRYSOL (Section 3.3) to form the initial pool of available models. Parents (ensembles of MD structures) are selected, exchanged and mutated randomly from the initial population to produce children (best ensembles) for the next generation. The minimum discrepancy between the experimental and average calculated scattering of the ensemble *Chi-square* is defined as the selection criteria for the best ensembles. The algorithm is iterated for a set number of generations.

Once the ensembles of structures best representing the experimental data are selected, a clustering of the models can be performed considering defined structural parameters such a, R_g, solvent accessible surface (SASA), root-mean square deviation (rmsd), helicity etc. As opposed to photoreceptor studies, where a sequential model is assumed, a kinetic model cannot be easily derived from the difference data due to the non-homogeneous mixtures of states. To extract the time constants of the conformational changes occurring during the unfolding process, it is possible to perform a temporal decomposition of each cluster population (**Paper I**). This allows the identification of "true" kinetic intermediates.

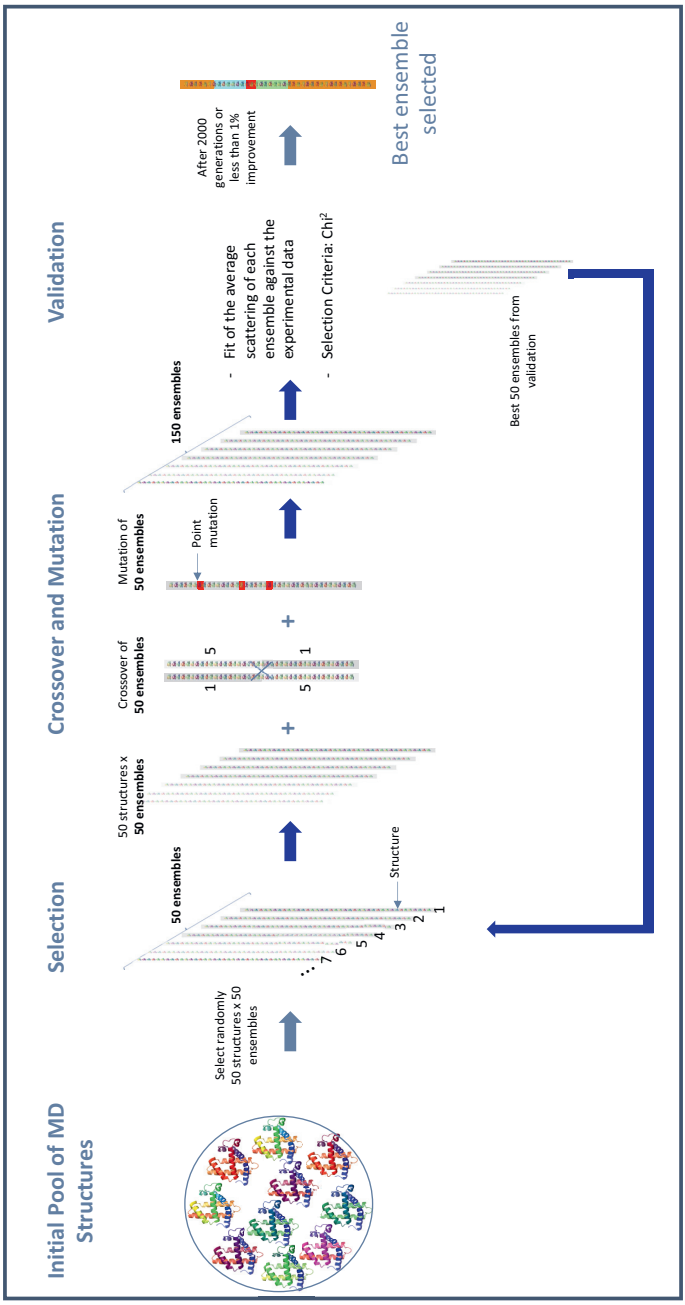

Figure 3.9: Genetic algorithm workflow for the selection of MD structures by ensemble fitting.

3.4 Summary

Time resolved X-ray solution scattering is a powerful method to obtain direct information of the particle shape, size, molecular weight, and protein flexibility. Since TRXSS is a low-resolution technique, it is often combined with another biophysical technique such as molecular dynamics simulations. The kinetics and dynamics of essential biological mechanisms can be obtained at atomic level and in great detail, using TRXSS in combination with MD simulations. I will present a few examples of application of these methods in the next chapters.

Chapter 4

Apomyoglobin

In **Paper I**, we applied time-resolved X-ray scattering to investigate and comprehend the unfolding mechanism of apomyoglobin. The three-dimensional structure of myoglobin, protein responsible of the dioxygen storage in muscle, was the first protein structure to be solved by X-ray crystallography [15]. Weighing only 17 kDa, and stable at a wide range of temperatures, apomyoglobin (without the heme) is particularly suited for three-dimensional structure and dynamics investigation. Apomyoglobin is a well-characterized α-helical globular protein that has been extensively used as a model system for protein folding and stability studies [94–99]. The folding field has particularly been reinforced by the growing interest in neurodegenerative diseases that result from protein unfolding, misfolding and aggregation, which range from Alzheimer's disease [100] to cancer [101, 102]. As most of the available biophysical techniques are only providing indirect structural information, or only focus on the equilibrium of stable intermediate states, TRXSS is notably adapted for the characterization of protein folding transient states. TRXSS can ultimately gain insights into the kinetic and thermodynamic relationships between protein folding and unfolding.

4.1 Apomyoglobin structure

Myoglobin is a protein which belongs to the globin family, and is found in the muscle tissue of many animals. Similar to hemoglobin, it contains a heme prosthetic group that binds dioxygen. Myoglobin accepts oxygen from the red blood cells and transports it to the mitochondria of muscle

cells. Apomyoglobin (apoMb), corresponds to myoglobin in which the porphyrin ring (heme) has been removed. ApoMb consists of eight alpha helices (80% helicity [96, 103]) folding into a compact and globular structure (Figure 4.1). It has a R_g of ca. 18 Å and a D_{max} of 56 Å [103].

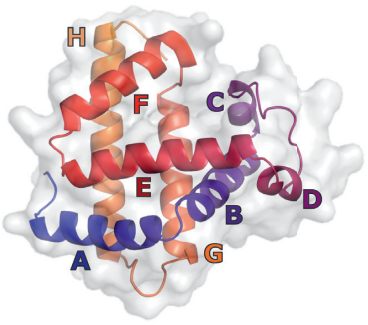

Figure 4.1: Apomyoglobin three-dimensional structure and surface obtained by X-ray crystallography (PDB code: 1BVC after biliverdin ligand removal)

Composed of 154 amino acids, many residues are conserved among globins [104] and especially between myoglobins in animals (Figure 4.2). Indeed, Ptitsyn and Ting [104] have analyzed 728 globin sequences and found two groups of residues. One group composed of 10 conserved functional positions is involved in the binding of the heme (Figure 4.2). Another group, more distant does not show any involvement in primary oxygen-carrying function. This group comprises three pairs of non-functional residues (V11 and W15, I112 and L116, and M132 and L136), all located on α-helices. The residues of each pair are precisely separated by four residues, forming a complete helix-turn. Interestingly, they are also located on helix A, G and H that form a compact hydrophobic core in the early event of apoMb folding [97]. Several mutation studies showed that the stability of the native state was greatly reduced when those non-functional residues were substituted [105, 106]. Therefore, they have been suggested to promote the rapid and correct folding process to the native tertiary structure.

Figure 4.2: Multi-sequence alignment of different myoglobins and globins among animals and bacteria.

4.2 Folding intermediate states in apomyoglobin

Proteins spontaneously fold from randomly unfolded conformations to biologically active structures in a hierarchical manner. The folding pathway of apoMb has been intensely studied mainly by NMR [94–99] and circular dichroism [107, 108], under kinetic and equilibrium conditions. A three-step mechanism of unfolding has been proposed from a folded (**F**) through an intermediate (**I**) to an unfolded (**U**) state:

Folded State *Molten Globule* *Unfolded State*

The intermediate state, found under equilibrium conditions, has been characterized at the atomic level as a molten globule [107]. This latter corresponds to an extended but condensed form of the native-state configuration, with a reduced content of native-like secondary structures [103]. The topology of a molten globule is also similar to that of the native state. The residue side-chains are more mobile, so that a molten globule is generally seen as a looser native state. Molten globules have been undetected for a long time due to their transience and the limitations of the available biophysical techniques to attain early timescales, such as micro- or even nano-seconds. With the technological advances, new examples of molten globules have been seen and they are now believed to be essential intermediates in the folding pathway of proteins [62, 109, 110].

For example, the molten globule existing in the apoMb unfolding pathway, has been described under equilibrium conditions at low pH [97, 103, 111–114]. It has been depicted as compact but lacking secondary structure. Even though it remains relatively folded and condensed, the apoMb molten globule, has a helicity of only 40-50% [96, 103]. The R_g of the molten globule have been characterized by X-ray solution scattering to 23 Å [103, 113].

4.3 Structural characterization of apomyoglobin unfolding

In **Paper I**, we combined TRXSS and a T-Jump to investigate the unfolding process of apoMb. Due to the experimental set-up, the system temperature could only be raised by 10°C with the provided IR-laser. To select a reasonable equilibrium temperature, from which a 10°C deviation could sample unfolding, we evaluated the folding state of apoMb by measuring the intrinsic fluorescence under thermal denaturation. It was mostly folded at 50°C, while more than 50% of apoMb was unfolded at 60°C. Therefore, we measured steady-state SAXS on apoMb at 50°C for a dilution series. We then collected TRXSS data, after a 7 ns IR-laser pulse, raising the system temperature by 10°C, inevitably shifting the equilibrium towards unfolding. The results showed three states in the unfolding pathway of apoMb as previously observed: a folded, a molten globule and an unfolded state (Figure 4.3 A-C). By combining MD simulations and TRXSS, we could create models to interpret and characterize the unfolding process of apoMb. The time-resolution of the events allowed us to order the events with great details.

We found that the molten globule, prevailing in the solution 100 µs after the T-Jump, qualitatively resembles the molten globule intermediates found under equilibrium conditions at low pH. However, the kinetic molten globule that we observe is also quantitatively different. Indeed, it is folded (ca. 60%) with a slightly expanded R_g, only 6% increase compared to the resting state. Furthermore, we observe that helix E is substantially folded in the kinetic molten globule and docked on helix A. The MD simulations provided us with great detail of the atomic positions. While the surface of the protein is getting looser and most of the helices are unfolded, especially helices B, C and D, the molten globule keeps a tight and compact hydrophobic core.

We observed that even though the protein has the same shape and is relatively folded, less native-contacts persist in the molten globule compared to the native state. Particularly, we discovered that the remainder of the contacts are precisely the conserved non-functional residues highlighted by Ptitsyn already in 1999 [104].

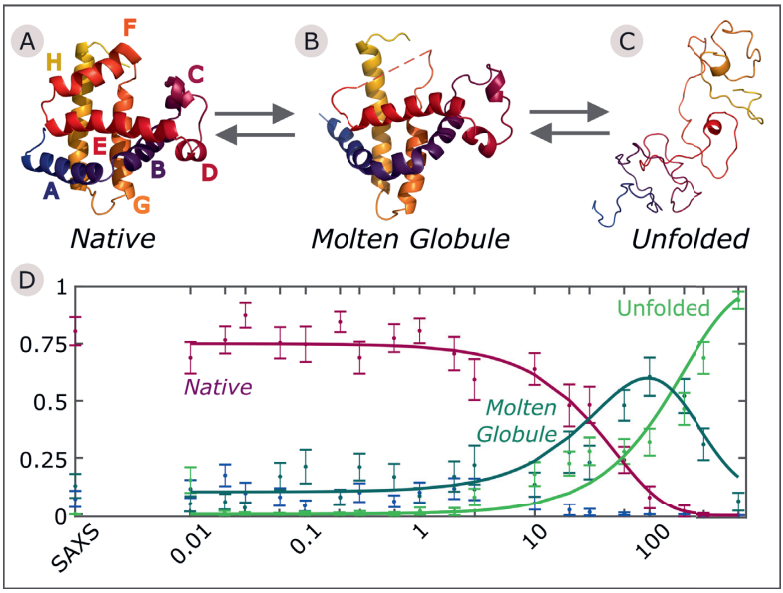

Figure 4.3: Unfolding pathway of apoMb. Upon a T-Jump, the native state of apoMb (A) expands to the molten globule state (B). Finally, the molten globule unfolds to form the unfolded state (C). As helix F has been shown not to be folded in myoglobin in the absence of heme, it is represented with dashed lines in B. The relative concentration of each state is represented in D. Adapted from **Paper I**.

In conclusion, the kinetic molten globule resembles the native state and can be seen as looser. As the molten globule unfolds, buried and destabilized residues start getting exposed to solvent. It is normal to observe some differences between the kinetic and the equilibrium molten globule as they have been observed under different experimental conditions. Furthermore, most of the studies considered the population at equilibrium to be the molten globule, but one can expect to also have a proportion of unfolded and folded in the mixture, affecting the values of the obtained structural parameters. By performing a decomposition of our data over time (Figure 4.3D), we could access the "pure" molten globule.

4.4 Summary

In **Paper I**, we have shown the application, potential and the advantages of using TRXSS in protein unfolding studies. TRXSS can identify transient intermediates on very early timescales and on wide length scales. By combining TRXSS and MD simulations, these techniques allow the identification and characterization of true kinetic intermediates at the atomic level. We have identified three states in the unfolding pathway of apoMb in accordance with previous studies: a native, a molten globule and an unfolded state. We have shown that the kinetic molten globule we observed is quantitatively different from the equilibrium molten globule previously described. Thereby, we have proposed and validated a powerful method to study unfolding mechanisms that can be applied to a wide range of proteins.

Chapter 5

LOV receptors

Blue-light photoreceptors, except cryptochromes, are by nature very modular in their architecture. These photoreceptors comprise typically one or two modules, conserved during evolution, that contain the chromophore and are responsible for the light absorption. Another module that exhibits light dependent catalytic activity or binding, serves as an output domain. The different modules are covalently linked and assembled in a relatively compact three-dimensional structure. The inherent modularity of blue-light photoreceptors makes them prime candidates for genetic manipulation and coupling to diverse output domains [115]. The elucidation of their photoreactions, structures, and signaling mechanisms have recently enabled applications in biotechnology and protein engineering, for example in optogenetics [45–47]. In the continuation of those new applications, TRXSS has been applied in **Paper II** and **Paper III** on light-oxygen-voltage (LOV) photoreceptors to understand their mechanism of signal transduction.

5.1 LOV photoreceptors

5.1.1 LOV domain structure

Light-oxygen-voltage (LOV) sensor proteins use a flavin mononucleotide (FMN) cofactor to absorb and respond to blue-light. They regulate many physiological processes including phototropism [40, 116] and plant growth optimization [41]. LOV domains, members of the large and diverse superfamily of PAS (Per, ARNT, Sim) domains, are structurally conserved from unicellular algae to higher plants. LOV domains (Figure 5.1) comprise five

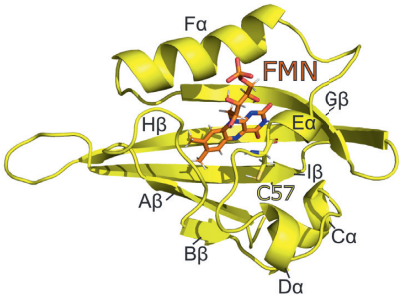

Figure 5.1: Structure of the LOV1 domain from *Chalmydomonas reinhardtii* in the dark state (PDB: 1N9L).

anti-parallel β-strands (Aβ, Bβ, Gβ, Hβ and Iβ) and four α-helices (Cα to Fα). Helices Eα and Fα pack against the β-sheet to shape a binding-pocket for the flavin isoalloxazine ring. Over the years, many residues have been identified to be essential for the phototransduction through modification of the hydrogen bonding [117]. They are mostly localized around the chromophore, within the binding pocket. Particularly, a cysteine located on Eα within the conserved sequence GXNCRFL(Q), forms a covalent bond with the FMN upon blue-light illumination.

5.1.2 Photoreceptor structure

LOV photoreceptors contain typically one or two LOV domains joined by linkers of different sizes and an output domain. Even though the LOV domain structure is very conserved, the spatial arrangement of the full-length protein and mechanism of phototransduction vary between photoreceptors. Phototropin from *Chlamydomonas reinhardtii* (*Cr*Phot) and the light photosensor YtvA from *Bacillus subtilis* (*Bs*YtvA) have been under investigation in **Paper II** and **Paper III**, respectively.

*Cr*Phot

Chlamydomonas reinhardtii possesses only one phototropin gene. *Cr*Phot acts as a monomer [118] and mediates blue-light dependent sexual differentiation [119] and photosynthetic gene expression [120]. Two

LOV domains (LOV1 and LOV2) in the N-terminal part of *Cr*Phot form the sensory domain and a serine threonine kinase (STK) domain composes the C-terminal effector domain (Figure 5.2 A). In agreement with previous studies (Figure 5.2 B, [118]), our SAXS data (Figure 5.2 C) shows that the *Cr*Phot full-length is arranged in tandem. However, the exact spatial positioning of the different domains is unknown, as no high-resolution structure has yet been provided for the phototropin full-length. Additionally, two helices, namely the A'α helix and the Jα helix, at the N- and C-terminal ends of the LOV2 domain, respectively, have been shown to be important for signal transduction [121]. The Jα helix, with conserved residues among phototropins, is embedded on the surface of the α/β-scaffold at the opposite side of the FMN binding pocket [122, 123].

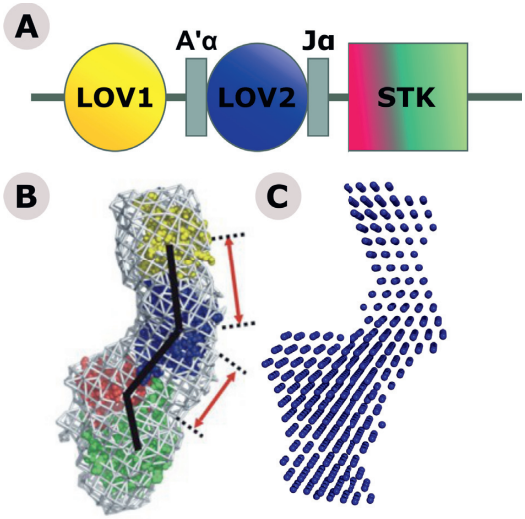

Figure 5.2: Organization of the three functional domains: LOV1, LOV2 and STK of *Cr*Phot full-length (A). Low resolution models of *Cr*Phot full-length in the dark obtained by SAXS are presented in B (adapted from [118]) and C (SAXS data measured at APS, not published). Homology modeled Ser/Thr kinase (N-lobe: red, C-lobe; green), LOV2 (blue) and the crystal structure of LOV1 (yellow) are fitted to the restored SAXS models in B [118].

*Bs*YtvA

The photoreceptor *Bs*YtvA is a dimer in solution. Each monomer consists of a LOV domain, a C-terminal Jα helix and a Sulfate Transporter and Anti-Sigma factor antagonist (STAS) domain. In contrast to *Cr*Phot, it is the N-terminal A'α helix that docks on the β-sheet (Figure 5.3). The receptor dimerizes in parallel through the LOV domain β-sheet surface and the A'α helix.

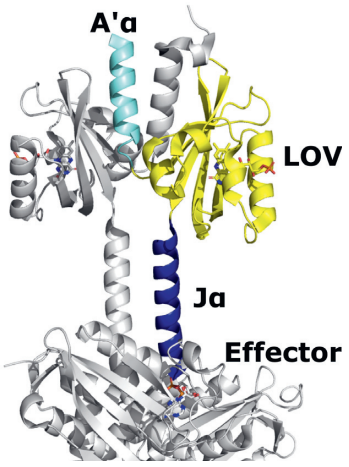

Figure 5.3: Crystal structure of the dimer *Bs*YtvA. The STAS output domain is replaced by a histidine kinase in this case (PDB: 4GZC)

5.2 LOV domain photochemistry

The photochemistry of LOV domains has been very well resolved [124–127]. In the dark state, the FMN chromophore (FMN^{450}, Figure 5.4) is fully oxidized and has a typical absorption spectrum with λ_{max} ranging from 445 to 450 nm (Figure 5.5). Upon absorption of a photon, an initial singlet-excited state (FMN^{*}, Figure 5.4) is formed within picoseconds, rapidly switching through intersystem crossing to a triplet-excited state (FMN^{T}, Figure 5.4).

An adduct is formed within microseconds, between the sulfur of a conserved cysteine residue and the C(4a) carbon of the isoalloxazine ring. This adduct, characterized by a maximum absorption at 390nm (Figure 5.5), constitutes the signaling state enabling the activation of the output domain. The formation of the cysteinyl adduct induces a tilt of the FMN chromophore and the protonation of the nitrogen N5. The proton could come from the cysteine itself [128]. Recent publications actually suggest that the protonation of N5 rather than the formation of the cysteinyl-adduct is required for the signal transduction [129]. Finally, the signaling state reverts through thermal process to the dark state within seconds to hours depending on the receptor (Figure 5.5). Recovery to the dark state requires the breakage of the flavin-C(4a)-cysteinyl adduct bond and the protonation of N5 and S of the FMN and cysteine respectively [130].

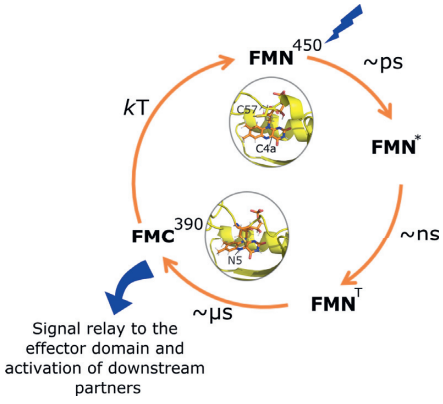

Figure 5.4: The photocycle of LOV domains. PDB: 1N9L

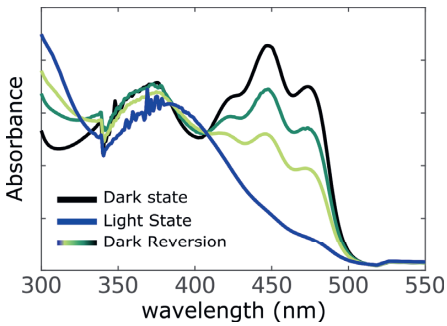

Figure 5.5: Typical absorption spectrum of a LOV domain in the dark (black) and after 445nm light illumination (blue). The protein reverts back to the dark state within seconds to minutes depending on the receptor.

5.3 Signal transduction in the LOV domain

Even though the photochemistry of the LOV domain is quite well understood, it is still unclear how the signal is relayed to the effector domain. Additionally, the vast majority of the mechanisms of signal relay are not conserved among blue-light photoreceptors [127]. Depending on the architecture and the output domain, LOV photoactivation can lead to photoreceptor dimerization [131, 132] or monomerization [133] and the re-positioning of the LOV domains with respect to each other [134]. Very similar in structures, the LOV domains may even have different functions within the same protein. In *Cr*Phot for example, LOV1 and LOV2 possess different roles. LOV2 has been shown to be responsible for blue-light dependent auto-phosphorylation of the kinase domain [135] through the release of the Jα helix upon adduct formation. While LOV1 is thought to act as a dimerization site in higher plants [132, 136], it must possess a different function in *Cr*Phot since only one isoform exists in the green algae and the photoreceptor functions as a monomer. As of today, the difference in functional role among LOV domains is still unclear.

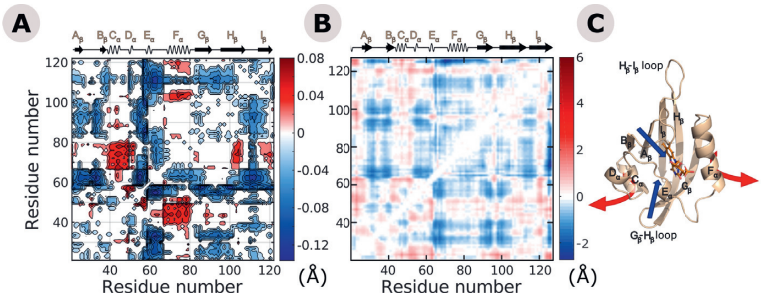

Figure 5.6: Cα distance difference maps between light and dark for *Cr*Phot (A) and *Bs*YtvA LOV domains (B). The conformational movements associated with photon absorption are shown in C (PDB: 1N9L).

We measured TRXSS on both *Cr*Phot LOV1 (**Paper II**) and *Bs*YtvA LOV domain (**Paper III**) to elucidate their mechanism of signal relay. For both samples, we observe small but global conformational changes in LOV that can be well explained by a mono-exponential kinetic model, with a time constant of 1 to 2 µs. The time scale of these events corresponds to

the formation of the cysteinyl adduct as described by spectroscopy [124]. We assumed that the major conformational rearrangements that we see in the LOV domain, leading to the signaling state, are concomitant to the formation of the cysteinyl adduct. Structural modeling of the dark and light states based on TRXSS filtered MD simulations show that the initial events of signal relay from the chromophore to the β-sheet surface are conserved among LOV domains. Indeed, we observed small but significant and similar concerted movements upon adduct formation for *Cr*Phot (Figure 5.6 A) and *Bs*YtvA (Figure 5.6 B) that are characterized by the spacing of helix Cα and Fα, and the movement of Iβ, Aβ and Bα towards the FMN binding pocket. These movements are summarized in Figure 5.6 C.

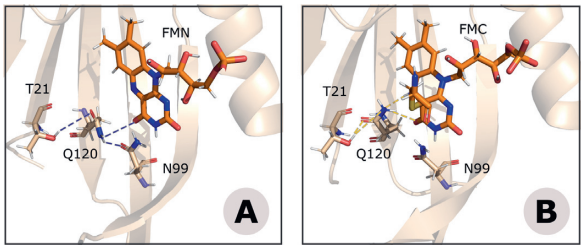

Figure 5.7: Hydrogen bonds involving the conserved GLN120 in the dark (A) and light state (B).

Through the generation of high resolution models of the *Cr*Phot LOV1 domain (**Paper II**), we could also gain atomic details on the mechanism of signal relay. It has been shown in previous studies that a rearrangement of the hydrogen bond network occurs upon adduct formation [137–139]. More specifically, a conserved glutamine has been proposed to play a key role in signal transduction by changing its conformation in the light state [140]. In our models representing the dark state, Q120 (in *Cr*Phot LOV1 numbering) forms a hydrogen bond with flavin-O4 (Figure 5.7 A). In the light state, the side chain flips to form a hydrogen bond with the protonated flavin-N5. The hydrogen bond that we see in the dark trajectory between Q120 and N99, on Gβ is broken in the light (Figure 5.7 B). As we observed major movements in the Iβ,Aβ/Bα region, we concluded that the change of conformation of Q120 is responsible of this movement,

in agreement with other studies [141]. Likewise, we found that the movement of Iβ also induces a change of electrostatic surface on the outer surface of the β-sheet precisely where Jα lies on the *Cr*Phot LOV2 (Figure 5.8).

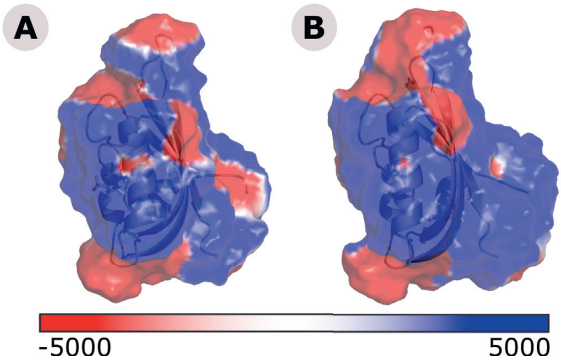

Figure 5.8: Electrostatic surface computed for the dark (A) and the light state (B). Blue surfaces indicate a positive surface charge while the red displays a negative surface charge.

In **Paper II**, we have demonstrated that even though LOV domains possess different roles, the initial events of signal transduction from the absorption of a photon to the conformational changes are conserved among LOV domains. Blue-light induced chain movements alter the hydrogen bond network and electrostatic surface. We suggest that the difference in role and function of LOV domains is associated with the nature of the linkers and the structural arrangement and placement of the LOV domains. Indeed, the binding site on the surface of the β-sheet can be occupied in LOV2 by the C-terminal Jα helix, the N-terminal extension or the A'α helix as seen for *Bs*YtvA [142]. Phototropin from the green algae *Chlamydomonas reinhardtii* functions as a monomer and is organized in tandem [118]. However, the exact positioning of the LOV1 domain with respect to the LOV2 domain is not known. In the light of the TrXSS and MD simulations results, we suggest that the LOV1 and LOV2 domains are in close proximity or interact through their β-sheets. As LOV1 presents similar conformational changes and feature than the LOV2 domain upon blue-light absorption [143], we do not exclude that LOV1 also interacts

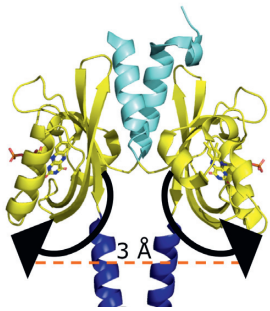

Figure 5.9: Light induced conformational change in *Bs*YtvA. (PDB: 2PR5)

with the A'α and the Jα helix. The *Cr*Phot LOV1 domain, even though not essential for the activation of the STK domain, has been suggested to extend the lifetime of the signaling state [144]. We propose that LOV1 participates in *Chlamydomonas reinhardtii* to the destabilization of the A'α and Jα helices.

5.4 Structural photoactivation of the YtvA-LOV domain

LOV activation alters a binding site on the surface of a β-sheet adjacent to the chromophore. While the modification of this binding site in LOV2 leads to the release of Jα in *Cr*Phot, our TRXSS data does not show any undocking of the A'α in *Bs*YtvA. Instead, the scattering data together with MD simulations revealed a light-induced separation of the two LOV monomers (Figure 5.9). The two domains splay apart at the level of the transition between the C-terminal part of the LOV domain and the Jα helix, while the A'α helices barely move. The separation of the two domains was only about 3 Å, in agreement with previous results [145]. As blue-light absorption leads to a change in the electrostatic surface of the β-sheet in *Cr*Phot, we can suggest that the same phenomenon happens in *Bs*YtvA leading to the repulsion of the two LOV domains. Increased separation of the Jα helices causes supercoiling of the effector domain regulating its kinase activity [53].

5.5 Summary

Despite the difference in function and structural arrangements of the different LOV domains we show with TRXSS that the light-induced conformational changes are conserved among LOV domains. Through the generation of high-resolution models by MD simulations we could analyze and identify crucial residues in the mechanism of signal transduction in LOV domains. As the LOV domains associate with different linkers and display different functions in spite of their similar activation mechanism, we suggest that their function is associated with their three-dimensional arrangements and the nature of their surrounding linkers. This is particularly promising for further photoreceptor manipulation. Added to their modularity and ubiquity, LOV domains could easily be recombined with different output domains to engineer new photoreceptors, like the light-gated histidine kinase YF1 for example [146].

Chapter 6

Cryptochromes

Cryptochromes are blue-light sensing photoreceptors present in all kingdoms of life. Playing a key role in development and growth, they also participate in the animal circadian clock and are proposed as magnetoreceptors in migratory birds. Interestingly, cryptochrome structures are closely related to photolyases (45%), evolutionary ancient flavoproteins that catalyze light-dependent DNA repair. Cryptochromes, which have lost the ability to repair DNA, function instead as signaling molecules, that regulate various biological responses. In **Paper IV**, we employed TRXSS to understand the mechanism of phototransduction and associated conformational changes, proffering the cryptochromes with this novel role.

6.1 The family of Photolyases/Cryptochromes

The Photolyases/Cryptochromes family is broad, diversified, and present throughout the biological kingdom. This family contains three major categories of proteins: the cyclobutene pyrimidine dimer (CPD) photolyases, the (6-4) pyrimidine-pyrimidone adduct photolyases and the cryptochromes. Photolyases are proteins that utilize the light energy for the repair of UV-damaged DNA. Cryptochromes, descending from photolyases, were first discovered in *Arabidopsis thaliana* and subsequently found in higher plants and animals. Even though photolyases and cryptochromes are structurally closely related (Section 6.2), cryptochromes have lost the ability to repair DNA and have instead gained a novel role in signaling [38]. In the course of evolution, two classes of cryptochromes have

been differentiated. Cryptochromes from plants (cry1 and cry2) and cryptochromes from animals (type I in insects and II in mammals). However, since mammalian cryptochromes of type II participate in light-independent transcriptional regulation, they will not be further described. This thesis has focused on the insect cryptochrome from *Drosophila melanogaster* (*Dm*Cry) and comparison has been made with the (6-4) *Xenopus laevis* photolyase (*X*lPho).

6.2 Cryptochrome Structure

Cryptochromes and photolyases present a great sequence similarity in their N-terminal domain. This is reflected in the common protein folding and binding of the flavine adenosine mononucleotide (FAD) chromophore (Figure 6.1). The first crystal structure of a member of the Photolyases/Cryptochromes family was that of the *E. coli* class I photolyase [147]. This structure displayed a bilobal architecture divided in two subdomains. One domain forms the well-known Rossmann fold, with a mixed α/β topology, whilst the second domain is mostly helical. The helical domain binds the FAD chromophore. As of today, this structural fold appears to be analogous in all the refined structures of class I DNA photolyase [148], (6-4) photolyases [149, 150], plant [151] and animal [127] cryptochromes and DASH cryptochromes [152] (Figure 6.1). The length and nature of the linker, between the α/β domains and the helical domain, differ between proteins.

Compared to photolyases, cryptochromes also possess a carboxy-terminal tail (CTT) of typically 10-100 residues [38]. Interestingly, the CTT occupies precisely the DNA binding pocket present in photolyases. Through a Phe-Phe-Trp (FFW) conserved motif, and especially phenylalanine 534, the CTT positions analogously to the (6-4) photolesion on the photolyase homology domain (PHD) [127] (Figure 6.1 A and B). Upon blue-light illumination, limited proteolysis assays and SAXS [153] measurement have suggested that the CTT of *Dm*Cry undocks or unfolds from the PHD to educe a binding site for interacting partners. The FFW motif has also been indicated to be instrumental in this process. However, by which mechanism is the CTT released after the light absorption remains

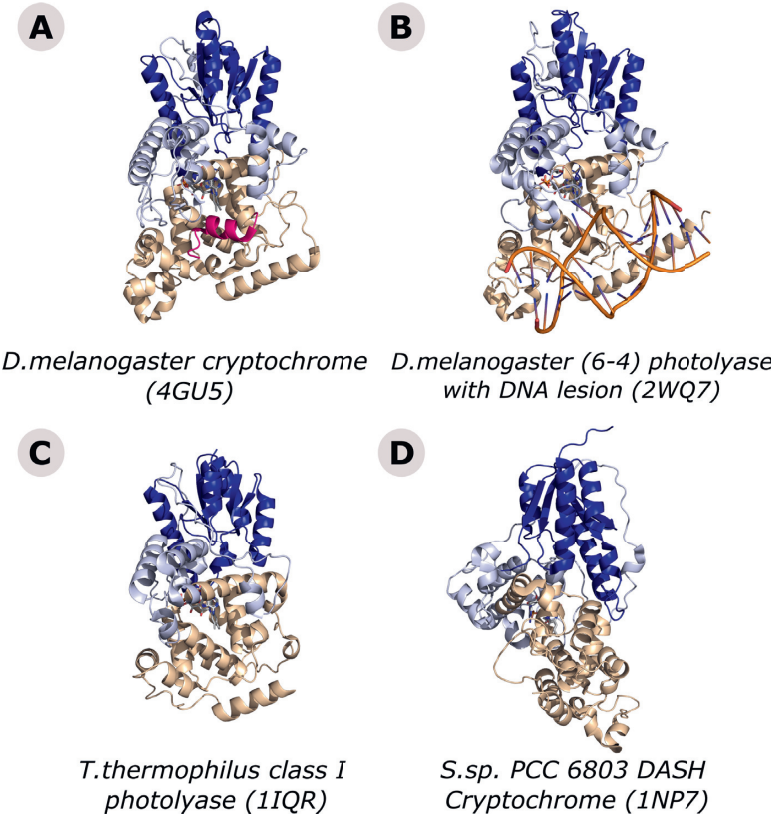

Figure 6.1: Crystal structures of *D.melanogaster* cryptochrome (A), *D.melanogaster*(6-4) photolyase in complex with the non-natural N4-methyl T(6-4)C lesion (orange) (B), *T.thermophilus* class I photolyase (C) and member of the newly identified class, *S. sp.* PCC 6803 DASH cryptochrome (D). The α/β and helical domain are colored in blue and beige respectively while the linker between the two domains is in light blue. The C-ter tail of cryptochrome is depicted in pink. The PDB access codes are given in parentheses.

largely unclear. Histidine 378 (H378) in *Dm*Cry has particularly been scrutinized, as its homolog in (6-4) photolyases is essential for the DNA repair activity [154, 155]. This H378 lies between the isoalloxazine moiety of the FAD chromophore and the CTT. Recently, a proton uptake by this histidine in *Dm*Cry has been suggested to alter the surrounding hydrogen bond network, ultimately leading in the release of the CTT [156].

6.3 Flavin photoreduction

Even though photolyases and cryptochromes present very similar structural features, they do not undergo the same reactions. While photolyases can perform a light-driven DNA repair and flavin-photoreduction, cryptochromes can only do the latter. Many studies have been carried out to understand these different functionalities. The FAD chromophore has be found in different redox states, namely fully oxidized (FAD_{ox}), fully reduced ($FADH^-$), and radical ($FADH^\bullet$). It appears that cryptochromes and photolyases have different photoactivation reactions. The FAD active form is fully reduced in photolyases, while it is fully oxidized in insect cryptochromes [38, 157]. In *Dm*Cry, an electron transfer is triggered upon blue-light illumination, from the FAD_{ox} along a series of tryptophan residues [147, 151]. Three of the tryptophan residues are conserved among photolyases. In animal cryptochromes, a fourth tryptophan has been recently discovered to participate in the electron transfer [127]. The electron transfer is very fast (few nanoseconds) and go through W420, W397, W342, and W536 eventually leading to the CTT release in *Dm*Cry.

6.4 Signal transduction

*Dm*Cry regulates the circadian function by inducing the light-dependent ubiquitin degradation of the timeless protein (TIM). It was suggested that the release of the CTT, after photoactivation, opens a binding site for TIM. This hypothesis is reinforced by the fact that *Dm*Cry CTT deletion mutant (ΔCTT) is constitutively active [158]. Light also activates *Dm*Cry degradation but on a longer timescale. In **Paper IV**, we used TRXSS to study the conformational changes of *Dm*Cry upon blue-light absorption,

and assess the role of H378 in signal transduction. We recorded TRXSS for wild-type *Dm*Cry and a histidine 378 to alanine mutant (*Dm*Cry H378A) at pH 7 and 9. If H378 was to uptake a proton, the solution pH should then have an effect on the formation of the signaling state.

The difference scattering signals of *Dm*Cry and *Dm*Cry H378A, evolved over time and were very similar at pH 7 (Figure 2, A and B in **Paper IV**). Five intermediates were found for *Dm*Cry by spectral kinetic decomposition namely, $DmCry_{\alpha}$, $DmCry_{\beta}$, $DmCry_{\gamma}$, $DmCry_{\delta}$, $DmCry_{\epsilon}$ (Figure 6.2 A,B and E).

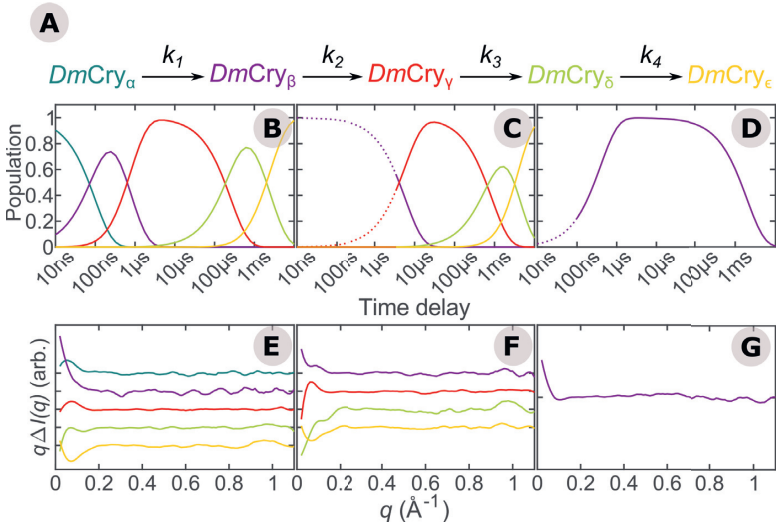

Figure 6.2: Kinetic analysis of the TRXSS data for wild-type *Dm*Cry (A) at pH 7. The time-dependent contributions of the different structural intermediates are presented for wild-type *Dm*Cry (B), *Dm*Cry(H378A) (C) and wild-type *Xl*Pho (D). The model outside the probed time region is shown as dotted lines. The basis spectra of the different intermediates are shown for *Dm*Cry (E), *Dm*Cry(H378A) (F) and wild-type *Xl*Pho (G), with the same color code.

Conformational transitions were calculated at ca. 100 ns, 770 ns, 259 µs and 2.5 ms. Only four species were found for ΔCTT (Figure 6.2 C and F), resembling $DmCry_{\beta}$ - $DmCry_{\epsilon}$. Since TRXSS data was recorded from 3 µs for ΔCTT, we assumed that $DmCry_{\alpha}$ was not directly observed due to the experimental time scale and was convoluted together

with *Dm*Cry$_\beta$. Clear difference signals, up to 300 µs, were observed in the low q region (q<0.1 Å^{-1}) indicating small but concerted changes of the global protein structure. From 1 ms, difference signal started to appear in the intermediate q range (0.1<q<0.2) for *Dm*Cry$_\varepsilon$, indicating major secondary structure rearrangements that we associated with the conformational change of the CTT. We thus attributed *Dm*Cry$_\varepsilon$ to the signaling state. This was supported by the fact that we did not observe such change in *Xl*Pho (Figure 2C in **Paper IV**), which does not possess a CTT. Only one state, resembling the early *Dm*Cry$_\beta$, was found in *Xl*Pho (Figure 6.2 D and G). This correlated well and confirmed that the two proteins undergo similar electron transfer reactions after photoactivation.

Together, these results demonstrated that H378 is not essential for the formation of the signaling state. We proposed that H378 is instead stabilizing the CTT in the resting state, by forming a hydrogen bond with W536 in the conserved FFW motif (**Paper IV**, Figure 4F). This would prevent a signal leaking in the dark state and a constitutive activation of TIM. Nonetheless, *Dm*Cry$_\varepsilon$ was not observed at pH 9 for *Dm*Cry and *Dm*Cry H378A (Figure S4 A and B in **Paper IV**). This indicated that a proton uptake was still required for the formation of the signaling state and that this reaction is very much pH dependent. The identity of this proton donor is still unknown.

Finally, in the light of the results of **Paper IV** we propose the following sequence of events in the photoactivation of *Dm*Cry (Figure 6.3). In the first events of photoreaction (*Dm*Cry$_{\alpha\text{-}\gamma}$), the hydration layer rearranges. Simultaneously, the photo-excited state corresponding to the radical pair FAD$^{\bullet-}$...W394$^{\bullet}$ is formed. The protein corresponds then to *Dm*Cry$_\gamma$, which is slightly more compact than the resting state. As a consequence of the reduction of FAD, the hydrogen bond between H378 and W536 is broken. Subsequently, the CTT loosens up or becomes unstructured (*Dm*Cry$_\delta$). The protonation of a yet unknown residue leads eventually to the release of the CTT (*Dm*Cry$_\varepsilon$). The formation of this signaling state finally enables the interaction with downstream partners.

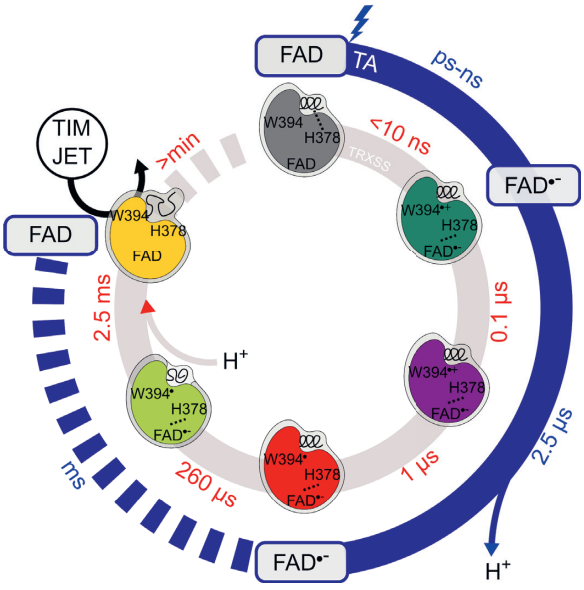

Figure 6.3: Proposed structural photocycle for *Dm*Cry (in grey) and structural rearrangements in the different intermediates based on our TRXSS data. The associated photochemical photocycle of the FAD chromophore (in blue) as determined by transient absorption (TA) spectroscopy is presented in the outer circle.

6.5 Summary

Through **Paper IV**, we have shown another example of application for TRXSS in combination with MD simulations. We could study the global conformational rearrangements of *Dm*Cry upon blue-light illumination. Thanks to the atomic details provided by MD simulations, we could highlight the role of H378 in the stabilization of the CTT in the dark state, eventually gaining new insights in the signal transduction mechanism in the cryptochrome family. As deregulation of the circadian clock constitutes a significant factor in tumorigenesis for example [159], this study could bring new perspectives into the origin of non-functional copies of cryptochromes.

Chapter 7

Conclusions and Outlook

This work has presented the potential and power of time-resolved solution X-ray scattering to study protein structure and dynamics. This method has allowed us to address various biological questions on different time and length scales.

For example, this technique has brought new insights into the "protein folding problem" of small globular proteins (**Paper I**). Partially folded and misfolded proteins that escape the cellular quality-control mechanism have a high tendency to form inter-molecular hydrogen bonds resulting in aggregation. This could have drastic consequences and lead to neurodegenerative diseases and cancer. With this work, we have demonstrated that apomyoglobin unfolds in a succession of defined steps. We have also characterized the molten globule intermediate at the atomic level. Finally, we have suggested that conserved, non-functional residues initiate the mechanism of apomyoglobin folding. We hope that these results contribute to the broader understanding of conformational diseases.

In this thesis, we also have applied TRXSS to study the signal transduction mechanism in blue-light photoreceptors (**Paper II-IV**). In **Paper IV**, we have identified new structural intermediates in the photoactivation of *Drosophila melanogaster* cryptochrome. Photoreceptors not only govern diverse adaptive responses in organisms, but as light-regulated actuators, they also provide the foundation for optogenetics. The modular nature of blue-light photoreceptors has been particularly attractive for the design and engineering of new photoreceptors. Particularly, understanding the signal relay from the photosensory domain to the effector domain of

blue-light photoreceptors could help the fusion of new photoreceptor domains exhibiting different properties. In **Paper II** and **Paper III** we have examined the signal transduction from the chromophore to the outer surface of the light-oxygen-voltage domains from *Chlamydomonas reinhardtii* and *Bacillus subtilis*. We have shown that the essence of the conformational changes occurring upon photon absorption are conserved among LOV domains. We have proposed a new structural arrangement of the full-length *Cr*Phot and suggested a new role of the LOV1 domain. In the dimer *Bs*YtvA, we have demonstrated that the conformational changes lead to the separation of the two LOV domains at the anchoring of the effector domain. We suggest that the movement of the LOV domains enable the activation of the effector domain.

A fundamental asset of X-ray solution scattering is that it provides the overall structure, both in architectural arrangements and conformations, in the 10–50 Å resolution range. It can be used in near-physiological conditions. With this work, we have demonstrated that TRXSS has a wide field of applications. TRXSS offers also significant potential advantages for the structural analysis of membrane proteins in lipid bilayers and detergents. Due to their size and level of complexity, membrane proteins represent a great challenge for X-ray crystallography and NMR, which can be addressed with by X-ray solution scattering [34]. However, TRXSS also possesses major drawbacks that hopefully can be overcome in the future.

For example, the low-resolution retrieval of information has already been surmounted by the combination of TRXSS with high-resolution techniques such as X-ray crystallography or MD simulations (as used in this thesis). With the constant advances of computational power, we expect to see tremendous progress in the understanding but also the prediction of the structure and function of biological macromolecules. We will be able to simulate *in-silico* larger systems on even broader time-scales. The emergence of X-Ray Free-Electron Laser Facility appears to be particularly promising for TRXSS as well [160]. Nevertheless, there is also much room for improvement in the TRXSS data collection and processing, especially in automation. Aside from the practical problem of triggering a reaction, TRXSS could benefit from the development of sample delivery devices, which are up to this day, mostly "homemade". The development

of even faster read-out detectors could potentially allow investigations on even shorter time-scale [75]. Furthermore, there is a need for the development of high-throughput structural analyses, which have already been implemented for X-ray crystallography [161].

To conclude, I hope that TRXSS together with technological advances will become a widely adopted tool. This technique can play a great role in cutting-edge structural research. I hope that this technique will not only provide answer to existing problems but also explore new challenges and discoveries in structural biology.

Acknowledgements

It feels only like yesterday when I walked Lundberg's corridors for the first time. I got to meet a lot of wonderful people over those five years and I am very thankful! Whether it was around an Äkta, a spectrophotometer, a fika or a beer, you certainly have contributed to make this PhD a great and memorable experience.

First of all, I would like to thank my supervisor, **Sebastian**, for giving me the opportunity to work in his group. This has certainly been an eventful PhD and I learned a lot. Thank you for all the freedom in leading and running my projects. I definitely went out of my comfort zone.

I also want to thank **Richard**, my examiner, for all the support and fun/awkward moments in the lab. If you feel like renovating another house and ordering floors, you can still give "Theo" a call!

Huge and special thanks to my mentors, **Linnéa** and **Oskar**. Thank you, **Linnéa** for the support and help you have given me during my PhD and all the failed attempted projects. You taught me a lot and always had my back, but you also made work and trips so much fun! Thank you, **Oskar**, for taking me on board with your projects, for all your patience and help, especially with my exemplary math skills... I really appreciated your honesty and organization skills, particularly during beamtimes.

All the work in this thesis would not have been possible without our collaborators, **Ryan**, **Erik**, **Ralph**, and **Andreas**, who provided the cryptochrome and YtvA proteins. I would also like to acknowledge the beamline scientists at APS, ESRF and PSI who always helped us, even at indecent hours. The beamtimes and holidays around the world with **Oskar**, **Linnéa**, **Elin**, **Matthijs**, **Joachim**, **Ashley**, **Stephan**, and **Janne** were intense and

sleep-depraving but also hilarious and just brilliant. Thanks for the great Strahlzeits!

To all the nerdy scientists in the Westenhoff group, it has been an honor and a great pleasure to work with you all. Thank you, **Elin**, we started and will finish this together. I am really happy that I got to know you, partner in crime, lab, and partays. We'll see each other again at the lost and found counter for sure. Pardon my French **Matthijs** ! Thanks for the thesis proof-read, the great midnight taco-pies and all the theoretical help. To my dear Germans **Amke Pamke** and **Joachim**, I really enjoyed our Japan trip together (Apart for that last day. . .). Thank you **Andrea** and **Weixiao** for carrying on the timeless blue-light photoreceptor projects. Thank you to the Woldies **Emil**, **Ashley**, **Stephan**, **Petra E** and our new group members **Laras** and **Lidija**. I wish you all the best!

Thank you to all the other PIs, **Michal** (For the constructive comments), **Gisela** (the Gothia towers would have been nice!), **Gergely**, **Björn**, **Kristina**, and **Johanna**, for hiring inspiring scientists. Thank you to all the administrators and especially **Valida**, **Bruno** and **Lars**, who always made our experiments and equipments run smoothly.

My thanks extend to the rest of the great pets of Lundberg, who challenged my OCDs every day around the lab, a cup of coffee or during beerclub: **Cecilia S** (The most caring and loving person in the lab), **Ann** ("C'est toi"), **Tinna** (Great poker face!), **Rebecka** (The Lundberg powerful source of information and great roomie), **Laura** (Toujours aussi saignante), **Andreas** (DJ Dunge), **Doris** (For all the chocolate), **Maja** (Wear that nerdy tee-shirt proudly!), **Florian** (Can't wait to see your next performance!), **Jessica** (Good luck with the new 2021 vintage), **Petra B** (For taking good care of our äktas), **Per** (Rule that beerclub with an iron fist), **Daniel "Sarabi"**, **Ylber** ("Peww peewww. Schwoof.. aaarrrrggg (wookie screaming)"), **Yosh** and **Lisa** (Congratulations for the baby!), **Giorgia**, **Greger**, and **Viktor**.

I don't forget all the Lundberg pensionärer, **Majo** (My beloved frozen-margaritas and dance partner), **Rob D**, **David**, **Rajiv** and **Elin D**! You

have surely been missed!

Manoop, **Swagatha**, **Emilie**, **Darius**, **Damasus**, **Irena**, **Ashish**, **Katarina**, **Davide**, **Jake**, **Dimitra** and **Lisa**, keep up the good work ! I wish a warm welcome and good luck to all the newbies, **Jens**, **Owens**, **Jonatan**, **Analia**, **Lucija**, **Adina** and **Adam**.

Le meilleur pour la fin. Je tiens à remercier Éamon (##BigUpToLenaDafgård), ma famille et mes amis pour le soutien qu'ils m'ont apporté durant ce doctorat. Je suis fière d'en être arrivée là et c'est grâce à vous! Merci de m'avoir écoutée râler, changer d'avis et parler de choses que vous ne compreniez pas avec un intérêt certain. Vous êtes mon bol d'air et m'avez permis de finir ce doctorat "à peu près" normale. Tout simplement, Merci !

Bibliography

[1] R. L. Baldwin, "In search of the energetic role of peptide hydrogen bonds," 5 2003.

[2] R. Aurora, T. P. Creamer, R. Srinivasan, and G. D. Rose, "Local interactions in protein folding: Lessons from the α-helix," 1997.

[3] W. Kauzmann, "Some Factors in the Interpretation of Protein Denaturation," *Advances in Protein Chemistry*, vol. 14, no. C, pp. 1–63, 1959.

[4] C. Levinthal, "Are there pathways for protein folding?," *Journal de Chimie Physique*, vol. 65, pp. 44–45, 5 1968.

[5] C. B. ANFINSEN, E. HABER, M. SELA, and F. H. WHITE, "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, pp. 1309–1314, 1961.

[6] B. Nölting, *Protein Folding Kinetics*. Springer-Verlag, 2006.

[7] C. B. Anfinsen, "Principles that govern the folding of protein chains," 1973.

[8] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein folding: A synthesis," *Proteins: Structure, Function, and Bioinformatics*, vol. 21, no. 3, pp. 167–195, 1995.

[9] K. A. Dill, "Theory for the Folding and Stability of Globular Proteins," *Biochemistry*, vol. 24, pp. 1501–1509, 3 1985.

[10] S. W. Englander and L. Mayne, "The case for defined protein folding pathways," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. 8253–8258, 8 2017.

[11] W. Hu, B. T. Walters, Z. Y. Kan, L. Mayne, L. E. Rosen, S. Marqusee, and S. W. Englander, "Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 7684–7689, 5 2013.

[12] W. Hu, Z. Y. Kan, L. Mayne, and S. W. Englander, "Cytochrome c folds through foldon-dependent native-like intermediates in an ordered pathway," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. 3809–3814, 4 2016.

[13] Y. Xu, L. Mayne, and S. W. Englander, "Evidence for an unfolding and refolding pathway in cytochrome c.," *Nature structural biology*, vol. 5, pp. 774–8, 9 1998.

[14] H. J. Dyson and P. E. Wright, "Coupling of folding and binding for unstructured proteins," 2 2002.

[15] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis," *Nature*, vol. 181, no. 4610, pp. 662–666, 1958.

[16] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 1 2000.

[17] G. Rhodes, *Crystallography Made Crystal Clear*. Elsevier, 2006.

[18] H. Formanek and S. Formanek, "Protein crystallography by T. L. Blundell and L. N. Johnson ," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 33, pp. 2702–2702, 8 1977.

[19] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young, "Protein states and proteinquakes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 82, no. 15, pp. 5000–5004, 1985.

[20] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins.," *Nature*, vol. 450, pp. 964–72, 12 2007.

[21] R. Ishima and D. A. Torchia, "Protein dynamics from NMR," 9 2000.

[22] L. E. Kay, "REVIEW / SYNTHÈSE Protein dynamics from NMR 1," *Cell*, vol. 152, no. Palmer 1993, pp. 145–152, 1998.

[23] H. Yu, "Extending the size limit of protein nuclear magnetic resonance," 1 1999.

[24] R. H. Callender and R. B. Dyer, "Protein Dynamics: Time-Resolved Spectroscopic Studies," in *Encyclopedia of Biophysics*, pp. 2007–2015, Springer Berlin Heidelberg, 2013.

[25] M. A. Hough, "Choosing the optimal spectroscopic toolkit to understand protein function," 6 2017.

[26] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjær, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. Liang, A. V. Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjöhamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. H. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman, and R. Neutze, "Visualizing a protein quake with time-resolved {X}-ray scattering at a free-electron laser," *Nat. Methods*, vol. 11, pp. 923–926, 9 2014.

[27] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. Deponte, U. Weierstall, R. B. Doak, F. R. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt,

L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Ströder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. Spence, "Femtosecond X-ray protein nanocrystallography," *Nature*, vol. 470, pp. 73–78, 2 2011.

[28] M. Karplus and J. Kuriyan, "Molecular dynamics and protein function," 5 2005.

[29] L. A. Feigin, D. I. Svergun, and G. W. Taylor, "Direct Methods," in *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, pp. 147–183, Boston, MA: Springer US, 1987.

[30] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.," *Quarterly reviews of biophysics*, vol. 40, pp. 191–285, 8 2007.

[31] D. Svergun, "Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing," *Biophysical Journal*, vol. 76, pp. 2879–2886, 6 1999.

[32] M. V. Petoukhov and D. I. Svergun, "Global rigid body modeling of macromolecular complexes against small-angle scattering data," *Biophysical Journal*, vol. 89, no. 2, pp. 1237–1250, 2005.

[33] M. Cammarata, M. Levantino, F. Schotte, P. A. Anfinrud, F. Ewald, J. Choi, A. Cupane, M. Wulff, and H. Ihee, "Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering," *Nature Methods*, vol. 5, pp. 881–886, 10 2008.

[34] M. Andersson, J. Vincent, D. van der Spoel, J. Davidsson, and R. Neutze, "A Proposed Time-Resolved X-Ray Scattering Approach to Track Local and Global Conformational Changes in Membrane Transport Proteins," *Structure*, vol. 16, pp. 21–28, 1 2008.

[35] J. G. Kim, T. W. Kim, J. Kim, and H. Ihee, "Protein Structural Dynamics Revealed by Time-Resolved X-ray Solution Scattering," *Accounts of Chemical Research*, vol. 48, pp. 2200–2208, 8 2015.

[36] D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, I. Kosheleva, and L. X. Chen, "Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering.," *The journal of physical chemistry letters*, vol. 8, pp. 4413–4418, 9 2017.

[37] A. Björling, S. Niebling, M. Marcellini, D. van der Spoel, and S. Westenhoff, "Deciphering solution scattering data with experimentally guided molecular dynamics simulations," *J Chem Theory Comput*, vol. 11, no. 2, pp. 780–787, 2015.

[38] I. Chaves, R. Pokorny, M. Byrdin, N. Hoang, T. Ritz, K. Brettel, L.-O. Essen, G. T. J. van der Horst, A. Batschauer, and M. Ahmad, "The cryptochromes: blue light photoreceptors in plants and animals.," *Annual review of plant biology*, vol. 62, pp. 335–364, 2011.

[39] A. Möglich, X. Yang, R. A. Ayers, and K. Moffat, "Structure and Function of Plant Photoreceptors," *Annual Review of Plant Biology*, vol. 61, pp. 21–47, 6 2010.

[40] K. S. Conrad, C. C. Manahan, and B. R. Crane, "Photochemistry of flavoprotein light sensors," *Nature Chemical Biology*, vol. 10, pp. 801–809, 9 2014.

[41] A. Takemiya, S. I. Inoue, M. Doi, T. Kinoshita, and K. I. Shimazaki, "Phototropins promote plant growth in response to blue light in low light environments," *Plant Cell*, vol. 17, no. 4, pp. 1120–1127, 2005.

[42] O. Berntsson, R. P. Diensthuber, M. R. Panman, A. Björling, E. Gustavsson, M. Hoernke, A. J. Hughes, L. Henry, S. Niebling, H. Takala, J. A. Ihalainen, G. Newby, S. Kerruth, J. Heberle, M. Liebi, A. Menzel, R. Henning, I. Kosheleva, A. Möglich, and S. Westenhoff, "Sequential conformational transitions and α-helical supercoiling regulate a sensor histidine kinase," *Nature Communications*, vol. 8, p. 284, 8 2017.

[43] S. Bannister, E. Böhm, T. Zinn, T. Hellweg, and T. Kottke, "Arguments for an additional long-lived intermediate in the photocycle of the full-length aureochrome 1c receptor: A time-resolved small-angle X-ray scattering study," *Structural Dynamics*, vol. 6, 5 2019.

[44] J. S. Lamb, B. D. Zoltowski, S. A. Pabit, B. R. Crane, and L. Pollack, "Time-resolved dimerization of a PAS-LOV protein measured with photocoupled small angle X-ray scattering," *Journal of the American Chemical Society*, vol. 130, pp. 12226–12227, 9 2008.

[45] H. M. Beyer, S. Juillot, K. Herbst, S. L. Samodelov, K. Müller, W. W. Schamel, W. Römer, E. Schäfer, F. Nagy, U. Strähle, W. Weber, and M. D. Zurbriggen, "Red Light-Regulated Reversible Nuclear Localization of Proteins in Mammalian Cells and Zebrafish," *ACS Synthetic Biology*, vol. 4, pp. 951–958, 3 2015.

[46] K. Deisseroth, G. Feng, A. K. Majewska, G. Miesenböck, A. Ting, and M. J. Schnitzer, "Next-generation optical technologies for illuminating genetically targeted brain circuits," in *Journal of Neuroscience*, vol. 26, pp. 10380–10386, 10 2006.

[47] T. Ziegler and A. Möglich, "Photoreceptor engineering," 6 2015.

[48] B. E. Warren, "Perfect crystal theory," in *X-ray diffraction*, ch. 14, p. 381, Dover Publications, 1990.

[49] R. W. Harrison, "Phase problem in crystallography," *Journal of the Optical Society of America A*, vol. 10, p. 1046, 5 1993.

[50] C. E. Blanchet and D. I. Svergun, "Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution," *Annual Review of Physical Chemistry*, vol. 64, pp. 37–54, 4 2013.

[51] L. Makowski, "Characterization of proteins with wide-angle X-ray solution scattering (WAXS).," *Journal of structural and functional genomics*, vol. 11, pp. 9–19, 3 2010.

[52] M. Andersson, E. Malmerberg, S. Westenhoff, G. Katona, M. Cammarata, A. B. Wöhri, L. C. Johansson, F. Ewald, M. Eklund, M. Wulff, J. Davidsson, and R. Neutze, "Structural dynamics of light-driven proton pumps," *Structure*, vol. 17, pp. 1265–1275, 9 2009.

[53] O. Berntsson, R. P. Diensthuber, M. R. Panman, A. Björling, A. J. Hughes, L. Henry, S. Niebling, G. Newby, M. Liebi, A. Menzel, R. Henning, I. Kosheleva, A. Möglich, and S. Westenhoff, "Time-Resolved X-Ray Solution Scattering Reveals the Structural Photoactivation of a Light-Oxygen-Voltage Photoreceptor," *Structure*, vol. 25, pp. 933–938, 6 2017.

[54] D. I. Svergun, M. V. Petoukhov, and M. H. Koch, "Determination of domain structure of proteins from X-ray solution scattering.," *Biophysical journal*, vol. 80, pp. 2946–53, 6 2001.

[55] D. I. Svergun and M. H. Koch, "Small-angle scattering studies of biological macromolecules in solution," *Reports on Progress in Physics*, vol. 66, no. 10, pp. 1735–1782, 2003.

[56] G. Rhodes, "Chapter 6 - {Obtaining} {Phases}," in *Crystallography {Made} {Crystal} {Clear} ({Third} {Edition})*, Complementary {Science}, pp. 109–143, Burlington: Academic Press, 2006.

[57] R. P. Rambo and J. A. Tainer, "Accurate assessment of mass, models and resolution by small-angle scattering," *Nature*, vol. 496, pp. 477–481, 4 2013.

[58] A. Björling, O. Berntsson, H. Takala, K. D. Gallagher, H. Patel, E. Gustavsson, R. St Peter, P. Duong, A. Nugent, F. Zhang, P. Berntsen, R. Appio, I. Rajkovic, H. Lehtivuori, M. R. Panman, M. Hoernke, S. Niebling, R. Harimoorthy, T. Lamparter, E. A. Stojković, J. A. Ihalainen, and S. Westenhoff, "Ubiquitous {Structural} {Signaling} in {Bacterial} {Phytochromes}," *J Phys Chem Lett*, vol. 6, pp. 3379–3383, 9 2015.

[59] D. T. Clarke, "Circular Dichroism in Protein Folding Studies," *Current Protocols in Protein Science*, vol. 70, 11 2012.

[60] A. L. Serrano, M. M. Waegele, and F. Gai, "Spectroscopic studies of protein folding: Linear and nonlinear methods," 2 2012.

[61] M. Zeeb and J. Balbach, "Protein folding studied by real-time NMR spectroscopy," *Methods*, vol. 34, no. 1, pp. 65–74, 2004.

[62] A. Reiner, P. Henklein, and T. Kiefhaber, "An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 4955–60, 3 2010.

[63] H. Roder, K. Maki, and H. Cheng, "Early Events in Protein Folding Explored by Rapid Mixing Methods," *Chemical Reviews*, vol. 106, pp. 1836–1861, 5 2006.

[64] B. Schuler and W. A. Eaton, "Protein folding studied by single-molecule FRET," 2 2008.

[65] Y. Bai, "Protein Folding Pathways Studied by Pulsed- and Native-State Hydrogen Exchange," *Chemical Reviews*, vol. 106, pp. 1757–1768, 5 2006.

[66] H. S. Choa, N. Dashdorj, F. Schotte, T. Graber, R. Henning, and P. Anfinrud, "Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 16, pp. 7281–7286, 2010.

[67] H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen, and S. Westenhoff, "Signal amplification and transduction in phytochrome photosensors," *Nature*, vol. 509, no. 7499, pp. 245–248, 2014.

[68] O. Berntsson, R. Rodriguez, L. Henry, M. R. Panman, A. J. Hughes, C. Einholz, S. Weber, J. A. Ihalainen, R. Henning, I. Kosheleva, E. Schleicher, and S. Westenhoff, "Photoactivation of Drosophila melanogaster cryptochrome through sequential conformational transitions," *Science Advances*, vol. 5, p. eaaw1531, 7 2019.

[69] T. W. Kim, J. H. Lee, J. Choi, K. H. Kim, L. J. van Wilderen, L. Guerin, Y. Kim, Y. O. Jung, C. Yang, J. Kim, M. Wulff, J. J. van Thor, and H. Ihee, "Protein Structural Dynamics of Photoactive Yellow Protein in Solution Revealed by Pump–Probe X-ray Solution Scattering," *Journal of the American Chemical Society*, vol. 134, pp. 3145–3153, 2 2012.

[70] H. S. Cho, F. Schotte, N. Dashdorj, J. Kyndt, R. Henning, and P. A. Anfinrud, "Picosecond Photobiology: Watching a Signaling Protein Function in Real Time via Time-Resolved Small- and Wide-Angle X-ray Scattering," *Journal of the American Chemical Society*, vol. 138, pp. 8815–8823, 7 2016.

[71] M. Cammarata, M. Lorenc, T. K. Kim, J. H. Lee, Q. Y. Kong, E. Pontecorvo, M. Lo Russo, G. Schiró, A. Cupane, M. Wulff, and H. Ihee, "Impulsive solvent heating probed by picosecond x-ray diffraction," *Journal of Chemical Physics*, vol. 124, no. 12, 2006.

[72] K. S. Kjær, T. B. Van Driel, J. Kehres, K. Haldrup, D. Khakhulin, K. Bechgaard, M. Cammarata, M. Wulff, T. J. Sørensen, and M. M. Nielsen, "Introducing a standard method for experimental determination of the solvent response in laser pump, X-ray probe time-resolved wide-angle X-ray scattering experiments on systems in solution," *Physical Chemistry Chemical Physics*, vol. 15, pp. 15003–15016, 9 2013.

[73] N. L. Opara, I. Mohacsi, M. Makita, D. Castano-Diez, A. Diaz, P. Juranić, M. Marsh, A. Meents, C. J. Milne, A. Mozzanica, C. Padeste, V. Panneels, M. Sikorski, S. Song, H. Stahlberg, I. Vartiainen, L. Vera, M. Wang, P. R. Willmott, and C. David, "Demonstration of femtosecond X-ray pump X-ray probe diffraction on protein crystals," *Structural Dynamics*, vol. 5, 9 2018.

[74] J. C. Spence, "XFELs for structure and dynamics in biology," *IUCrJ*, vol. 4, pp. 322–339, 2017.

[75] I. Kantor, J. C. Labiche, E. Collet, L. Siron, J. J. Thevenin, C. Ponchut, J. Borrel, T. Mairs, C. Marini, C. Strohm, O. Mathon, and S. Pascarelli, "A new detector for sub-millisecond EXAFS spectroscopy at the European Synchrotron Radiation Facility," *Journal of Synchrotron Radiation*, vol. 21, pp. 1240–1246, 11 2014.

[76] A. Casanas, R. Warshamanage, A. D. Finke, E. Panepucci, V. Olieric, A. Nöll, R. Tampé, S. Brandstetter, A. Förster, M. Mueller, C. Schulze-Briese, O. Bunk, and M. Wang, "EIGER detector: Application in macromolecular crystallography:," *Acta Crystallographica Section D: Structural Biology*, vol. 72, pp. 1036–1048, 9 2016.

[77] O. Glatter, "A new method for the evaluation of small-angle scattering data," *Journal of Applied Crystallography*, vol. 10, pp. 415–421, 10 1977.

[78] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 7 1948.

[79] P. V. Konarev and D. I. Svergun, "A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems," *IUCrJ*, vol. 2, pp. 352–360, 4 2015.

[80] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, "Challenges in protein-folding simulations," *Nature Physics*, vol. 6, no. 10, pp. 751–758, 2010.

[81] R. Schneider, A. R. Sharma, and A. Rai, "Introduction to molecular dynamics," 2008.

[82] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models," *Journal of Chemical Theory and Computation*, vol. 6, pp. 459–466, 2 2010.

[83] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," 2002.

[84] L. T. Da, F. K. Sheong, D. A. Silva, and X. Huang, "Application of markov state models to simulate long timescale dynamics of biological macromolecules," *Advances in Experimental Medicine and Biology*, vol. 805, pp. 29–66, 2014.

[85] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," 4 2009.

[86] H. Lei and Y. Duan, "Improved sampling methods for molecular simulation," 4 2007.

[87] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," *Computer Physics Communications*, vol. 180, pp. 1961–1972, 2 2009.

[88] M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. M. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer,

D. W. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, and A. White, "Promoting transparency and reproducibility in enhanced molecular simulations," 8 2019.

[89] S. Ahn, K. H. Kim, Y. Kim, J. Kim, and H. Ihee, "Protein Tertiary Structural Changes Visualized by Time-Resolved X-ray Solution Scattering," *The Journal of Physical Chemistry B*, vol. 113, pp. 13131–13133, 10 2009.

[90] O. Fisette, P. Lagüe, S. Gagné, and S. Morin, "Synergistic applications of MD and NMR for the study of biological systems," 2012.

[91] H. Liu, A. Hexemer, and P. H. Zwart, "The Small Angle Scattering ToolBox (SASTBX): An open-source software for biomolecular small-angle scattering," *Journal of Applied Crystallography*, vol. 45, pp. 587–593, 6 2012.

[92] C. J. Knight and J. S. Hub, "WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics," *Nucleic Acids Research*, vol. 43, pp. W225–W230, 7 2015.

[93] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, "Structural characterization of flexible proteins using small-angle X-ray scattering," *Journal of the American Chemical Society*, vol. 129, no. 17, pp. 5656–5664, 2007.

[94] T. Uzawa, C. Nishimura, S. Akiyama, K. Ishimori, S. Takahashi, H. J. Dyson, and P. E. Wright, "Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 13859–64, 9 2008.

[95] J. T. Lecomte, S. F. Sukits, S. Bhattacharya, and C. J. Falzone, "Conformational properties of native sperm whale apomyoglobin in solution," *Protein Science*, vol. 8, no. 7, pp. 1484–1491, 1999.

[96] J. Yao, J. Chung, D. Eliezer, P. E. Wright, and H. J. Dyson, "NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding," *Biochemistry*, vol. 40, no. 12, pp. 3561–3571, 2001.

[97] F. M. Hughson, P. E. Wright, and R. L. Baldwin, "Structural characterization of a partly folded apomyoglobin intermediate.," *Science (New York, N.Y.)*, vol. 249, pp. 1544–8, 9 1990.

[98] D. W. Meinhold and P. E. Wright, "Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 9078–83, 5 2011.

[99] D. Eliezer and P. E. Wright, "Is Apomyoglobin a Molten Globule? Structural Characterization by NMR," *Journal of Molecular Biology*, vol. 263, pp. 531–538, 11 1996.

[100] J. W. Kelly, "Towards an understanding of amyloidogenesis," *Nature Structural Biology*, vol. 9, pp. 323–325, 5 2002.

[101] H. Gong, X. Yang, Y. Zhao, R. Petersen, X. Liu, Y. Liu, and K. Huang, "Amyloidogenicity of p53: A Hidden Link Between Protein Misfolding and Cancer," *Current Protein & Peptide Science*, vol. 16, no. 2, pp. 135–146, 2014.

[102] A. Rodriguez-Gonzalez, T. Lin, A. K. Ikeda, T. Simms-Waldrip, C. Fu, and K. M. Sakamoto, "Role of the aggresome pathway in cancer: Targeting histone deacetylase 6-dependent protein degradation," 4 2008.

[103] D. Eliezer, J. Yao, H. J. Dyson, and P. E. Wright, "Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding," *Nature Structural Biology*, vol. 18, pp. 148–155, 2 1998.

[104] O. B. Ptitsyn and K.-L. H. Ting, "Non-functional conserved residues in globins and their possible role as a folding nucleus 1 1Edited by F. E. Cohen," *Journal of Molecular Biology*, vol. 291, pp. 671–682, 8 1999.

[105] M. S. Kay and R. L. Baldwin, "Packing interactions in the apomyglobin folding intermediate," *Nature Structural Biology*, vol. 3, pp. 439–445, 5 1996.

[106] A. E. Dyuysekina, D. A. Dolgikh, E. N. Samatova Baryshnikova, E. I. Tiktopulo, V. A. Balobanov, and V. E. Bychkova, "pH-induced equilibrium unfolding of apomyoglobin: substitutions at conserved Trp14 and Met131 and non-conserved Val17 positions.," *Biochemistry. Biokhimiia*, vol. 73, pp. 693–701, 6 2008.

[107] F. M. Hughson and R. L. Baldwin, "Use of site-directed mutagenesis to destabilize native apomyoglobin relative to folding intermediates," *Biochemistry*, vol. 28, pp. 4415–4422, 5 1989.

[108] P. A. Jennings and P. E. Wright, "Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin," *Science*, vol. 262, no. 5135, pp. 892–896, 1993.

[109] T. Kiefhaber and R. L. Baldwin, "Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, pp. 2657–61, 3 1995.

[110] S. K. Jha and J. B. Udgaonkar, "Direct evidence for a dry molten globule intermediate during the unfolding of a small protein.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 12289–94, 7 2009.

[111] B. D. Armstrong, J. Choi, C. López, D. A. Wesener, W. Hubbell, S. Cavagnero, and S. Han, "Site-specific hydration dynamics in the nonpolar core of a molten globule by dynamic nuclear polarization of water," *Journal of the American Chemical Society*, vol. 133, pp. 5987–5995, 4 2011.

[112] D. Barrick and R. L. Baldwin, "Three-state analysis of sperm whale apomyoglobin folding.," *Biochemistry*, vol. 32, pp. 3790–6, 4 1993.

[113] M. Kataoka, I. Nishii, T. Fujisawa, T. Ueki, F. Tokunaga, and Y. Goto, "Structural characterization of the molten globule and native states of apomyoglobin by solution x-ray scattering," *Journal of Molecular Biology*, vol. 249, no. 1, pp. 215–228, 1995.

[114] K. Gast, H. Damaschun, R. Misselwitz, M. Müller-Frohne, D. Zirwer, and G. Damaschun, "Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate," *European Biophysics Journal*, vol. 23, pp. 297–305, 10 1994.

[115] Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, and K. M. Hahn, "A genetically-encoded photoactivatable {Rac} controls the motility of living cells," *Nature*, vol. 461, pp. 104–108, 9 2009.

[116] J. M. Christie, M. Salomon, K. Nozue, M. Wada, and W. R. Briggs, "LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 8779–8783, 7 1999.

[117] R. Fedorov, I. Schlichting, E. Hartmann, T. Domratcheva, M. Fuhrmann, and P. Hegemann, "Crystal {Structures} and {Molecular} {Mechanism} of a {Light}-{Induced} {Signaling} {Switch}: {The} {Phot}-{LOV}1 {Domain} from {Chlamydomonas} reinhardtii," *Biophysical Journal*, vol. 84, pp. 2474–2482, 4 2003.

[118] K. Okajima, Y. Aihara, Y. Takayama, M. Nakajima, S. Kashojiya, T. Hikima, T. Oroguchi, A. Kobayashi, Y. Sekiguchi, M. Yamamoto, T. Suzuki, A. Nagatani, M. Nakasako, and S. Tokutomi, "Light-induced conformational changes of LOV1 (light oxygen voltage-sensing domain 1) and LOV2 relative to the kinase domain and regulation of kinase activity in Chlamydomonas phototropin," *Journal of Biological Chemistry*, vol. 289, pp. 413–422, 1 2014.

[119] K. Huang and C. F. Beck, "Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga Chlamydomonas reinhardtii," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 6269–6274, 5 2003.

[120] C. S. Im, S. Eberhard, K. Huang, C. F. Beck, and A. R. Grossman, "Phototropin involvement in the expression of genes encoding chlorophyll and carotenoid biosynthesis enzymes and LHC apoproteins in Chlamydomonas reinhardtii," *Plant Journal*, vol. 48, pp. 1–16, 10 2006.

[121] S. M. Harper, L. C. Neil, and K. H. Gardner, "Structural {Basis} of a {Phototropin} {Light} {Switch}," *Science*, vol. 301, pp. 1541–1544, 9 2003.

[122] A. S. Halavaty and K. Moffat, "N- and {C}-{Terminal} {Flanking} {Regions} {Modulate} {Light}-{Induced} {Signal} {Transduction} in the {LOV}2 {Domain} of the {Blue} {Light} {Sensor} {Phototropin} 1 from {Avena} sativa,," *Biochemistry*, vol. 46, pp. 14001–14009, 12 2007.

[123] A. S. Halavaty and K. Moffat, "Coiled-coil dimerization of the LOV2 domain of the blue-light photoreceptor phototropin 1 from Arabidopsis thaliana," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 69, pp. 1316–1321, 12 2013.

[124] E. Schleicher, R. M. Kowalczyk, C. W. M. Kay, P. Hegemann, A. Bacher, M. Fischer, R. Bittl, G. Richter, and S. Weber, "On the Reaction Mechanism of Adduct Formation in LOV Domains of the Plant Blue-Light Receptor Phototropin," *Journal of the American Chemical Society*, vol. 126, pp. 11067–11076, 9 2004.

[125] T. Kottke, J. Heberle, D. Hehn, B. Dick, and P. Hegemannt, "Phot-LOV1: Photocycle of a blue-light receptor domain from the green alga Chlamydomonas reinhardtii," *Biophysical Journal*, vol. 84, pp. 1192–1201, 2 2003.

[126] D. Matsuoka and S. Tokutomi, "Blue light-regulated molecular switch of {Ser}/{Thr} kinase in phototropin," *Proc Natl Acad Sci U S A*, vol. 102, pp. 13337–13342, 9 2005.

[127] B. D. Zoltowski, A. T. Vaidya, D. Top, J. Widom, M. W. Young, and B. R. Crane, "Structure of full-length Drosophila cryptochrome.," *Nature*, vol. 480, pp. 396–9, 11 2011.

[128] J. T. M. Kennis, I. H. M. van Stokkum, S. Crosson, M. Gauden, K. Moffat, and R. van Grondelle, "The {LOV}2 domain of phototropin: a reversible photochromic switch," *J. Am. Chem. Soc.*, vol. 126, pp. 4512–4513, 4 2004.

[129] E. F. Yee, R. P. Diensthuber, A. T. Vaidya, P. P. Borbat, C. Engelhard, J. H. Freed, R. Bittl, A. Möglich, and B. R. Crane, "Signal transduction in light-oxygen-voltage receptors lacking the adduct-forming cysteine residue," *Nat Commun*, vol. 6, p. 10079, 2015.

[130] A. Losi and W. Gärtner, "The Evolution of Flavin-Binding Photoreceptors: An Ancient Chromophore Serving Trendy Blue-Light Sensors," *Annual Review of Plant Biology*, vol. 63, pp. 49–72, 6 2012.

[131] A. T. Vaidya, C. H. Chen, J. C. Dunlap, J. J. Loros, and B. R. Crane, "Structure of a light-activated LOV protein dimer that regulates transcription," *Science Signaling*, vol. 4, 8 2011.

[132] M. Nakasako, K. Zikihara, D. Matsuoka, H. Katsura, and S. Tokutomi, "Structural Basis of the LOV1 Dimerization of Arabidopsis Phototropins 1 and 2," *Journal of Molecular Biology*, vol. 381, pp. 718–733, 9 2008.

[133] K. S. Conrad, A. M. Bilwes, and B. R. Crane, "Light-induced subunit dissociation by a light-oxygen-voltage domain photoreceptor from Rhodobacter sphaeroides," *Biochemistry*, vol. 52, pp. 378–391, 1 2013.

[134] U. Heintz and I. Schlichting, "Blue light-induced LOV domain dimerization enhances the affinity of aureochrome 1a for its target DNA sequence," *eLife*, vol. 5, 1 2016.

[135] J. M. Christie, T. E. Swartz, R. A. Bogomolni, and W. R. Briggs, "Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function," *The Plant Journal*, vol. 32, pp. 205–219, 10 2002.

[136] H. Katsura, K. Zikihara, K. Okajima, S. Yoshihara, and S. Tokutomi, "Oligomeric structure of LOV domains in Arabidopsis phototropin," *FEBS Letters*, vol. 583, pp. 526–530, 2 2009.

[137] S. Crosson and K. Moffat, "Structure of a flavin-binding plant photoreceptor domain: {Insights} into light-mediated signal transduction," *Proc Natl Acad Sci U S A*, vol. 98, no. 6, pp. 2995–3000, 2001.

[138] M. Dittrich, P. L. Freddolino, and K. Schulten, "When light falls in LOV: A quantum mechanical/molecular mechanical study of photoexcitation in phot-LOV1 of chlamydomonas reinhardtii," *Journal of Physical Chemistry B*, vol. 109, pp. 13006–13013, 7 2005.

[139] D. Nozaki, T. Iwata, T. Ishikawa, T. Todo, S. Tokutomi, and H. Kandori, "Role of Gln1029 in the photoactivation processes of the LOV2 domain in Adiantum phytochrome3," *Biochemistry*, vol. 43, pp. 8373–8379, 7 2004.

[140] A. I. Nash, W. H. Ko, S. M. Harper, and K. H. Gardner, "A conserved glutamine plays a central role in LOV domain signal transmission and Its duration," *Biochemistry*, vol. 47, pp. 13842–13849, 12 2008.

[141] P. L. Freddolino, K. H. Gardner, and K. Schulten, "Signaling mechanisms of LOV domains: New insights from molecular dynamics studies," *Photochemical and Photobiological Sciences*, vol. 12, pp. 1158–1170, 6 2013.

[142] R. P. Diensthuber, M. Bommer, T. Gleichmann, and A. Möglich, "Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators.," *Structure (London, England : 1993)*, vol. 21, pp. 1127–36, 7 2013.

[143] J. P. Zayner, C. Antoniou, and T. R. Sosnick, "The amino-terminal helix modulates light-activated conformational changes in AsLOV2," *Journal of Molecular Biology*, vol. 419, pp. 61–74, 5 2012.

[144] H. Guo, T. Kottke, P. Hegemann, and B. Dick, "The {Phot} {LOV}2 {Domain} and {Its} {Interaction} with {LOV}1," *Biophys J*, vol. 89, no. 1, pp. 402–412, 2005.

[145] C. Engelhard, R. P. DIensthuber, A. Möglich, and R. Bittl, "Blue-light reception through quaternary transitions," *Scientific Reports*, vol. 7, 12 2017.

[146] A. Möglich, R. A. Ayers, and K. Moffat, "Design and Signaling Mechanism of Light-Regulated Histidine Kinases," *Journal of Molecular Biology*, vol. 385, pp. 1433–1444, 2 2009.

[147] H. W. Park, S. T. Kim, A. Sancar, and J. Deisenhofer, "Crystal structure of DNA photolyase from Escherichia coli," *Science*, vol. 268, no. 5219, pp. 1866–1872, 1995.

[148] H. Komori, R. Masui, S. Kuramitsu, S. Yokoyama, T. Shibata, Y. Inoue, and K. Miki, "Crystal structure of thermostable DNA photolyase: pyrimidine-dimer recognition mechanism.," *Proc.Natl.Acad.Sci.USA*, vol. 98, pp. 13560–13565, 2001.

[149] K. Hitomi, L. DiTacchio, A. S. Arvai, J. Yamamoto, S.-T. Kim, T. Todo, J. a. Tainer, S. Iwai, S. Panda, and E. D. Getzoff, "Functional motifs in the (6-4) photolyase crystal structure make a comparative framework for DNA repair photolyases and clock cryptochromes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 6962–7, 4 2009.

[150] A. F. Glas, E. Kaya, S. Schneider, K. Heil, D. Fazio, M. J. Maul, and T. Carell, "DNA (6-4) photolyases reduce Dewar isomers for isomerization into (6-4) lesions.," *Journal of the American Chemical Society*, vol. 132, pp. 3254–3255, 3 2010.

[151] C. A. Brautigam, B. S. Smith, Z. Ma, M. Palnitkar, D. R. Tomchick, M. Machius, and J. Deisenhofer, "Structure of the photolyase-like domain of cryptochrome 1 from Arabidopsis thaliana.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 12142–7, 8 2004.

[152] R. Brudler, K. Hitomi, H. Daiyasu, H. Toh, K. I. Kucho, M. Ishiura, M. Kanehisa, V. A. Roberts, T. Todo, J. A. Tainer, and E. D. Getzoff, "Identification of a new cryptochrome class: Structure, function, and evolution," *Molecular Cell*, vol. 11, no. 1, pp. 59–67, 2003.

[153] A. T. Vaidya, D. Top, C. C. Manahan, J. M. Tokuda, S. Zhang, L. Pollack, M. W. Young, and B. R. Crane, "Flavin reduction activates Drosophila cryptochrome.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 20455–60, 12 2013.

[154] K. Hitomi, H. Nakamura, S. T. Kim, T. Mizukoshi, T. Ishikawa, S. Iwai, and T. Todo, "Role of Two Histidines in the (6-4) Photolyase Reaction," *Journal of Biological Chemistry*, vol. 276, no. 13, pp. 10103–10109, 2001.

[155] E. Schleicher, K. Hitomi, C. W. M. Kay, E. D. Getzoff, T. Todo, and S. Weber, "Electron nuclear double resonance differentiates complementary roles for active site histidines in (6-4) photolyase," *Journal of Biological Chemistry*, vol. 282, no. 7, pp. 4738–4747, 2007.

[156] A. Ganguly, C. C. Manahan, D. Top, E. F. Yee, C. Lin, M. W. Young, W. Thiel, and B. R. Crane, "Changes in active site histidine hydrogen bonding trigger cryptochrome activation," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 10073–10078, 9 2016.

[157] A. Sancar, "Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors," *Chemical Reviews*, vol. 103, pp. 2203–2237, 6 2003.

[158] N. Ozturk, C. P. Selby, Y. Annayev, D. Zhong, and A. Sancar, "Reaction mechanism of Drosophila cryptochrome.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 516–21, 1 2011.

[159] K. Uth and R. Sleigh, "Deregulation of the circadian clock constitutes a significant factor in tumorigenesis: A clockwork cancer. Part I: Clocks and clocking machinery," 2014.

[160] R. Neutze, "Opportunities and challenges for timeresolved studies of protein structural dynamics at X-ray free-electron lasers," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, 7 2014.

[161] T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty, and H. N. Chapman, "CrystFEL: A software suite for snapshot serial crystallography," *Journal of Applied Crystallography*, vol. 45, pp. 335–341, 2 2012.