Thesis for the degree of doctor of philosophy in the natural sciences

Detecting and Identifying Solution-structural Change in Photoactive Proteins

Alexander Björling

UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology
Göteborg, Sweden
2015

Thesis for the degree of doctor of philosophy in the natural sciences

Detecting and Identifying Solution-structural Change in Photoactive Proteins

Alexander Björling

ISBN 978-91-628-9651-5 (Print)
ISBN 978-91-628-9650-8 (Pdf)
Available online at http://hdl.handle.net/2077/40649

Department of Chemistry and Molecular Biology
Lundbergslaboratoriet
SE-40530 Göteborg
Sweden
Telephone: +46 (0)31–786 00 00

Printed by Ale tryckteam AB
Göteborg, Sweden 2015

Abstract

Conformational dynamics allow proteins to fulfil their biological roles, yet the understanding of these molecular machines is largely limited to structural snapshots. Time-resolved X-ray scattering provides a possible way of uncovering such dynamics, but hinges on the ability to trigger reactions, and most critically on the ability to interpret the resulting data. This thesis explores how such experiments can be carried out and analysed.

First, a computational tool for the interpretation of time-resolved X-ray scattering data is presented and tested. It is found that the method works for systems which undergo concerted domain movements, but fails for those where the structure changes in more subtle ways. The method is a potential starting point for systematic, molecular dynamics based interpretation schemes, and provides the most general and unbiased scheme yet.

Further, the time-resolved X-ray scattering technique is applied to the phytochrome family of light-sensing proteins. The results establish a comprehensive mechanism of signal transduction in a particular member of the family. A comparative study shows that this mechanism is also found in several other protein relatives, implying that a core structural apparatus has been conserved and reused for various purposes during evolution.

Foreword

In the ten or so years of studies and research leading up to this moment, a number of people have given me professional opportunities, large and small, that have truly changed the course of my career. These people are, in order of appearance, Patrik Andersson, Elisabet Ahlberg, Juan Feliu, and, of course, Sebastian Westenhoff.

Without Sebastian, I would not have entered the world of biophysics and biochemistry which, I think, would have been a shame. He has been a fantastic supervisor in every way, and without his professionalism and personality, I would probably not have finished what I started.

I owe a great deal to the people who have coauthored my papers. In particular, working with Stephan Niebling, Heikki Takala, and Oskar Berntsson has been both productive and extremely rewarding. I would also like to acknowledge that most of the results in this thesis would not have come about without the initiatives of Janne Ihalainen, and to thank David van der Spoel for having patience with a programming dilettante.

Along with these individuals, my colleagues at Lundbergslaboratoriet, past and present, have provided a great environment in which to do research, and looking back, I which that I had got to know more of you better.

Erik Werner, Rob Dods, and Oskar Berntsson have all spent a good chunk of time proof-reading this thesis and greatly improved it. It turns out that thesis blindness is a serious condition, and I really appreciate their help.

Lastly, my wife, Sara, and my kids, Nils and Mette, have paid the price of having an absent-minded researcher in the family, yet always been supportive, flexible, and understanding. For this I am most grateful.

Alexander Björling,

Malmö, November 2015

Publications

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

Paper I — S. Niebling, **A. Björling** and S. Westenhoff, "MARTINI bead form factors for the analysis of time-resolved X-ray scattering of proteins", *J. Appl. Cryst.* 47 (2014) 1190.

Paper II — **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.* 11 (2015) 780.

Paper III — H. Takala*, **A. Björling***, O. Berntsson, *et al.*, "Signal amplification and transduction in phytochrome photosensors", *Nature* 509 (2014) 245.

Paper IV — **A. Björling**, O. Berntsson, H. Takala, *et al.*, "Ubiquitous signaling in bacterial phytochromes", *J. Phys. Chem. Lett.* 6 (2015) 3379.

Paper V — **A. Björling**, H. Lehtivuori, *et al.*, "Structural photoconversion dynamics of a bacterial phytochrome", *In preparation.*

In addition, I have co-authored the following article, which is not included in this thesis owing to my relatively small contribution.

Paper VI — H. Takala, **A. Björling**, M. Linna, S. Westenhoff and J. I. Ihalainen, "Light-induced Changes in the Dimerization Interface of Bacteriophytochromes", *J. Biol. Chem.* 290 (2015) 16383.

Contribution report

Some of these studies (Papers III, IV, and V) involve samples of protein, cloned and produced in bacteria and purified in the laboratory at great effort. This work was carried out by other people, and my contribution was instead to plan, execute and analyse synchrotron experiments, as detailed here.

Paper I	I co-designed the project, contributed to evaluating its results, and had a major role in straightening out the methodology. I wrote most of the "Theory and methods" section.
Paper II	My supervisor and I conceived the idea together. We developed several other schemes and chose this route over the alternatives based on tests and reasoning. I implemented a working prototype, and contributed to the final implementation. I designed and executed all tests, and wrote the paper.
Paper III	I had a major role in the planning and execution of the various solution X-ray experiments. I designed, implemented and ran the analysis of the resulting data, and made a major contribution to writing the manuscript.
Paper IV	I had full responsibility for the planning and execution of the synchrotron experiments. I analysed the data and wrote the paper.
Paper V	These experiments were carried out over a long period of time, and I was involved and responsible for data collection to a varying degree. I did all X-ray data analysis, and wrote the paper.

Contents

Abbreviations

Agp1/2	Phytochrome from the bacterium *A. tumefaciens*
APS	Advanced photon source [Chicago]
ATP	Adenosine triphosphate
BioCARS	The 14-ID-B beamline at the APS
BM29	A SAXS beamline at the ESRF
BPhP	Bacterial phytochrome
CCD	Charge-coupled device, a detector type
Cph1	A phytochrome from the cyanobacterium *Synechocystis*
CpxA	A histidine kinase from *E. coli*
CRYSOL	A scattering prediction program
cSAXS	A beamline as the SLS
*Dr*BphP	*D. radiodurans* bacterial phytochrome
EnvZ	A histidine kinase from *E. coli*
ESRF	European synchrotron radiation facility [Grenoble]
GAF	A chromophore-binding domain in phytochromes, short for c**G**MP-specific phosphodiesterase / **a**denylyl cyclase / **F**hlA
GROMACS	A MD software package
Hb	Haemoglobin
HK	Histidine kinase
HK853	A histidine kinase from *T. maritima*
ID09b	A beamline at the ESRF
LAO	Lysine/arginine/ornithine-binding protein

LCLS	Linac coherent light source [Stanford]
LED	Light-emitting diode
`libwaxs`	A set of python classes for data reduction and analysis
Mb	Myoglobin
MD	Molecular dynamics
NMR	Nuclear magnetic resonance
PAS	A chromophore-binding domain in phytochromes, short for **P**er-**A**RNT-**S**im
PDB	Protein data bank
PHY	A domain in the PAS-GAF-PHY photosensory core of phytochromes, short for **Phy**tochrome-specific
RMSD	Root-mean-square deviation
RR	Response regulator
*Sa*BphP1/2	*S. aurantiaca* bacterial phytochromes 1 and 2
SASE	Self-amplified spontaneous emission
SASTBX	Small-angle scattering toolbox, a computer program
SAXS	Small-angle X-ray scattering
SLS	Swiss light source
VicK	A histidine kinase from *S. mutans*
WAXS	Wide-angle X-ray scattering
XFEL	X-ray free electron laser

Chapter 1

Introduction

This thesis deals with structural change in proteins, the process by which a macromolecule transfers from one conformation to another. This happens all the time in living organisms. At equilibrium, proteins wander aimlessly along pre-determined conformational paths, performing perpetual exercises which are written into the very genetic code that also prescribes their structure and their function [1, 2]. Biomolecular catalysts, which are vital to all organisms, undergo well-defined movements which allow them to fill their purposes with efficiency and specificity [3]. Yet, structural biology has rather little to say about the subject. At the time of writing, the Protein data bank (PDB) contains 113331 structures, most of which were determined from X-ray crystallography. Determining a hundred thousand protein structures is no mean feat, and these structures have revolutionised how we understand the building blocks of life. But if one wishes to understand how these complex molecular machines work, then static snapshots are not enough. It is then necessary to directly and experimentally study their dynamics [4].

In this work, my colleagues and I have contributed to this effort. We have developed experimental and analytical techniques which can be used for studying the structural dynamics of a subset of all natural proteins, those which depend on light stimuli. Before us, others have carried out similar studies of proteins which are incidentally light-sensitive in ways unrelated to their biological function [5], or which serve to harvest the energy content of incident light [6]. In contrast, we apply these methods to the signalling proteins which inform plants, fungi, and bacteria of their surroundings, the "eyes" of these blind organisms.

The dynamics of light-dependent proteins are usually studied by means of time-resolved absorption spectroscopy. While certainly useful, spectroscopic methods report on molecules' electronic or vibrational energy levels, which are only indirectly related to molecular structure. X-rays, on the other hand, interact with matter in a way which depends directly on the spatial arrangement of its electrons [7]. If X-ray scattering experiments are done with the protein

in solution, then the protein is free to move as it would, for example, in the cell itself. At certain X-ray facilities, scattered X-rays can also be measured with very high time resolution. This makes the technique of time-resolved X-ray scattering ideal for experimental studies of protein structural dynamics. It is, however, a rather new and untested method, and the aim of this work is to further its applicability.

The work that I describe here can be divided into two parts, in the title referred to as *detection* and *identification*. The first is the experimental process of observing structural change as a time-dependent X-ray scattering signal. For light-sensitive proteins, this amounts to setting up an experiment where light in the form of laser pulses can be delivered to a protein sample, which is simultaneously monitored in a controlled way using an X-ray beam. All this must be done at large-scale synchrotron facilities, where the high X-ray intensities needed for detecting these weak signals are available. The second aspect encompasses what comes after the experiment, when data encrypted by the interference process and by the random nature of liquid samples are to be understood. The task is to translate data with a low information content into structurally meaningful insights, and is currently the bottleneck in this type of study.

The research leading up to this thesis was carried out in an erratic and iterative way, with the dead ends and surprising findings that seem to be inherent to basic research. While I in principle think that science should not be narrated [8], I fear that a chronological account of my PhD work would be both uninteresting and hard to follow. I have therefore chosen to summarise and discuss my research papers in a way which is intuitive to me, and interwoven this with theory and background in the following thematic chapters.

Chapter 2 explains how X-ray scattering works, and why measuring scattered X-rays gives structural information on molecular length scales. It describes how X-ray scattering patterns can be calculated from trial structures, and explains why this is a necessary but resource-intensive operation. In this context, Paper I, which explores how coarse-grained representations of protein structures can be used to accelerate structural interpretation, is summarised.

Chapter 3 describes how time-resolved X-ray scattering can be *detected*. The chapter is mainly practical, and describes how lasers and X-rays are synchronised, how dynamics can be triggered, and how time-resolved data can be recorded and treated. The sensitivity of the technique is discussed, as are its major experimental pitfalls. The choices made in the included research papers are motivated.

Chapter 4 picks up where the last chapter left off, and outlines what can be done with the data once recorded. The theoretical possibility of correctly *identifying* structural movements from information-poor scattering data is discussed, and an overview is given over the approaches previously taken. This is a natural place to summarise Paper II, in which a general and rather unbiased method for interpreting time-resolved solution X-ray data is proposed and implemented. A set of limitations to the method are encountered and

explored.

Chapter 5 describes the application of the technique to the phytochrome family of light-sensing proteins. Building on the previous chapters, it summarises and puts in context Papers III, IV, and V, all of which are solution-structural investigations into phytochrome signalling.

At the end, **Chapter 6** is an attempt to sum up, to conclude, and to provide an outlook.

Chapter 2

X-ray scattering

This chapter introduces X-ray radiation, and explains why it is such a powerful probe for molecular structure. The chapter then explores how to predict the scattering from a solution of protein molecules, and how this computationally costly operation can be sped up, as required in iterative refinement schemes. At the end of the chapter, I show that the complicated task of X-ray scattering prediction is simplified when dealing with time-resolved data. In this way, the chapter provides theoretical background and an overview of published methods, as well as a summary of Paper I.

2.1 Fundamentals

X-rays are electromagnetic waves with wavelengths in the range of 0.1 Å to 100 Å, which corresponds to photon energies of around 100 eV to 100 keV. This highly energetic type of radiation interacts with matter in several ways, and is a useful probe throughout science and medicine. One of its most useful properties is that the wavelength is on the same length scale as interatomic distances and molecular dimensions. Therefore, radiation scattered from a sample carries information on the internal atomic or molecular structure of the material, and scattering techniques are frequently used for structural studies in physics, chemistry and biology.

In what follows, I will only consider *elastic* scattering of X-rays, where no energy is transferred from the electromagnetic wave to the sample, that is, where the wavelengths of the scattered and incident waves are equal. Further, although the phase shift of anomalous dispersion is a useful effect in crystallography [9] and even for solution studies [10, 11], the following treats only normal, instantaneous scattering.

Simple scattering can be conveniently thought of in classical terms. Essentially, an electron struck by an X-ray beam is accelerated by its electric field. According to classical electromagnetism, the resulting oscillating charge will emit a scattered wave with an amplitude given by fundamental constants [12].

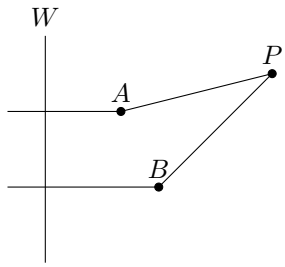


Figure 2.1: Scattering from two points

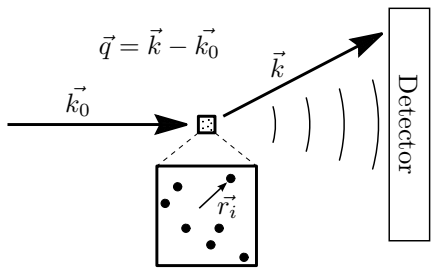


Figure 2.2: Scattering from a general sample.

When scattering originates from different parts of the sample, such as different atoms in a molecule, these scattered waves add and interfere. Consider the geometry shown in Fig. 2.1. At the point of observation P, the observed wave is the sum of the scattered waves from particles A and B. The two waves will reach P with different phases, because each has travelled a different distance from the plane front W to the point of observation. The actual phase shift will depend on the difference $|\overrightarrow{\mathrm{WA}}| + |\overrightarrow{\mathrm{AP}}| - (|\overrightarrow{\mathrm{WB}}| + |\overrightarrow{\mathrm{BP}}|)$, which for larger three-dimensional systems is a complicated expression.

Consider instead the geometry shown in Fig. 2.2, where the distance between the sample and detector is much larger than the size of the sample. As it turns out, the geometrical problem of Fig. 2.1 simplifies a great deal under this condition. An incident plane wave $\vec{k_0}$ with $|\vec{k_0}| = \frac{2\pi}{\lambda}$ (where λ is the wavelength) is scattered and observed in the direction $\vec{k}$. With all phases taken into account, a collection of spherical point scatterers at the sample position produce a scattered wave of amplitude

$$F(\vec{q}) = \sum_k f_k(\vec{q}) \exp(i\vec{q} \cdot \vec{r}_k) \quad , \tag{2.1}$$

where $\vec{r}_k$ is the position of atom k relative to some arbitrary origin, and $\vec{q} = \vec{k} - \vec{k_0}$ is the momentum transfer vector. The quantity $f_k(\vec{q})$ is the atomic scattering factor of particle k, a tabulated quantity derived from quantum-

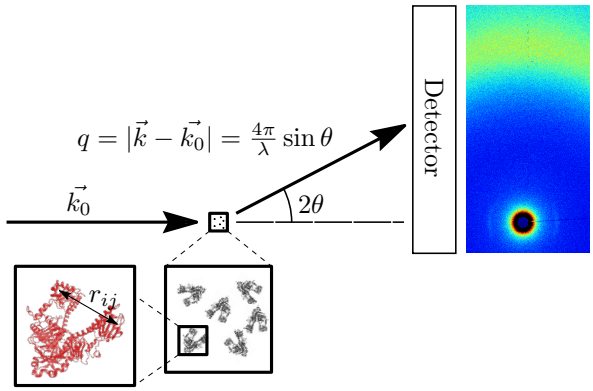


Figure 2.3: Scattering from a collection of identical but randomly oriented particles, together with a detector image of protein solution scattering. The spot on the detector image shows the beam block which catches the primary beam, and the image axes are perpendicular to the direction of the primary beam.

chemical calculations [13–15].

In any actual experiment, what is measured is not scattered amplitude but intensity, which is obtained by multiplying $F(\vec{q})$ by its complex conjugate. If $\vec{r}_{kl}$ is the vector from atom l to atom k, the total scattered intensity becomes

$$\begin{aligned} I(\vec{q}) &= F(\vec{q})F(\vec{q})^* \\ &= \left[\sum_k f_k(\vec{q}) \exp(i\vec{q}\cdot\vec{r}_k)\right]\left[\sum_l f_l(q) \exp(-i\vec{q}\cdot\vec{r}_l)\right] \\ &= \sum_k \sum_l f_k(\vec{q}) f_l(\vec{q}) \exp(i\vec{q}\cdot\vec{r}_{kl}) \quad . \end{aligned} \tag{2.2}$$

What is more, for solution experiments, one typically measures a large number of identical particles, such as protein molecules. In solution, these are usually randomly and uniformly oriented, as illustrated in Fig. 2.3. This means that one simultaneously measures all possible orientations, and the resulting detector image is spherically symmetric and essentially one-dimensional. It is convenient to represent the scattered angle as $q = |\vec{k} - \vec{k_0}| = \frac{4\pi}{\lambda} \sin\theta$, such that it is independent of the actual wavelength and detector distance used. The one-dimensional intensity is obtained by taking a spherical average, which yields the well-known Debye equation (2.3). Note that all atoms are assumed to scatter spherically, so that the quantities $f_k(\vec{q}) = f_k(q)$ are independent of

the orientation of $\vec{q}$.

$$\begin{aligned} I(q) = \langle I(\vec{q}) \rangle &= \left\langle \sum_k \sum_l f_k(\vec{q}) f_l(\vec{q}) \exp(i\vec{q} \cdot \vec{r}_{kl}) \right\rangle \\ &= \sum_k \sum_l f_k(q) f_l(q) \langle \exp(i\vec{q} \cdot \vec{r}_{kl}) \rangle \\ &= \sum_k \sum_l f_k(q) f_l(q) \frac{\sin(q r_{kl})}{q r_{kl}} \end{aligned} \tag{2.3}$$

A final complication is that, in most experiments, one measures a whole solution of the compound of interest, including a solvent and possible additives. In practice, one then records the scattered intensity both from the solution of interest (I_{soln}) and from the background solvent or buffer (I_{bg}). It is the difference of these, $I(q) = I_{\text{soln}}(q) - I_{\text{bg}}$ that must be compared to structural hypotheses for interpretation. Therefore, two additional terms must be included before taking the intensity and spherical average of Eq. (2.1). The first is the contribution from the solvent or buffer displaced by the solute. The second is that of the solvent or buffer close to the solute, where its structure most likely differs from that of the bulk [16]. Taken together, the measured intensity can be expressed as

$$\begin{aligned} I(q) &= I_{\text{soln}}(q) - I_{\text{bg}}(q) \\ &= \left\langle (F_{\text{vac}} - F_{\text{excl}} + F_{\text{solv}})(F_{\text{vac}} - F_{\text{excl}} + F_{\text{solv}})^* \right\rangle \quad , \end{aligned} \tag{2.4}$$

where the terms in parentheses represent scattered amplitudes from the molecule itself in vacuum (F_{vac}), from the excluded solvent (F_{excl}), and from the solvation layer (F_{solv}). This final expression is rather complicated, and involves individual contributions as well as all of their cross terms. In the following section, we shall see several practical approaches to overcoming this complexity.

2.2 X-ray scattering prediction

Structural interpretation of macromolecular X-ray scattering patterns is a reverse modelling process, meaning that theoretical scattering curves must be repeatedly calculated (as discussed in Chapter 4). This demands both accurate and reasonably fast ways of predicting the scattering from a given structure or structural distribution. The available methods mainly differ in how they represent the solvent contributions in Eq. (2.4), and in how they carry out the spherical average. Both these aspects appear to be critical to the quality and computational cost of the prediction. This section describes today's dominant software package, and justifies why our studies in Paper III required an alternative approach. At the end of the section, I describe state of the art methods for *ab initio* estimates of the terms of Eq. (2.4), and give an outlook.

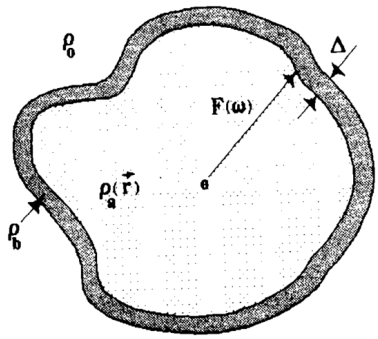


Figure 2.4: The protein solution model used in CRYSOL. Directly copied from Svergun *et al.* [17].

2.2.1 The industry standard: CRYSOL

The first major software package for predicting small-angle X-ray scattering (SAXS) patterns was CRYSOL, published in 1995 [17]. Adapted to the limited computing resources available at the time, it was designed to be fast rather than accurate. Still, it is probably the most widely used program today.

CRYSOL takes its departure in Eq. (2.4), and constructs a model where all three terms are accounted for. This model is displayed in Fig. 2.4, which is copied from the original CRYSOL paper [17]. The first term (F_{vac}) is simply a sum over all atoms as in Eq. (2.1). For the second, excluded solvent term (F_{excl}), CRYSOL uses an approximation originally introduced by Fraser, where the excluded solvent is taken into account by placing a spherical Gaussian of negative electron density at the position of each scatterer [18]. This can be done at the level of the atomic scattering factor and also gives a spherically Gaussian term in q space.

$$f_k^{\text{excl}}(q) = f_k(q) - \nu_k \rho_s \exp\left[-\nu_k^{2/3} \frac{q^2}{4\pi}\right] \quad , \tag{2.5}$$

where ν_k is a volume associated with the scatterer. This correction to the atomic form factors is both useful and popular, and we shall see it again in other scattering prediction strategies.

In the CRYSOL model, the solvation layer is described by a continuous layer of uniform density, described by the envelope function $F(\omega)$ (see Fig. 2.4). The solvation layer density is simply defined as

$$\rho(\vec{r}) = \begin{cases} \rho_b & \text{for } F(\omega) \leq |\vec{r}| \leq F(\omega) + \Delta \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

What makes CRYSOL so fast is what happens in the next step. Taking intensities and orientational averages is hugely accelerated by expanding each

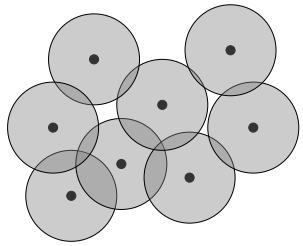


Figure 2.5: Corrected scattering factors cause overlaps and gaps.

term of Eq. (2.4) in Bessel functions j_l and spherical harmonics Y_{lm}. For example, the exponential in Eq. (2.1) is expanded as

$$\exp(i\vec{q}\cdot\vec{r}_k) = 4\pi \sum_{l=0}^{L} \sum_{m=-l}^{l} i^l j_l(qr) Y_{lm}^* Y_{lm} \quad . \tag{2.7}$$

Owing to the mutual orthogonality of spherical harmonics, the final intensity evaluation will involve nested sums over the atoms, and over the spherical harmonic indices.

$$I(q) = \sum_{k=1}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} \ldots \tag{2.8}$$

Since a decent representation of a protein requires $L \approx 15$, the l and m sums involve $(L+1)^2 = 256$ terms. Thus for N atoms, the sum has $L^2N \approx 256N$ terms, and scales with the number of atoms as $\mathcal{O}(N)$. For large molecules this evaluation, even including the overhead of the expansion, is quite efficient. It is much less time-consuming than evaluating the Debye equation (2.3), for which the number of terms scales roughly as $\mathcal{O}(N^2)$.

CRYSOL has several shortcomings, and perhaps the worst danger is the risk of over fitting. In the implementation, six quantities are treated as free fitting parameters: the solvent electron densities of the bulk solvent (ρ_0) and of the hydration shell (ρ_b), the volume displaced by each atom (ν_k), and the overall displaced volume. In addition, the experimental data can be scaled and shifted. Much of the error caused by a bad model therefore risks being absorbed into these fitting parameters [19, 20].

Further, the scattering model itself suffers from some of its approximations. Firstly, the excluded solvent correction introduces artefacts, since the spheres will both overlap and leave gaps uncovered, as schematically illustrated in Fig. 2.5. This introduces structure where there should be none. The excluded solvent amplitude should be an average over all possible solvent configurations, ultimately a homogeneous density. Instead, the excluded solvent scatterers cause artificial scattering features at high q [21]. Secondly, CRYSOL has an over-simplified view of the hydration layer, which is considered as continuous and independent of the local protein surface properties. Actually, the solvation

layer treatment has very little physical support. Thirdly, the surface representation $F(\omega)$ is far too simple to represent proteins which are not roughly spherical. For example, the surfaces of proteins which have internal cavities or external protrusions, or which consist of multiple domains connected together, cannot be represented in this way.

2.2.2 An alternative: SASTBX

A more recent development is the method introduced by Liu *et al.* [33], and implemented in the SASTBX package [32]. By this method, the exponential in Eq. (2.1) is expanded in terms of three-dimensional Zernicke polynomials (which are just convenient basis functions). Similarly to the spherical harmonics expansion described above, this greatly reduces the computational cost of the intensity evaluation, and reduces the scaling to $\mathcal{O}(N)$. In this method, the displaced solvent and the hydration layer are modelled as continuous regions on a grid of finite volume elements [33]. One result of this is that the overlaps-and-gaps problem discussed above (Fig. 2.5) is solved. Another is that the shape of the hydration layer region becomes more flexible, so that SASTBX can accommodate much more complex molecular shapes.

For the structural studies in Paper III, this last point became critical. As will be seen in later chapters, the phytochrome protein studied is strongly non-spherical, with a dimeric interface that opens up to form a solvated cleft. CRYSOL does a poor job of describing such a molecular envelope, and we used SASTBX with Zernicke expansion for all scattering calculations.

2.2.3 The bleeding edge: explicit solvent methods

Many alternatives to the spherical harmonics (or Zernicke polynomial) expansion methods have been proposed, and Table 2.1 summarises most of them. But the most ambitious development is to calculate the scattering of a biomolecule solution directly from all-atom molecular simulations. Almost accomplishing this, Grishaev *et al.* used independent frames of an equilibrium water box to find representations of the displaced solvent and hydration layer contributions [19]. They then scaled up the scattering from the bulk-structured water in the hydration layer to account for the assumed excess density there. The results were scattering curves which better matched experimental data. But more importantly, their method was better able to discriminate good models from randomly generated bad ones, a crucial requirement to avoid the over fitting which inevitably happens when too many free parameters are allowed to vary [20].

In an attempt to numerically explore how much information can be extracted from static SAXS and WAXS curves, Park *et al.* were the first to calculate the scattering from a complete, realistically hydrated protein molecule. They simulated two droplets, one of pure water and one which also contained a conformationally locked protein molecule, and calculated the time-averaged

Table 2.1: Overview of scattering prediction methods.

Name	Spherical averaging	Scaling	Excluded solvent	Hydration	Authors	Journal	Year	Ref.
CRYSOL	Spherical harmonics	$\mathcal{O}(N)$	Corrected $f(q)$	Uniform layer, near-spherical	Svergun *et al.*,	J. Appl. Cryst.	1995	[17]
SASSIM	Spherical harmonics	$\mathcal{O}(N)$	Corrected $f(q)$	Explicit *simulated*	Merzel *et al.*,	Acta Cryst. D	2002	[22]
ORNL_SAS	Monte Carlo	$\mathcal{O}(N^2)$	Continuous, arbitrary shape	Uniform layer, arbitrary shape	Tjioe *et al.*,	J. Appl. Cryst	2007	[23]
no name	Numerical	$\mathcal{O}(N)$	Explicit *simulated*	Explicit *simulated*	Park *et al.*,	J. Chem. Phys.	2009	[24]
Fast-SAXS	Coarse-grained Debye	$\mathcal{O}(N^2)$	Corrected $f(q)$	Scaled explicit *bulk* water	Yang *et al.*,	Biophys. J.	2009	[25, 26]
SoftWAXS	Numerical	$\mathcal{O}(N)$	Continuous, arbitrary shape	Uniform layer, arbitrary shape	Bardhan *et al.*,	J. Appl. Cryst.	2009	[21, 27]
FoXS	Debye	$\mathcal{O}(N^2)$	Corrected $f(q)$	Corrected $f(q)$	Schneidman-Duhovny *et al.*,	Nucleic Acids Res.	2010	[28, 29]
AXES	Numerical	$\mathcal{O}(N)$	Scaled explicit *bulk*	Scaled explicit *bulk*	Grishaev *et al.*,	JACS	2010	[19]
AquaSAXS	Numerical	$\mathcal{O}(N)$	Corrected $f(q)$	Poisson-Boltzmann density	Poitevin *et al.*,	Nucleic Acids Res.	2011	[30]
HyPred	Numerical	$\mathcal{O}(N)$	Continous, arbitrary shape	Semi-empirical density	Virtanen *et al.*,	Biophys. J.	2011	[31]
SASTBX	Zernicke polynomials	$\mathcal{O}(N)$	Continuous, arbitrary shape	Uniform layer, arbitrary shape	Liu *et al.*,	Acta Cryst. A	2012	[32, 33]
WAXSiS	Numerical	$\mathcal{O}(N)$	Explicit *simulated*	Explicit *simulated*	Chen *et al.*,	Biophys. J.	2014	[20, 34]

signal. In this way, all terms of Eq. (2.4) were approximated by considering the actual solvent which would have occupied the molecule's place, and that which is accumulated or depleted at various positions around the molecule. Their approach uses no free parameters that can be varied to reproduce data, and which might therefore absorb the model's flaws. Instead, the quality of the structure, together with the quality of the simulation performed, the force field used, etc, completely determines the predicted scattering.

In a similar approach, Chen *et al.* recently launched a web server which performs MD simulations and returns the predicted scattering from simulated, completely explicit models [20, 34]. Their method is mostly equivalent to that of Park *et al.*, with the improvement that a dynamic envelope is used, so that protein flexibility can be taken into account. Thus, the idea is to validate MD trajectories by comparing the time-averaged scattering intensity to experiment.

It is easy to conclude that the ambitious methods presented in this section provide the way forward in scattering calculations. But for a SAXS or WAXS practitioner, the choice will ultimately be a practical one, with the accuracy gained by using a more advanced method weighed against the computational cost and logistical aspects. Sometimes, a rough picture is simply enough. Other times, MD simulation might not even be an option, due to exotic amino acid components or chromophores that may not be parametrised in standard MD packages. Therefore it is entirely possible that CRYSOL, from 1995, will continue to be the most commonly used program in the field.

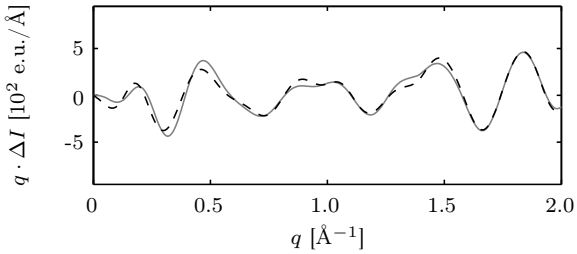
2.3 Structural change and difference scattering

This thesis deals with the structural dynamics of molecules. In solution X-ray experiments, such dynamics are tracked by measuring the change in scattered intensity which occurs during the dynamic process of interest. The process is triggered by some perturbation, typically a laser pulse, at $t = 0$. The difference scattering at some later time $t = \tau$ is then calculated from the scattered intensity by subtracting that of the unperturbed sample at $t < 0$,

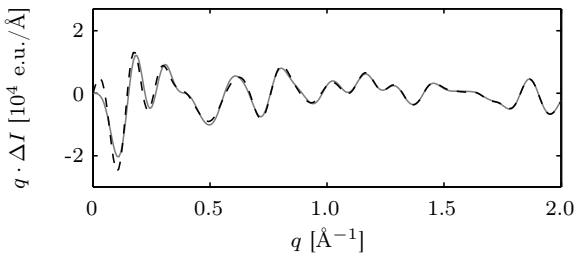
$$\Delta I(q, \tau) = I(q, \tau) - I(q, t < 0) \quad . \tag{2.9}$$

In time-resolved X-ray studies, solving the reverse problem of structural interpretation (Chapter 4) thus requires repeated calculation of the quantity $\Delta I(q, \tau)$. But now there are two trial structures, one for $t < 0$ and one for $t = \tau$. Obviously the most accurate way of calculating the difference in Eq. (2.9) is to calculate the predicted scattering from each, using the most accurate method available. But Eq. (2.9) actually affords some cancellation of error which can be exploited by using a simpler X-ray scattering calculation.

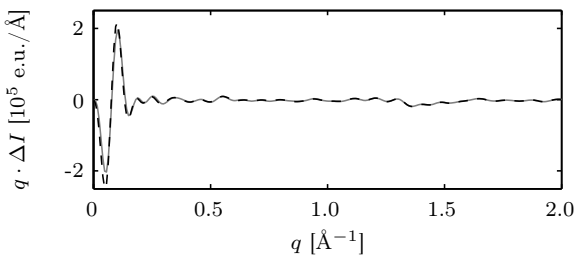
In particular, the effect of the hydration layer is negligible in the quantity $\Delta I(q, \tau)$, as shown for three test systems in Fig. 2.6. The test systems are sets of published structure pairs which represent conformational change of varying magnitudes, at 0.2 Å, 4.0 Å and 10.1 Å (RMSD). The difference scattering from



(a) Sperm whale myoglobin: CO photolysis [35]. RMSD = 0.2 Å.



(b) Human haemoglobin: CO photolysis [36, 37]. RMSD = 4.0 Å.



(c) *D. radiodurans* phytochrome: Pr→Pfr photoconversion (Paper III). RMSD = 10.1 Å.

Figure 2.6: Predicted solution difference scattering of three test systems which undergo conformational change at different length scales, calculated from atomistic models using the CRYSOL program [17]. Solid lines are calculated with the displaced solvent correction in Eq. (2.5), dashed lines additionally account for the hydration layer using the default CRYSOL settings. Reproduced from Paper I.

each system was predicted with CRYSOL, both by ignoring the hydration layer, and by accounting for it with the program's default parameters but at the highest possible resolution ($L = 50$ in Eq. (2.7)). Although the treatment of the hydration layer in CRYSOL is primitive, it is useful as a prototype for estimating the effect on $\Delta I(q, \tau)$. Evidently, at all scales of conformational change, the hydration layer scattering more or less cancels out.

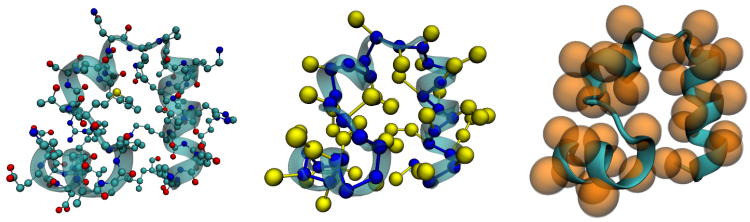
The results of Fig. 2.6 justify disregarding the hydration layer scattering in difference intensity calculations, but this must be done with prudence. If, for example, solvent-accessible surface was created or lost in the conformational change, then the error cancellation would probably become worse. This could happen if a buried interface was formed or broken in the process. One could also imagine an effect on $\Delta I(q, \tau)$ if the hydration layer remains unchanged, but that the internal charge density of the protein is rearranged, such that the protein-hydration layer cross terms are altered. For most systems, however, Fig. 2.6 is a reassuring reminder that the complexities of scattering predictions are somewhat relaxed for difference scattering.

2.4 Speeding up the calculation

Predicting scattering data from a molecular model is riddled with the above challenges and complexities. But as we shall see in Chapter 4, the reverse operation – determining structure from scattering data – is even worse. For proteins, the information encoded in a scattering curve is nowhere near what is needed to uniquely solve a structure, so structural interpretation is a reverse problem which requires testing a large number of trial structures. Needless to say, such schemes easily become computationally overwhelming, and scattering evaluations must be accelerated.

There are a number of ways to approach this problem. As we have already seen, expansion of the three-dimensional form factor (2.1) in some convenient basis set greatly accelerates calculations by simplifying the spherical average. Another strategy is to do Monte Carlo sampling of the molecule's internal distance distribution [23, 38]. There are also methods that directly speed up the analytical (but computationally expensive) spherical average of the Debye equation (2.3). One example is the approximation by Gumerov *et al.*, which reduces the $\mathcal{O}(N^2)$ complexity to $\mathcal{O}(N)$ [39], although to my knowledge, this has not been implemented in distributed software. The Debye sum can also be simplified by distance binning, or by assuming that all atomic scattering factors have the same q dependence, both of which decrease the number of time-consuming $\sin(qr)$ evaluations in Eq. (2.3) [29, 40].

Perhaps the most intuitive way of speeding up the Debye sum is to cluster atoms together and consider a coarse-grained representation of the protein. This is justified by the low-resolution nature of protein X-ray scattering, which is in practice rarely used for learning about short-range structure in proteins. Scattering calculations have been coarse-grained before, usually at amino-acid



(a) All-atom **(b)** MARTINI **(c)** Amino-acid

Figure 2.7: Coarse-grained protein representations, exemplified by the headpiece sub domain of chicken villin, PDB code 3TRW [44].

resolution [25, 40–43]. In Paper I, we developed a finer coarse-graining scheme, and evaluated the resolution limits and gain in applicability as compared to other representations. We now turn to a detailed discussion of this study.

2.4.1 MARTINI scattering factors

Our coarse-graining scheme is coincident with MARTINI, an increasingly popular representation and force field for biomolecules [45–51]. MARTINI splits each amino acid into 1 to 4 coarse beads, depending on size, as illustrated in Fig. 2.7b. A total of 20 bead types are used, and their pairwise interaction potentials parametrised based on thermodynamic data, with focus on polarity and hydrogen bonding behaviour. Although MARTINI simulations are far less reliable than those based on atomistic models, they have been applied to various biomolecular problems, particularly those involving protein-membrane interactions, and have recently resulted in several influential studies [52–54].

We chose the MARTINI level of representation for two main reasons. Firstly, as we shall see, amino acid-level resolution is not sufficient to refine structures in the range $0.4\,\text{Å}^{-1} < q < 0.8\,\text{Å}^{-1}$, where a lot of time-resolved information is encoded. Secondly, simulations in the MARTINI model is possibly a very useful vehicle for structural fitting, because coarse-graining greatly speeds up the simulations themselves (and not just the scattering calculations). Whereas the 20 MARTINI particle types are designed to cover the necessary properties in terms of size and polarity, for X-ray calculations, the number of electrons in a coarse bead and possibly the bead's internal structure is more important. Paper I therefore defines one type of scatterer for each constituent bead of all the amino acids. This defines a set of 49 scatterers, and the task is to define the scattering properties of each.

The mathematics of coarse-graining are discussed in detail in Paper I. Essentially, the form of the Debye equation (2.3) is maintained,

$$I(q) = \sum_m \sum_n F_m(q) F_n(q) \frac{\sin(qR_{mn})}{qR_{mn}} \quad , \tag{2.10}$$

where m and n now run over the coarse beads. The capital letters $F_m(q)$ denote coarse-grained scattering factors and, having decided on the form of Eq. (2.10), the task remains to calculate these. As Paper I shows, this can be done in a number of increasingly accurate ways, which takes more and more of each bead's internal structure into account. In the simplest terms possible, a form factor can be calculated by making sure that a single bead will give roughly the same scattered intensity as its constituent atoms would, with an average taken over their possible conformations. Since intensity is amplitude squared, the square of the form factor can then be expressed as follows.

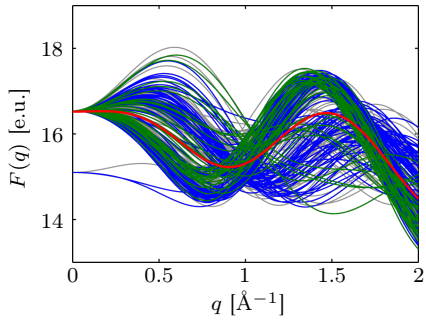
$$F_m{}^2(q) = \left\langle \sum_{k\in m} \sum_{l\in m} f_k(q) f_l(q) \frac{\sin(qr_{kl})}{qr_{kl}} \right\rangle_{\text{PDB}} \tag{2.11}$$

Here, the average $\langle ... \rangle_{\text{PDB}}$ runs over a number of instances in published structures, chosen so that all commonly occurring conformations of each bead's constituent atoms is included in the average. Such protein libraries exist, and we chose to average over RaSP50, a basis set designed to span a wide range of possible protein folds [55].

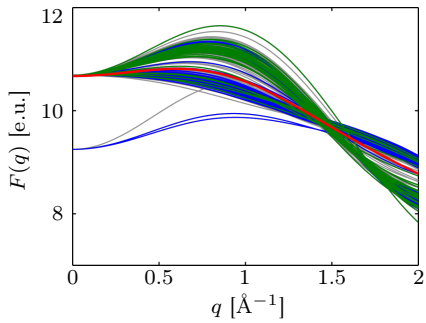
A caveat of Eq. (2.11) is that, for each q, $F_m(q)$ can be both positive and negative (but not complex, since elastic scattering is described by real scattering factors). This is important when considering interference terms between beads: all-positive scattering factors would overestimate the total intensity. The problem arises when the displaced solvent correction (2.5) is applied, because some parts of amino acids have a smaller electron density than water, and therefore the corrected amplitude will be negative. This was not acknowledged by Yang *et al.* [25], probably because it never occurred in their study, which considered entire amino acids where the electron-rich backbone always gives a net positive corrected $F(q)$. In practice, this can be appropriately corrected for. The square root of Eq. (2.11) can be taken, and the all-positive $F(q)$ corrected to produce actual form factors, as described in detail in Paper I.

The effect of the library averaging says a lot about how rough the coarse-graining approximation is. If the $F(q)$ calculated from individual instances differ considerably from the library average, then a lot of structure has been averaged away in (2.11). Fig. 2.8 shows how individual instances deviate for a particular amino acid, methionine. It clearly shows that the MARTINI representation conserves much more of the structural diversity than the amino-acid approach does. Indeed, for methionine, all side chain conformations give nearly identical form factors, while backbone beads differ a little more. On the other hand, the amino acid representation averages very different structures into the same function, inevitably worsening the approximation. Not surprisingly, this shows that the diversity in the relative arrangements of the backbone and side chain beads is more important than the structural diversity of the constituent atoms within beads.

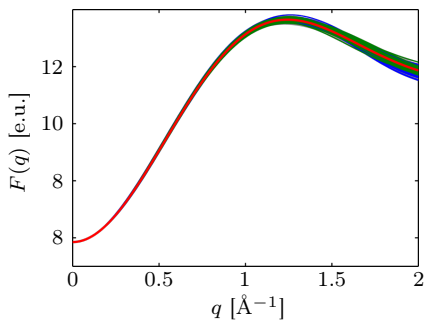
We now turn to an examination of how the effects just described affect the actual modelling of time-resolved data. As an example, in Paper I we use two



(a) Amino acid



(b) MARTINI backbone beads



(c) MARTINI side chain beads

Figure 2.8: Coarse-grained scattering factors for methionine, before and after library averaging. Red: library average, blue: from α-helix-dominated proteins, green: from β-sheet-dominated proteins, gray: from proteins with mixed content. Reproduced from Paper I.

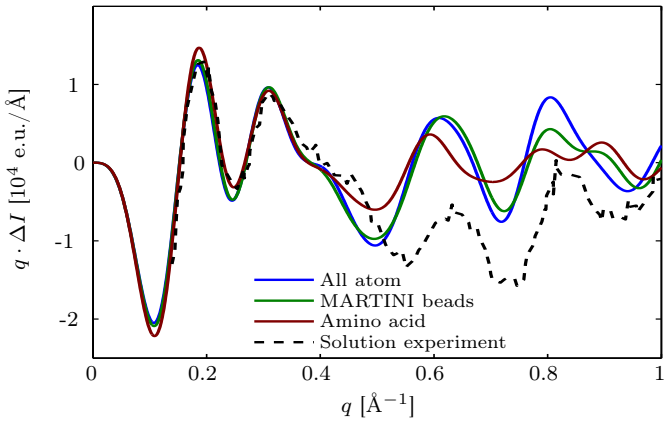


Figure 2.9: Calculated difference in solution scattering intensity between human carbon monoxy- and deoxy-haemoglobin, based on published crystal structures ($\Delta I = I_{\text{2HHB}} - I_{\text{1BBB}}$ [36, 37]). Reproduced from Paper I.

published crystal structures of human haemoglobin for a trial calculation. In one structure, the protein's heme group has a bound carbon monoxide (CO), in the other, CO has dissociated and the protein has adjusted its structure slightly. We calculated the difference of the scattering intensities from the bound and unbound crystal structures. Fig. 2.9 examines how the three representations (atomistic, MARTINI and amino acid based) compare, and also shows experimental data collected in solution.

Fig. 2.9 demonstrates that none of the crystal structure based difference curves reproduce the experimental solution data. This is not surprising, because crystal structures represent momentary protein conformations which may or may not be relevant in solution. It also shows that if one wished to refine crystal structures against this solution data, one would need a protein representation which is accurate enough in the q range used. So, since the MARTINI curve follows the atomistic curve up to around $0.8\,\text{Å}^{-1}$, this would be an acceptable level of coarse-graining. The amino acid representation, on the other hand, deviates already at $0.4\,\text{Å}^{-1}$ and is therefore too rough for this particular refinement. Clearly, the best choice of representation depends on the q range of interest and the computational complexity of the problem, and this should be an active decision when setting out to interpret scattering data.

2.5 Summary

- Predicting X-ray scattering patterns from molecular models is riddled with complications, and a host of methods have been devised to circumvent these problems.

- For difference scattering, as encountered in time-resolved studies, the situation is slightly less complicated because errors associated with the solvation layer tend to cancel out.
- Coarse-grained molecular models can be used to accelerate X-ray scattering calculations, and different levels of coarse-graining are appropriate for different situations.

Chapter 3

The time-resolved scattering experiment

This chapter explores how and why time-resolved solution X-ray experiments work. All X-ray measurements hinge on the availability of intense radiation, to allow collecting the necessary statistics in a reasonable amount of time. But this requirement is especially strict for time-resolved experiments as, in principle, the same amount of data must be collected independently for each time point. This chapter therefore begins with a brief description of how intense X-rays are generated, before outlining how the experiments are set up and performed, and how time resolution is achieved. I then describe the implications of measuring difference scattering $\Delta I(q, \tau)$, and discuss how to handle experimental pitfalls such as laser damage, X-ray damage, and concentration effects. The last section defines a set of software classes for reducing and analysing the large amounts of data spawned by the experiments, which I have developed to facilitate my PhD work.

3.1 Generation of intense X-Rays

Since the dawn of X-ray crystallography in 1912, brought about by the discoveries of Laue and coworkers [56, 57] as well as those of the Bragg family [58, 59], a lot has changed in the way that X-rays are produced. The vacuum tubes from which X-rays were first discovered in 1895 have developed over the last century, and are still used for in-house diffractometers, as well as for medical and security applications. But the biggest leap in X-ray production occurred when the need for intense radiation were aligned with the development of particle physics. It so happens, that as charged particles are accelerated, they emit electromagnetic waves according to classical physics. Thus in circular electron accelerators such as cyclotrons or synchrotrons, the orbiting electrons inevitably lose energy as radiation, probably much to the dismay of high-energy

Table 3.1: Synchrotron beamlines used for the studies included in this thesis. U, W, and BM stand for undulator, wiggler and bending magnet, respectively, whereas $\Delta\lambda/\lambda$ is the approximate spread in X-ray wavelength. BioCARS and cSAXS refer to the beamlines 14-ID-B and X12SA.

Beamline	Facility	Electron energy	Source type	Typical flux	$\Delta\lambda/\lambda$
ID09b	ESRF	6 GeV	U	$1 \cdot 10^{10}$ ph/pulse	3%
BM29	ESRF	6 GeV	BM	$1 \cdot 10^{12}$ ph/s	1%
ID23-1	ESRF	6 GeV	U	$5 \cdot 10^{12}$ ph/s	
cSAXS	SLS	2.4 GeV	U	$1 \cdot 10^{12}$ ph/s	$< 0.02\%$
BioCARS	APS	7 GeV	U	$4 \cdot 10^{10}$ ph/pulse	5%
I911-2	Max II	1.5 GeV	W	$1 \cdot 10^{11}$ ph/s	0.1%

physicists. But this radiation can also be harnessed, and light from accelerators was first utilised in 1956 [60]. By the 1980:s, synchrotron facilities started being built for the sole purpose of producing intense light for research [61, 62], and in that same decade, synchrotron X-rays became a key tool for structural biology [63].

The designated synchrotron light sources of the 1980:s, termed "2:nd generation" facilities, used bending magnets both for steering the electron beam and for producing radiation. The 1990:s saw the construction of several larger facilities which operated on a different principle. These "3:rd generation" synchrotrons were designed to hold straight magnet assemblies, called *undulators* and *wigglers*, which produce coherent, polarised and intense radiation with a tunable peak wavelength. These so-called *insertion devices* are arrays of magnets which periodically bend the electron beam path, and are designed such that the waves from each bend interfere constructively in the forward-going direction. At modern synchrotrons, electrons travelling around the storage ring are organised into bunches, with a well-defined time structure around the ring [64, Fig. 3]. One effect of this is that the radiation produced in magnetic devices is coherent, as entire electron bunches are accelerated together. Another is that the resulting X-rays emerge in pulses, usually around 100 ps in duration, which can be exploited for time-resolved studies. The work described in Papers III, IV, and V was conducted at several 3:rd generation synchrotron facilities, as summarised in 3.1.

X-ray production has continued to develop beyond synchrotron sources, as X-ray free-electron lasers (XFEL:s) are coming into operation [65–67]. These machines are linear accelerators coupled with very long undulators (132 m at the LCLS), which produce intense X-ray flashes through a process known as self-amplified spontaneous emission. This SASE process results in very short pulses, with duration on the order of 100 fs.

The XFEL sources are often praised as generally superior to synchrotrons, primarily on the basis of their peak brilliance [68, p. 4], a quantity defined

as the number of photons per second, per unit area, per unit solid angle, per unit spectral bandwidth. This quantity, which is fundamentally invariant for a given source, usually carries units of photons/s/mm^2/mrad^2/(0.1%bandwidth), and is about a billion (10^9) times higher at today's XFEL sources compared to state-of-the-art synchrotron beamlines. High peak brilliance can certainly be useful in some areas, especially for studying nonlinear phenomena, for time-resolved experiments at ultrafast timescales, or for experiments banking on the diffract-and-destroy principle [69, 70]. But in many cases, the experimental demand is to get as many total photons as possible onto the sample during the course of an allotted beam time. For that purpose, XFEL:s are perhaps better but not in a world of their own. As an example, the LCLS delivers roughly 10^{12} X-ray photons per pulse, at a repetition rate of 120 Hz, which adds up to around 10^{14} photons per second. Comparing with the synchrotron beamlines listed in Table 3.1, this is higher by two orders of magnitude, which for many experiments will not justify competing for scarce and coveted XFEL beam time, let alone the cost of building and running the machine. Hence, the many synchrotron facilities around the world are likely to stay in business for the foreseeable future.

3.2 Experimental setup

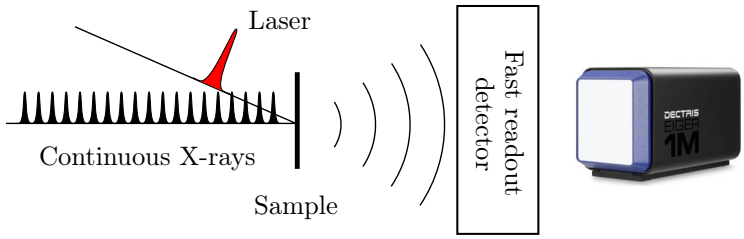
3.2.1 Detection schemes

In the work included in this thesis, we have employed two methods for recording time-resolved solution scattering at synchrotron facilities, both illustrated in Fig. 3.1. The conceptually simplest is what I will refer to as the *rapid readout* mode of acquisition, while the *pump-probe* mode offers more efficient acquisition at short time points, and therefore higher time resolution in practice.

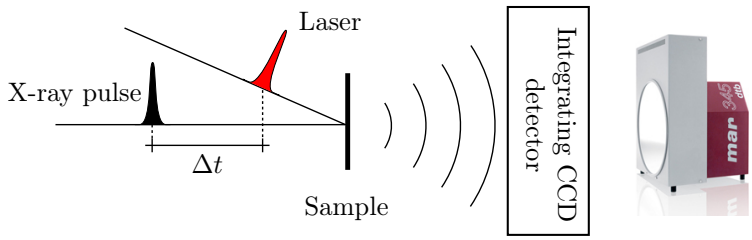
Rapid readout

In this mode, time resolution simply arises from the speed of the X-ray detector, which operates much like a video camera. The X-ray shutter is opened to let the beam through to the sample, and the fast detector records a series of frames. As illustrated in Fig. 3.1a, $t = 0$ is defined by a perturbation (usually a laser pulse) somewhere in the exposure train. As also shown in Fig. 3.1a, the X-rays are pulsed according to the time structure of the storage ring. But in practice, the pulses arrive at a rate much higher than the frame rate of the detector, and the beam can be considered to be entirely continuous in time.

The rapid readout mode is made possible by fast photon-counting detectors, which have become available in the last decade [71, 72]. Each pixel is a photoelectric silicon sensor equipped with its own amplification and counting electronics, and can therefore detect single quanta of scattered radiation. These detectors are free of dark currents and readout noise. Time resolution



(a) Rapid readout acquisition.



(b) Pump-probe acquisition.

Figure 3.1: Modes of time-resolved X-ray scattering data acquisition. The photographs show Eiger and MarCCD detectors.

is strictly limited by the time required after each exposure, to read out the counts from each pixel. For the PILATUS series of detectors [71], this is on the order of 3 ms, while the newer Eiger detectors [72] read out in as little as 4 µs. This means that these detector types can in principle record a frame every 10 ms and 10 µs, respectively, and still maintain a photon-counting duty cycle above 50%.

The next limitation to rapid readout time resolution is data transfer, the rate at which the assembled images can be transferred over the network to a storage server. The PILATUS detectors transfer the last exposure as a new one is acquired on the pixels, and so the transfer limit is the rate at which an entire detector image can be uploaded. For smaller detectors such as the 300K-W model (300 000 pixels) used in Papers III and V, this is much faster than 10 ms, and does not limit the time resolution. For the larger 2M model (2×10^6 pixels) used in Paper IV, the images were larger, and data transfer limited the time resolution to 40 ms.

The newer Eiger detectors have a built-in memory, so that frames can be acquired faster than they can be transferred. This allows for much higher frame rates than for PILATUS instruments, although too high average frame rates will eventually fill the memory and limit the overall duty cycle of the experiment. After a fast series of frames, the experiment will have to wait for transfer anyway. A solution is to use an external timing system, and to trigger

the Eiger detector at unevenly spaced time points, with high resolution at the beginning and lower at the end of the frame series. Although my PhD work has involved one experiment using a prototype single-module Eiger detector with 500 000 pixels, this data was ultimately not used in any of the included papers.

Despite these technical limitations, for both types of rapid readout detector, the actual limit is perhaps that imposed by the overall repetition rate, the rate at which the process under study is actually triggered. For high time-resolution to be meaningful, enough photons must be collected from each time point to properly resolve the signal. Since a short time point can only be measured in a short window of time, this means that the experiment has to revisit that time point frequently in order to resolve it. Thus, continuous X-ray exposure with fast readout is best suited for measuring kinetics on the range of time scales from milliseconds to seconds. The first rapid-readout measurement of protein dynamics, applied to proteorhodopsin [73, 74], had a time resolution of 10 ms. In papers III, IV, and V, this mode of acquisition was used for measurements ranging from 10 ms all the way up to the long time limit of the steady state. These measurements were all conducted at the cSAXS beamline (see Table 3.1).

Pump-probe

In the pump-probe acquisition mode, the perturbation is a laser flash which is synchronised with the X-ray source in order to achieve time resolution, as illustrated in Fig. 3.1b. The mode is implemented at the beamlines ID09b and BioCARS (Table 3.1), where intense and slightly polychromatic (so-called pink) beams are produced in undulators [64, 75]. The beam is passed through an optical line which consists of a heat load chopper, a millisecond shutter, and a rotating chopper which is fast enough to select single X-ray pulses from the storage ring's time structure [76]. Advanced control systems keep the rotating chopper in phase with the storage ring, continuously correcting for drifts and monitoring its phase jitter. A femto-, pico-, or nanosecond laser system is synchronised with the X-ray pulses, and offset backwards by a delay Δt which defines the time point of interest. By measuring over several delays, a time-resolved data series is produced.

In this way, all the measuring power of the beamline can be directed at just one delay. This circumvents the problem just discussed for the rapid readout mode, where a small fraction of the X-ray photons are devoted to each short time point. Since all scattered photons reflect the same dynamic state, they can be recorded on an integrating CCD detector, which only has to be read out once it reaches the end of its dynamic range. CCD detectors are not photon-counters, and are affected by dark currents and readout noise. However, the X-ray pulses in these setups are so intense that the counting electronics of the PILATUS or Eiger detectors could not keep up with the peak photon flux.

At both the BioCARS and ID09b beamlines, single-bunch pulses can be

produced at 1 kHz. Even at this rate, only a very small fraction of the X-ray pulses produced in the undulator is transmitted. For example, the APS in 24-bunch mode operates at 6.7 MHz, so only one in around 6700 X-ray pulses ever reaches the sample. Table 3.1 shows that these facilities produce X-ray pulses containing around 10^{10} photons each. At 1 kHz, this amounts to a time-averaged flux higher than that available at monochromatic beamlines (such as cSAXS). For protein experiments, however, the photophysics of the sample itself usually limit the experiment to 10 Hz, which brings down the average flux to an order of magnitude below cSAXS. On the other hand, all these photons are focused at a specific time point. In addition, the fast chopper can be configured to transmit more than one X-ray pulse, which can add another order of magnitude in time-averaged flux for long delays, and again bring it up to 10^{12} photons/s. My experience from conducting several protein experiments under both regimes, is that the rapid-readout and pump-probe modes are roughly equivalent at the 10 ms time scales, with the former preferable at longer times and the latter at shorter times.

3.2.2 Sample delivery

A practical aspect of setting up an experiment is how to deliver the sample to the right spot on the X-ray beam path. If one or more laser triggers are used, these must overlap with the X-ray beam on exactly the right spot. The sample has to be held in a container which is transparent at all wavelengths involved, and it must be easily aligned. Also, as described below, both X-rays and lasers damage the protein samples, which must therefore be continuously replaced or cycled in order to dilute such damage.

The work leading up to Papers III, IV, and V has, in many ways, been a continuation of similar experiments at this department [6, 73, 77, 78]. Thus the experiments described here use thin-walled quartz capillaries ($\sim$1 mm in diameter), connected to tubes through which the sample is delivered with some sort of pump, with the whole capillary assembly mounted on motorised stages to facilitate precise positioning. But in the ten WAXS experiments leading up to this thesis (totalling 146 eight-hour shifts) the setup has been gradually refined.

Fig. 3.2 shows the current status of the sample delivery system, including the materials details. Although this aspect might seem trivial, a large fraction of the work required to prepare for time-resolved synchrotron X-ray experiments usually goes towards solving these types of practical issues. One key advance is the use of a peristaltic pump which operates in one direction. This avoids the nuisance of getting air bubbles into the system [79] as there is no air/liquid interface in the tubing. This type of pump also avoids pressure artefacts when changing pump direction, which occur with bidirectional syringe pumps and can be seen in scattered X-ray data. Lastly, it greatly simplifies sample changing, which with this configuration is just a matter of emptying the tubes and replacing the sample reservoir. Other useful details include join-

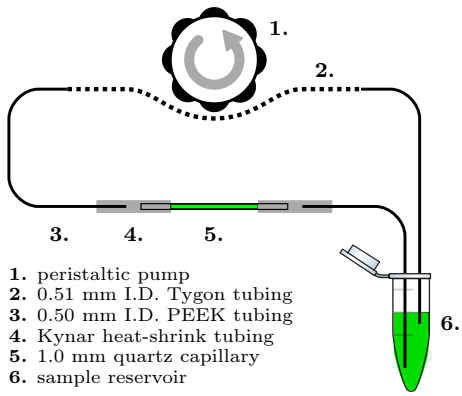


Figure 3.2: Sample environment and delivery. The total volume (reservoir excluded) is 150 µL. Reproduced from Paper IV.

ing the capillary and tubing using chemically inert heat-shrink tubing, which gives reproducible capillary assemblies that can be easily manufactured and replaced during synchrotron experiments.

With this setup a large sample volume is used (typically 500 µL), so that any damage from lasers and X-rays (see below) are diluted over a large volume. Ideally, each triggering event should strike a part of the sample which has not been been excited directly before. This assures that no kinetics are still going on in the interaction volume, and that its temperature is equilibrated so that no heat buildup occurs. With the reservoir setup sketched in Fig. 3.2, the sample is even mixed continually so that the sample is homogeneously irradiated.

In pump-probe acquisition, which in Papers III and V was done at 10 Hz repetition, the sample is continuously flowed, with the flow rate carefully matched to the repetition rate and the time points measured. If the flow is too fast, the excited sample will be flushed out of the scattering volume before the arrival of the X-rays. If the flow is too slow, the same sample will be repeatedly excited which may result in heat buildup and/or scrambled kinetics. In rapid readout acquisition, the long time points often used demand that the sample is stationary during each exposure series. Instead, the sample is moved between shutter openings by sending commands to the pump as part of the acquisition cycle. A short waiting period is necessary to let the sample stop moving and to let the pressure equilibrate, typically 1 s as used in Paper IV. For rapid readout acquisition, the sample delivery itself therefore adds dead time to the measurement, and typically decreases the duty cycle by 30-50% (see Paper IV, Supporting Information).

3.2.3 Triggering

Most time-resolved solution X-ray scattering studies published to date involve light-sensitive molecules [5–7, 73, 80–92], and the same is true for the studies presented in this thesis. The experiments described here cannot track equilibrium dynamics, but require that a large fraction of the molecules in a sample undergo a dynamic process at roughly the same time. Therefore, these experiments demand a way to suddenly trigger a dynamic chain of events. For light-driven systems, such triggering is simple. Either synchronise a laser with the X-ray source (for pump-probe acquisition), or program a delay generator to trigger a laser at a certain point during an exposure series (for rapid readout acquisition).

Moving beyond light-triggered systems is an important challenge for advancing the time-resolved X-ray technique [93]. For systems which can be triggered by the sudden addition of a new component, rapid mixing is an option. SAXS studies with moderate time resolution have been carried out in this way using stopped-flow [94] or continuous-flow mixers [95], and by using laminar diffusion mixers made from microfluidic devices [96–98]. While macroscopic flow mixers require prohibitively large samples for most protein experiments, the microfluidic approach will probably become increasingly important for understanding the dynamics of, for example, protein-ligand interactions.

Some systems which are not intrinsically light sensitive can be manipulated to react indirectly to light triggers. Again, if a system can be activated by the sudden addition of an additive, so-called *caged* compounds are sometimes an option. These molecules are engineered to release the desired component after the photolysis of a protecting group [99]. For example, caged Ca^{2+} ions can be released by illuminating a photo-labile derivative of ethylenediaminetetraacetic acid (EDTA), which then loses its affinity towards the ion. An obvious drawback of this approach, which also applies to the mixers discussed above, is that each part of the sample can only be measured once and then has to be discarded. In contrast, many (but not all) intrinsically light sensitive systems are reversible and can be measured repeatedly to resolve small signals.

Light-induced temperature jumps [100] can be used to suddenly shift temperature dependent equilibria [101], or conceivably to suddenly ramp up temperature dependent rate constants. As far as I know, this has not been successfully* applied to fast time-resolved solution scattering, but should be possible considering that the corresponding experiments have been carried out for slower processes [102, 103] and at equilibrium [79]. The conceivable time-resolution limit for this technique would be on the order of 100 ns to 1 µs, which is the time it takes for temperature and pressure to equilibrate in the interaction volume [104]. As discussed below, temperature jump experiments are in fact routinely done for subtracting the heating component of light-induced signals.

*Incidentally, the first year of my PhD was devoted to developing a model system for membrane fusion [105, 106], which could be triggered by a temperature jump. The scientific

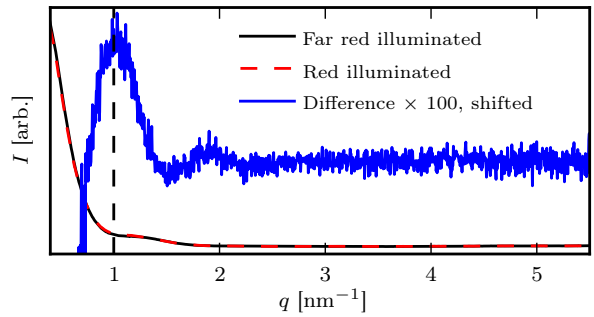


Figure 3.3: Difference scattering is a very sensitive technique, as illustrated here with data from the *D. radiodurans* phytochrome PAS-GAF-PHY fragment. Reproduced from Paper IV.

3.3 Difference scattering

3.3.1 Sensitivity and subtraction

The time-resolved scattering method measures difference scattering, as defined in Eq. (2.9). It is therefore possible to robustly detect very subtle changes in protein conformation. In fact, structural changes which are probably smaller than 1 Å can be very reproducibly detected [104]. This sensitivity is illustrated in Fig. 3.3, using a fragment of the *D. radiodurans* phytochrome (the main subject of Papers III, IV and V), where absolute detector readings of two conformational states are plotted together with their difference. In a conventional SAXS measurement, such a small difference would be prohibitively sensitive to uncertainties in concentration and scaling, and would not be resolvable.

This sensitivity has two main origins. Firstly, since the same sample and capillary position can be used for both terms of Eq. (2.9), there is no need for subtracting a separately measured (and appropriately scaled) buffer signal as in conventional SAXS, Eq. (2.4). This gives consistent data up to very high q. Secondly, any slow drifts (such as X-ray or laser damage, discussed below) are cancelled out if the difference is taken appropriately. It turns out that the best way to take this difference is to alternate measurements of the triggered process with measurements taken in the resting state. The appropriate interpolation between adjacent resting-state measurements then gives the reference to subtract for each triggered measurement. For haemoglobin photolysis, this can be accomplished by alternately running the experiment with and without laser excitation, or equivalently by interleaving positive and negative delays. For phytochrome photo switching, it can for example be done by alternating far-red and red illumination (cf. Fig. 3.3). This drift-cancellation

goal was to measure an X-ray fingerprint of the elusive intermediates of this biologically fundamental process. It never worked, as we could not instantaneously generate enough fusion events to accumulate detectable concentrations of intermediates.

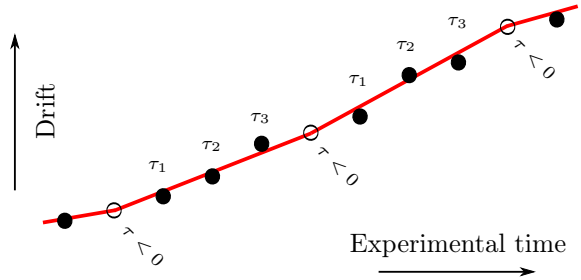


Figure 3.4: Toggled acquisition allows for near-perfect subtraction of background and drifts, as shown here for the pump-probe acquisition mode. Hollow circles represent negative time points where the sample is measured before the arrival of the excitation pulse, filled circles are positive time points. The red line represents the piece-wise interpolated reference, which is insensitive to linear drift.

in laser-toggled experiments is illustrated in Fig. 3.4.

3.3.2 Heating

Whenever a reaction is triggered by laser excitation, the sample warms up slightly. Thermal expansion in the liquid changes the distances present in the medium, and therefore affects the scattered signal. Fig. 3.5 shows the scattering from a quartz capillary filled with pure water at room temperature, as well as the difference scattering which results from moderate heating with an infrared laser. The absolute signal has a positive peak at $q = 2\,\text{Å}^{-1}$, which shifts very slightly and causes an isosbestic point in the difference curve just above this q value.

Proteins may also respond to heating, either by changing the average conformation or by broadening the distribution around it. It must therefore be shown that any signal induced by, for example, a visible laser pulse is due to relevant structural change, and not due to heating. One convenient way to accomplish this, is to carry out the same experiment but using an infrared laser as the trigger, with the laser energy adjusted so that it causes approximately the same amount of water heating signal as in the original experiment. This heating curve can then be subtracted from the light-induced difference curves at each time point, by scaling to the region $q > 1.5\,\text{Å}^{-1}$, where water scattering completely dominates over any protein contributions. What is left can then be considered to be the real structural component of the light-induced signal.

Fig. 3.6 shows a particularly demanding example. In this experiment, carbonmonoxy myoglobin (Mb–CO) was photolysed by intermittent steady-state illumination from a green laser ($\lambda = 532\,\text{nm}$). This causes the CO molecule to leave the Mb binding site, which in turn induces a tiny (RMSD <

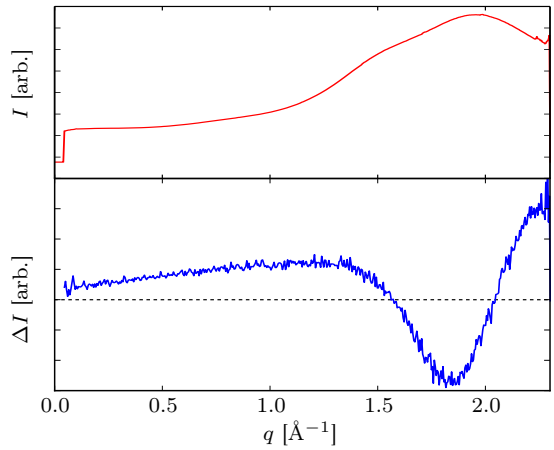


Figure 3.5: Scattering of water, and the difference scattering caused by moderate heating ($\Delta T < 10\,\mathrm{K}$).

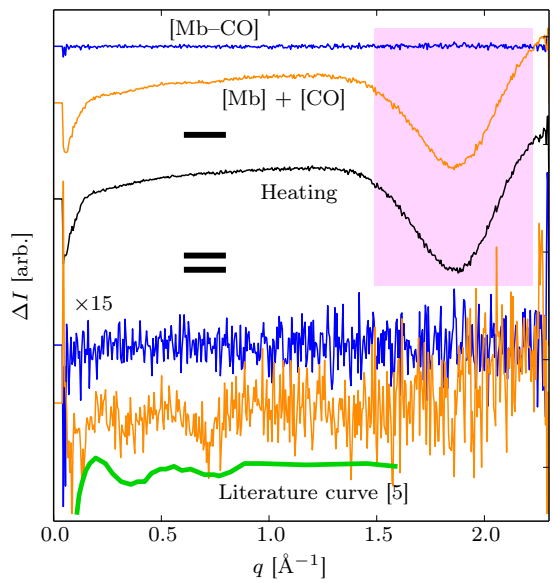


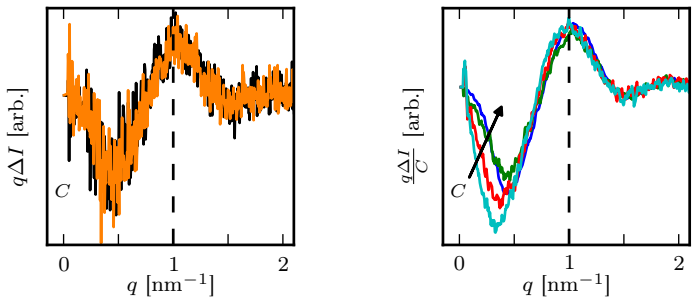
Figure 3.6: The difference scattering of a myoglobin (Mb) solution saturated with CO, before (blue) and during (orange) steady-state photolysis of the Mb–CO bond. The black curve is the pure heating response measured with an infrared laser. The noisy curves have had the heating contribution subtracted as described in the text. The green curve is a published reference from an experiment using the pump-probe acquisition mode. All curves are shifted for clarity.

1 Å) structural relaxation in the protein. However, as Mb and CO recombine very quickly [5], the sample is continuously excited and heated up. This results in two difference curves, a baseline for the [Mb–CO] complex, and a signal-containing curve for the separated solution of [Mb]+[CO], which contains a small structural contribution and a large heating signal. While the topmost orange curve in Fig. 3.6 contains weak wavy features at $q \approx 0.5\,\text{Å}^{-1}$, they are very hard to discern over the strong heating background. When the pure heating signal is subtracted, these very weak structural features appear in the orange curve, which at least resembles the reference curve shown at the bottom. In this case, pump-probe acquisition (used for acquiring the reference curve) is much more efficient, as all X-ray photons then arrive at a sample newly photolysed by a single laser pulse. However, as Fig. 3.6 shows, even this non-ideal experiment can detect very small (RMSD $< 1\,\text{Å}$) structural changes if heating subtraction is done carefully.

3.3.3 Radiation damage and concentration artefacts

An additional advantage of the difference technique is its insensitivity to X-ray radiation damage. The main effect of such damage is protein aggregation [107], often on the capillary wall, resulting in absolute data which is unfit for structural interpretation. But since only intact protein responds to illumination, the effect vanishes in the difference ΔI. Scattering from non-functional radiation-damaged protein on the wall becomes just another drift, and is cancelled by the toggled difference (Fig. 3.4). This is shown in Fig. 3.7a, where data from two individual transients of the *D. radiodurans* phytochrome fragment, separated in time by approximately 900 seconds of data acquisition and X-ray exposure, are plotted together. There is no significant difference between these curves, which means that radiation damage has not affected the outcome of the experiment.

A drawback of difference X-ray scattering is that small difference signals require high protein concentrations. In SAXS, artefacts due to inter-particle interference, repulsion or aggregation are typically seen at low q as the concentration is increased [108]. Such problems could also occur in difference measurements, and need to be ruled out for each studied system. As an example, Fig. 3.7b shows how the photoconversion signal in the *D. radiodurans* fragment changes as the sample is diluted. The difference curve clearly depends on concentration at low q. However, the main feature discussed in Papers III, IV and V, that at $q = 1/\text{nm}$, lies in a q range free from concentration artefacts. The conclusions of these studies are therefore insensitive to the high concentrations.



(a) X-ray beam damage is not a problem.

(b) High concentrations can be a problem, but mainly at low q.

Figure 3.7: Effects of radiation damage and high concentration, illustrated with data from the *D. radiodurans* phytochrome PAS-GAF-PHY fragment. See text for details. Reproduced from Paper IV.

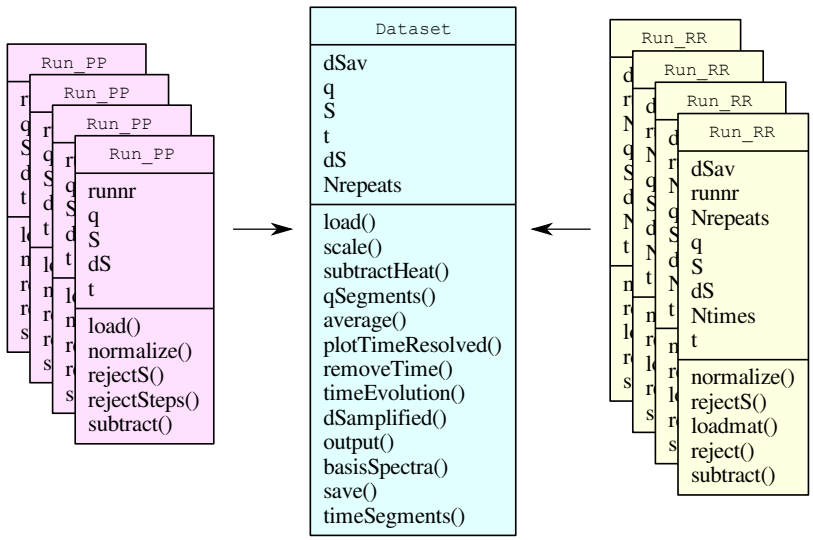


Figure 3.8: Container classes for time-resolved scattering data. Individual runs are stored in `Run_PP` (pump-probe) and `Run_RR` (rapid readout) objects, from which a universal `Dataset` object can be constructed.

3.4 Data reduction – a general implementation

To simplify the task of reducing and analysing time-resolved scattering data, I have developed a Python module[†] for data reduction, analysis and visualisation. The code, called `libwaxs`, ultimately organises all time-resolved scattering data sets into instances of the same class. Therefore, the same user scripts can be used to load and analyse data from any experiment without knowledge of the specific setup.

The `libwaxs` module consists of container classes for pump-probe and rapid-readout runs, as illustrated in Fig. 3.8. A `Run_xx` object is constructed with knowledge of the setup-specific parameters (X-ray energy, detector distance, etc). Corrected [109] and radially integrated data are loaded from beamline-specific files. Loading routines for new or modified file formats or setups are written or adapted for each synchrotron or XFEL experiment, and appended to those already in place. The `Run_xx` classes also carry methods to take the appropriate differences in Eq. (2.9), to reject outliers and to normalise the data if required.

A typical experiment generates tens or hundreds of runs before a signal is adequately resolved, and the next step in data reduction is to construct a universal `Dataset` object from such a set of runs. The final data can be represented as two-dimensional matrices $I(\tau, q)$ or $\Delta I(\tau, q)$, and vectors τ and q, regardless of the acquisition mode and experimental details. As Fig. 3.8 shows, the resulting `Dataset` object contains methods to further analyse and plot the data. For example, if a separate `Dataset` object representing heating data is available, the heating contribution can be easily removed. Data can be sliced and averaged across the τ or q axes to allow better plotting or analysis, and the time evolution of known basis signals can be extracted as done, for example, in Paper V.

3.5 Summary

- Modern synchrotron sources and detectors allow time-resolved acquisition in the "pump-probe" and "rapid readout" modes.
- The system under study must be triggered by a well-defined event. This is most easily done with a laser pulse, but other triggering methods should be developed in order to make time-resolved X-ray scattering applicable to a wider range of scientific problems.
- Difference scattering detection is very sensitive to structural change, but care must be taken to account for experimental artefacts such as heating and beam damage.
- A general software framework was developed to facilitate these analyses.

[†]Available at http://bitbucket.org/alexbjorling/libwaxs

Chapter 4

Structural interpretation

In this thesis, "structural interpretation" refers to the process of determining three-dimensional structures, at some spatial resolution, solely or partly based on solution X-ray scattering data. This chapter first discusses the information content of such data, and describes the conditions for deriving structural models from them. I then provide an overview of previous approaches to the interpretation problem for proteins, before turning to the MD-based methods used in the research leading up to this thesis. The first (from Paper III) is an approach where free or perturbed simulations are performed, and suitable structures are afterwards chosen from the trajectories. The second (from Paper II) is a more general method of biasing simulations towards structures that agree with experimental data. Finally I discuss when this method is and isn't applicable, based on two examples.

4.1 Is interpretation even possible?

When molecules in solution are measured at all orientations simultaneously (Chapter 2), the resulting scattering curves contain very little information. Exactly how small this information content is follows from communication theory [110, 111], suitably adapted to the scattering problem [112–115]. Essentially, the theory leads to an estimate of the maximal number of independent observations (n_S) held by a data set,

$$n_\mathrm{S} = \frac{q_\mathrm{max} \cdot r_\mathrm{max}}{\pi} \quad , \tag{4.1}$$

where q_max is the highest q value at which data was collected, and r_max is the largest internal distance of the scattering particle. As an example, a typical SAXS data set ($q_\mathrm{max} = 0.3\,\text{Å}^{-1}$) for bovine serum albumin ($r_\mathrm{max} = 93\,\text{Å}$) would contain at most $n_\mathrm{S} \approx 10$ data points. A time-resolved WAXS data set, such as that in Fig. 2.9 ($q_\mathrm{max} = 1\,\text{Å}^{-1}$) for human haemoglobin ($r_\mathrm{max} = 70\,\text{Å}$) contains no more than $n_\mathrm{S} \approx 20$ observations for each time point.

Under certain conditions, anisotropic X-ray scattering can be measured to squeeze slightly more information out of time-resolved experiments. This is possible because a linearly polarised laser selectively excites those molecules for which the transition dipole moment vector and the oscillating field of the laser beam are suitably aligned. Thus, if the laser energy is weak enough not to saturate excitation, and if detection is so fast that the protein molecules do not have time to tumble, then they will give rise to an anisotropic difference signal. A recent analysis shows that such data, under ideal conditions, carry exactly twice the information of an isotropic experiment [116].

Regardless of whether there are 10, 20 or 40 independent observations encoded in a data set, this number is dwarfed by the number of variables to be determined in structural interpretation. A molecule with N atoms has $3N - 5 \approx 3N$ degrees of freedom, which for any protein of relevance gives many thousands of unknowns. Of course, these degrees of freedom are not independent. For example, knowledge of a molecule's bonding topology (which for proteins mainly follows from the sequence) together with reference values for bonds and angles, can significantly reduce the number of unknowns.

Even for macromolecular crystallographic and NMR data, in which the information content is much higher than in solution X-ray data, prior chemical information is needed to find unique, acceptable structural solutions [117–121]. In fact, protein crystallography has developed into a robust framework in which to analyse the quality of the derived structural models and to avoid over-fitting [122–125].

Static SAXS data is commonly interpreted by so-called *ab initio* low-resolution modelling, or by assembling rigid high-resolution structural elements taken from crystallographic models [126, 127]. Both methods reduce the number of degrees of freedom, but usually not enough to allow unique solutions [128]. A more robust methodology for SAXS is starting to take shape, with the development of improved fitting metrics [114, 129], but it is still unclear how much structural information can actually be extracted from macromolecular SAXS curves. For time-resolved WAXS, where slightly more information is available, and where ideally the structural evolution of related intermediates is tracked, the situation is even less clear. But structural interpretation at any resolution will undoubtedly require complementary information to enable a unique solution, and the coming sections of this chapter discuss how it can be incorporated in practice.

4.2 Simple cases: small molecules

Some real-space structural information is directly available from scattering data, and if the system under study is simple enough, this might admit structural interpretation. Specifically, the distance distribution function $p(r)$ can be calculated from data by a simple transform. The continuous analogue of Eq. (2.1) gives, by taking intensities and spherically averaging, the following

complementary expressions [130].

$$I(q) \propto \int_0^\infty \mathrm{d}r\ p(r) \frac{\sin qr}{qr} \tag{4.2}$$

$$p(r) \propto \int_0^\infty \mathrm{d}q\ qr \cdot I(q) \sin(qr) \tag{4.3}$$

In practice the direct evaluation of Eq. (4.3) is an ill-posed problem which is better solved by expanding $p(r)$ in some basis set [112, 130, 131]. But in principle, some intuitively meaningful information can be straightforwardly extracted, even for large systems. For example, it can be immediately determined if a particle is compact or extended, and if a protein is folded or unfolded, or contains multiple separate domains, just by inspection of $p(r)$ [128].

For difference data, the corresponding transform gives analogous information on the *change* $\Delta p(r)$ in distance distribution [81]. Although the experimental q range is of course finite, a pragmatic approximation involving a damping factor is suggested by Kim *et al.* [92].

$$\Delta p(r) \propto \int_0^\infty \mathrm{d}q\ qr \cdot \Delta I(q,\tau) \sin(qr) \exp\left[-\alpha q^2\right] \tag{4.4}$$

Transforming time-resolved data according to Eq. (4.4) doesn't add anything new, it simply transforms $\Delta I(q,\tau)$ from q space to r space, but it does make the data more intuitive. If the molecule under study contains very few atoms, the $\Delta p(r)$ representation can directly guide interpretation. As an example, in the early study of light-induced dissociation of I_2 in CCl_4 solution by Plech *et al.* [81], $\Delta p(r)$ is directly interpreted. The first feature to appear after photolysis is the depletion of $p(r)$ at 2.7 Å, the length of the I–I bond in the ground state. At longer delays, as I_2 is re-formed and relaxed, the energy is transferred to the solvent and $\Delta p(r)$ mainly reflects how Cl–Cl distances change.

Even for marginally bigger systems, the situation is more complicated. For example, the study of 1,2-diiodoethane $(CH_2I)_2$ by Ihee *et al.* [83] uses electronic structure calculations and molecular dynamics together with spectroscopic background information, and still barely manages to distinguish between two structural hypotheses. A similar approach is taken in the more recent XFEL investigations of I_3^-, which also treat the solvent degrees of freedom with great care [88, 89].

4.3 *Whatever works:* protein approaches

Selected time-resolved solution X-ray studies of small molecules and proteins are listed in Table 4.1. These studies have one thing in common: they have all taken pragmatic routes to structural interpretation, and have included whatever system-specific knowledge that has been available. Thus, various *ad hoc*

approaches are described in the literature (including Paper III), and this section discusses some of them in order to illustrate how difficult this task actually is.

In a ground-breaking time-resolved WAXS study on Haemoglobin, Cammarata *et al.* were the first to show that such signals can even be detected for proteins. In the experiment, carbon monoxide bound to the four metal sites of the tetramer were dissociated by a green laser pulse. The ensuing structural relaxation of the protein, and the return to the initial conformation upon re-formation of the complex in the CO-saturated solution, were detected as a time-resolved set of scattering curves. While experimentally remarkable, this study only attempted rudimentary structural interpretation. For the 100 µs time point, where CO is dissociated and the protein relaxed, the data were simply compared to existing crystal structures. There are at least two such structures of carbon monoxy haemoglobin (R and R2), and one of the unbound protein (T), all of which were put through CRYSOL. The two possible unbound−bound model differences were compared with data, and the T−R2 difference curve matched qualitatively. For the 200 ns time point, where CO has dissociated but the protein is not yet relaxed, the crystal structure of the CO-bound protein (R2) was deformed as suggested by a previous model of *Myoglobin* [132], which is structurally similar to the haemoglobin monomer. Remarkably, this structural change also matched data in a rough way.

Of course, it is one thing to confirm previous structural models, and quite another to construct new ones. Thus when Andersson *et al.* [6, 74] compared data on the proton pump Bacteriorhodopsin to a wealth of crystallographic information on initial, intermediate, and late states, they found no satisfying matches. Going one step further than the aforementioned Haemoglobin study, these authors then used the available crystallographic data to construct a basis set of rigid helical movements, and refined the model by a hierarchical minimisation procedure in this low-dimensional space. In a similar study on Proteorhodopsin from the same group [74, 77], the approach was extended with MD simulations to sample physically relevant structures in the vicinity of the best-fitting rigid-body models.

A similar approach of moving secondary-structural elements around in a lower-dimensional parameter space was employed by Kim *et al.* [85] to analyse CO dissociation from dimeric (not tetrameric) Haemoglobin. These authors applied a chemically based potential energy, as well as a penalty term representing the mismatch between model and data, and minimised a weighted sum of the two with a Monte Carlo algorithm. In a similar scheme, applied to CO dissociation from Myoglobin, Ahn *et al.* [133] instead simulated the classical dynamics of such a rigid-body system. It should be noted that in both studies, refinement always starts from existing crystal structures. Thus, the solution structures of CO-unbound dimeric Haemoglobin and Myoglobin were refined from the corresponding crystal structures. It is not clear if the same structures would be found had only the initial CO-bound structures been known. This ultimately depends on the sampling power of the method, and

Table 4.1: Selected time-resolved scattering studies to date.

Compound	Time window	Interpretation	Author	Journal	Year	Ref.
I_2 in dichloromethane	100 ps – 600 ps	ΔI modeled from spectroscopic knowledge.	Neutze *et al.*,	PRL	2001	[80]
I_2 in tetrachloromethane	100 ps – 1 µs	$\Delta p(r)$ directly interpreted.	Plech *et al.*,	PRL	2004	[81]
CH_2I_2 in methanol	250 ps – 1 µs	ΔI modeled from spectroscopic knowledge.	Davidsson *et al.*,	PRL	2005	[82]
$(CH_2I)_2$ in methanol	100 ps – 3 µs	Distinction between previously suggested structural intermediates.	Ihee *et al.*,	Science	2005	[83]
Hemoglobin-CO	200 ns – 32 ms	Crude comparison to crystal structures (100 µs) and to previously suggested models (200 ns).	Cammarata *et al.*,	Nature Methods	2008	[5]
Cytochrome C	32 µs– 200 ms	Kinetics only.	"	"	"	"
Bacterio- & Proteorhodopsin	360 ns – 100 ms	Rigid body refinement in a basis set of movements defined by several crystal structures.	Andersson *et al.*,	Structure	2009	[6]
Myoglobin-CO	10 ns	Experiment-guided dynamics of rigid structural elements in a classical force field.	Ahn *et al.*,	JPC-B Lett.	2009	[133]
Myoglobin-CO	100 ps – 10 ms	Kinetics and SAXS considerations based on $I(q = 0) \propto \int \mathrm{d}q I(q)$.	Cho *et al.*,	PNAS	2010	[84]
Proteorhodopsin	10 ms – 100 ms	Kinetics only.	Westenhoff *et al.*,	Nature Methods	2010	[73]
Proteorhodopsin	2 µs– 100 ms	As in [6], followed by additional MD sampling.	Malmerberg *et al.*,	Biophys. J.	2011	[77]
Photoactive yellow protein	3 µs– 300 ms	*Ab initio* shape determination from reconstructed static $I(q)$ curves.	Kim, T.W. *et al.*,	JACS	2012	[87]
Hemoglobin *in cell*	316 ns – 10 ms	Kinetics only.	Spilotros *et al.*,	Soft Matter	2012	[86]
Hemoglobin dimer	100 ps – 10 ms	Monte Carlo sampling of rigid secondary structural elements in a hybrid potential.	Kim, H.K. *et al.*,	JACS	2012	[85]
I_3^- in three solvents	100 ps	Distinction between previously suggested structures.	Kim, H.K. *et al.*,	PRL	2013	[88]
I_3^- in methanol	100 ps – 3 µs	"	Kim, H.K. *et al.*,	ChemPhysChem	2013	[89]
Phytochrome	3 µs– 30 ms	Pairwise frame selection from MD simulations, perturbed as suggested from crystallography.	Takala *et al.*,	Nature	2014	[90]
Photosynthetic reaction center	500 fs – 280 ps	Pairwise frame selection from MD simulations, perturbed by heating the chromophore to simulate a laser-induced "protein quake".	Arnlund *et al.*,	Nature Methods	2014	[91]

the ability to uniquely select the right solution even when searching a larger region of conformational space, both of which are untested. Also, although these approaches sound very attractive, it is rather suspicious that each has not been applied to more than one data set.

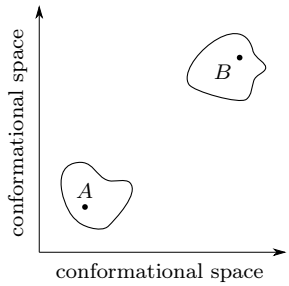
In a paper which mainly describes a new pump-probe X-ray setup at the BioCARS facility, Cho *et al.* [84] apply a more wary analysis to the Myoglobin-CO model system. Instead of refining explicit structures, they interpret the data in terms of standard SAXS parameters, specifically in terms of the integrated scattering. This quantity, proportional to $I(q = 0)$, carries information on the time-dependent volume change after excitation, and can also track the departure and re-binding of the CO molecule with reasonable precision. Kim *et al.* [87] go further in applying methods from conventional SAXS interpretation to time-resolved data, and construct low-resolution *ab initio* models for each intermediate species found in a kinetic analysis of the data. The validity of such an approach is, to the best of my knowledge, not tested on any other systems.

4.4 Selection from pre-generated ensembles

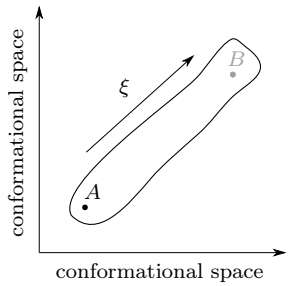
By dividing proteins into rigid elements, the above approaches reduce the excessive degrees of freedom of these molecules. An alternative way to accomplish this, is to use molecular dynamics as a vehicle for structural representation, and to let thermal sampling in a potential (a force field) remove all those conformations which are not physically accessible.

Ideally, one would like to simulate the whole dynamic process under study, and simply confirm the resulting trajectories by comparing with experiment. For the Haemoglobin-CO example, this would for instance mean running an equilibrium Hb-CO trajectory, and then removing the carbon monoxide to see what happens. In the example of light-sensing bacteriorhodopsin, it would entail perturbing the shape and charge distribution of the chromophore to mimic the excited state, and then to simulate the protein's response. In reality, this approach is unrealistic for two reasons. Firstly, it would require the empirical potential, the numerical algorithms and the starting structure to be almost perfect, which they are not. Secondly, while some processes would be accessible, most occur on time scales too long for simulation. Many quasi-independent runs would be needed, and the computational cost would simply be too high.

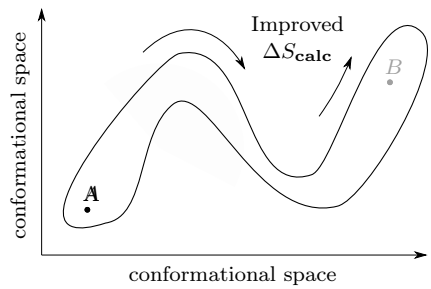
One notable exception is the recent study by Arnlund *et al.* [91], where the photosynthetic Reaction Center from *Blastochloris viridis* was violently excited with around 800 photons per molecule using a femtosecond laser, after which the ultrafast structural response was recorded at an XFEL. The process was simulated with MD by instantaneously heating up the chromophore, and then following the vibrational transfer of energy through the molecule and the solvent. The experimental and theoretical time dependencies were very



(a) Sampling around two known structures A and B.



(b) Sampling from A along a known reaction coordinate ξ.



(c) Scattering-guided MD (described below).

Figure 4.1: Structural refinement aided by MD. Encircled areas are those sampled in the trajectory.

similar, allowing a detailed interpretation of how the molecule rids itself of this excess energy.

Two other ways in which MD simulations can aid structural interpretation are illustrated in Fig. 4.1 a-b. Firstly, when two crystal structures that approximately represent the two states reflected by a solution $\Delta I(q)$ curve are known (such as for the haemoglobin model system discussed above), MD can be used for local sampling around them (Fig. 4.1a). This allows relaxing idiosyncrasies of the crystal models, which typically describe strained conformations [134]. This way of probing a small volume in conformational space in the vicinity of a given model was, for example, employed in the proteorhodopsin study mentioned above [77]. Secondly, if instead the starting model is known, and there is knowledge about the qualitative nature of structural change, then conformational space can be explored in a particular direction (Fig. 4.1b). The reaction coordinate to span can for instance be an internal distance, or the projection on some basis vector, and the MD package GROMACS contains code for running such directed simulations. Such knowledge might come from homologous proteins, from observed equilibrium fluctuations in a simulation of the starting model, or from other biophysical data such as crystallography. In this way, the solution-structural analysis in Paper III maps out a particular intra-molecular distance found from crystallography.

Once enough conformational space has been sampled through one of these schemes, theoretical difference curves can be obtained by calculating the absolute $I(q)$ for each frame of the trajectories. In Paper III, all possible pairwise differences between two such pools of curves were considered, and the best-fitting pairs selected. In reality, it is the time-averaged simulated signal which should be compared to experimental data [91]. However, averaging over a set of structures before comparison to data inevitably adds more unknowns to the problem and, as we found when performing the analysis leading up to Paper III, can actually lead to over-fitting in the space of possible structural choices.

4.5 Experimentally guided simulations

Most of the interpretation methods discussed in this chapter so far require substantial knowledge of the very process to be interpreted [5, 6, 77, 83, 88–90]. With this in mind, Paper II is an attempt to implement and test a generally applicable interpretation scheme, which requires only (i) an initial model, and (ii) difference X-ray scattering data. This scheme, which consists of biasing (or "guiding") MD simulations towards conformations which represent experimental data, is the topic of this section.

Simulations are guided towards a particular X-ray curve $\Delta I(q)$ by applying a pseudo energy on top of the normal MD potential (which typically accounts for bonds, angles, dihedrals, electrostatics, and van der Waals terms). It is important to understand that the time-resolved nature of the experiment and

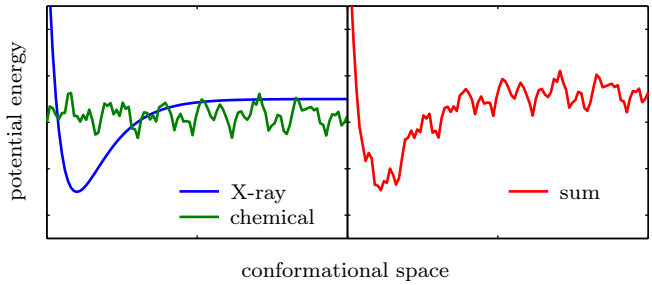


Figure 4.2: Schematic view of the energy landscape in guided MD refinement.

the dynamic nature of the simulations are separate. This scheme does not attempt to simulate the dynamics, but instead to find an artificial equilibrium which reproduces a particular scattering curve, either corresponding to a particular time point $\Delta I(q,\tau)$, or to a stable intermediate detected from kinetic decomposition of the data [6].

The experimental bias E_{XS} is defined as

$$E_{\mathrm{XS}} = \frac{k_\chi}{2} \sum_q \left[\frac{\Delta I_{\mathrm{exp}} - \alpha \Delta I_{\mathrm{calc}}}{\sigma_q} \right]^2 \quad , \tag{4.5}$$

where k_χ is a coupling parameter, σ_q the experimental error or inverse fitting weight at each q point, and α the relative triggering efficiency or yield in the experiment. The resulting force on each atom can be obtained by derivation of E_{XS} with respect to all coordinates of the molecule. This exercise is described in detail in Paper II, and yields the following force on atom k.

$$\begin{aligned} \vec{F}_{\mathrm{XS},k} =& 2k_\chi \alpha \sum_q \left\{ \frac{1}{{\sigma_q}^2} \left[\Delta I_{exp}(q) - \alpha \Delta I_{\mathrm{calc}}(q) \right] \right. \\ & \left. \times \sum_j f_k(q) f_j(q) \left[\cos qr_{kj} - \frac{\sin qr_{kj}}{qr_{kj}} \right] \frac{\vec{r}_{kj}}{{r_{kj}}^2} \right\} \end{aligned} \tag{4.6}$$

A schematic view of the resulting energy landscape is shown in Fig. 4.2, which attempts to illustrate how it is distorted or "warped" by the guiding energy. Amazingly, Eq. (4.6) says exactly in which direction to move each atom in order to reproduce experimental data. However, as we have seen, simply following these vectors in a steepest-descent minimisation would be futile as the data contains tens of observations and the model thousands of unknowns. What would happen is most likely that the molecule would assume some unphysical geometry while reproducing the input data to arbitrary precision. As already discussed, the role of the MD potential is to remove excessive degrees of freedom from the system, allowing only those with reasonable energies. But,

as also illustrated in Fig. 4.2, chemical force fields yield rough energy landscapes, and minimisation would instantly get stuck in a local minimum. Hence the use of the MD integrating algorithm, which integrates Newton's equations of motion at some relevant temperature.

In a qualitative way, the guided MD approach can be visualised as in Fig. 4.1c, where the simulation automatically makes the appropriate excursions in conformational space, based purely on the force field and the bias. So in contrast to the schemes depicted in Fig. 4.1a and Fig. 4.1b, the conformational-space direction in which to depart is not known beforehand.

Applying a bias towards X-ray scattering data is not an entirely new idea. Essentially, Grishaev *et al.* [135] have already described the pseudo-energy in Eq. (4.5) and force in Eq. (4.6). These authors described the addition of SAXS information to the analysis of NMR data, and did not investigate how Eq. (4.6) can be applied to structural change. On the other hand, Ahn *et al.* [133] applied the same equation to structural change in the carbon monoxy Myoglobin system using a rigid-body framework, but no thorough tests were published, and the implementation is not distributed. Our implementation is both tested, distributed*, and open to further development via the open-source GROMACS code base [136–138]. It should be noted here, that two papers rather similar to Paper II appeared shortly after it was published [139, 140], which shows that there is a general need for this type of interpretation tool.

A method such as guided MD refinement needs to be tested, both to make sure that the algorithm is correctly implemented and to evaluate its usefulness. Paper II describes three tests of the latter kind, and here I will discuss one of these as an example. Looking for models of global protein rearrangement, we learned that a large number of domain movements can be described in relatively simple mechanical terms [141]. The lysine/arginine/ornithine-binding protein (LAO) from *Salmonella typhimurium* is one such case, where lysine binding or detachment causes a domain to rotate around a well-defined hinge axis [142], as shown in Fig. 4.3a. The *holo* structure is crystallised with lysine, while the *apo* structure is crystallised without it.

The LAO test consisted of running simulations starting from the *holo* model, from which the lysine ligand had been manually removed. The refinement was guided by the theoretical *holo*-to-*apo* difference scattering curve, calculated using the Debye equation (2.3), and was run at several coupling strengths k_χ, chosen to give reasonable initial biasing energies $E_{\mathrm{XS}}(t = 0)$. The deviations from the starting and target structures were calculated from the finished trajectories, and are plotted in Fig. 4.3c. The panel for $30\,\mathrm{kJ\,mol^{-1}}$ shows most clearly how the structure transitions from *holo* to *apo*, with the corresponding RMSD curves simultaneously jumping up and down. At the same time, the biasing energy E_{XS} can be seen to drop in jumps. As expected, the transition occurs earlier for higher coupling strengths. At low coupling, where $E_{\mathrm{XS}}(t = 0)$ approaches the thermal energy ($N_{\mathrm{A}} \cdot k_{\mathrm{B}}T \approx 2.5\,\mathrm{kJ\,mol^{-1}}$ at

*At the time of writing, it is available as a patch at http://gerrit.gromacs.org, however it is being incorporated into the main GROMACS distribution.

298 K), the transition does not happen at all. The RMSD of the refined trajectories reach values of around 2 Å, which is typical for equilibrium fluctuations around a stable average structure [143]. The second half of the $30\,\mathrm{kJ\,mol^{-1}}$ trajectory is rendered on top of the crystallographic models in Fig. 4.3b.

The structural change of lysine detachment in LAO is rather simple. A hinge axis motion can be uniquely specified by three parameters†, which could perhaps be refined using the reaction coordinate procedure outlined above (Fig. 4.1b). The other test systems described in Paper II are of similar complexity, with essentially one intra-molecular distance describing the movement. On the other hand, this method does not presuppose knowledge of the type of movement happening, and instead this qualitative information is automatically found and output (Fig. 4.3b). Ultimately, the usefulness of guided MD refinement as a general tool depends on the ability to

(i) calculate X-ray scattering patterns with good enough accuracy,

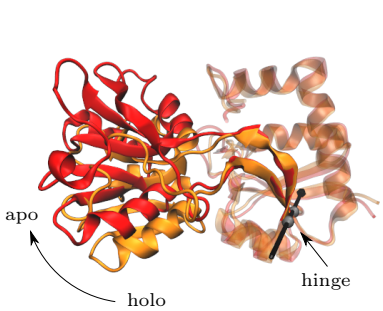
(ii) recognise the target structure, using the combined wisdom of the force field and the experimental data, and

(iii) sample so much conformational space in a reasonable amount of time, that the target structure can be found.

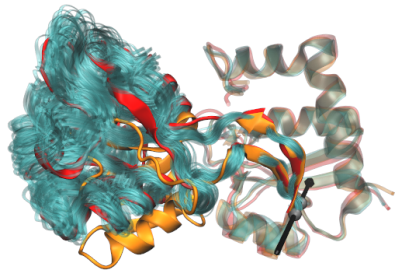
For difference scattering data, we have seen already that the Debye equation with corrected scattering factors is entirely adequate, even at coarse resolution (Chapter 2), which satisfies condition (i). At least for LAO and the other test systems in Paper II, condition (iii) is also satisfied, as 10 ns simulations converge on the target structures.

The uniqueness requirement (ii) is the most difficult to satisfy. As an example, we attempted to refine the solution structure of unbound Haemoglobin by starting from the carbon monoxy bound R2 form (with CO manually removed), and guided the simulation with difference scattering data (see Fig. 2.9). The result, shown in Fig. 4.4, was that the bias energy E_{XS} was relaxed, without convergence to the target conformation. This means that for the Haemoglobin test system, invalid structures exist which are favourable enough in the eyes of the force field, and that also approximate the input difference scattering data. Even if a good force field and extended sampling should in principle be able to weed out such false solutions, the uniqueness requirement was clearly not met for Haemoglobin in our tests. This is a discouraging result since our aim was to create the first generally applicable interpretation scheme for time-resolved scattering data. The uniqueness problem is discussed further in the next section.

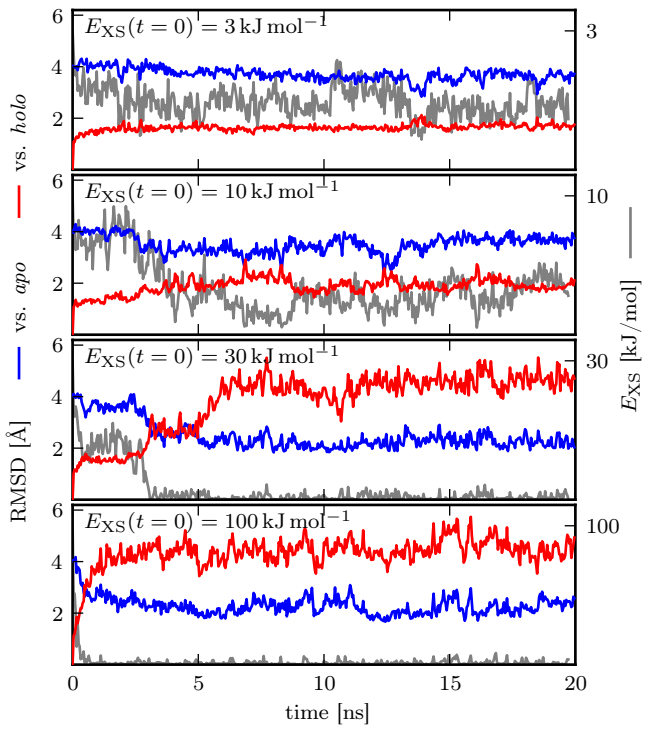
†Two points along the peptide backbone define the axis, one angle specifies the amount of rotation.



(a) Target structures [142].



(b) X-ray guided MD.



(c) Evolution of RMSD against the initial and target states, as well as of the bias energy. Different coupling k_χ as expressed as initial energies $E_{XS}(t=0)$.

Figure 4.3: Refinement of structural change upon binding of lysine to the test system, LAO.

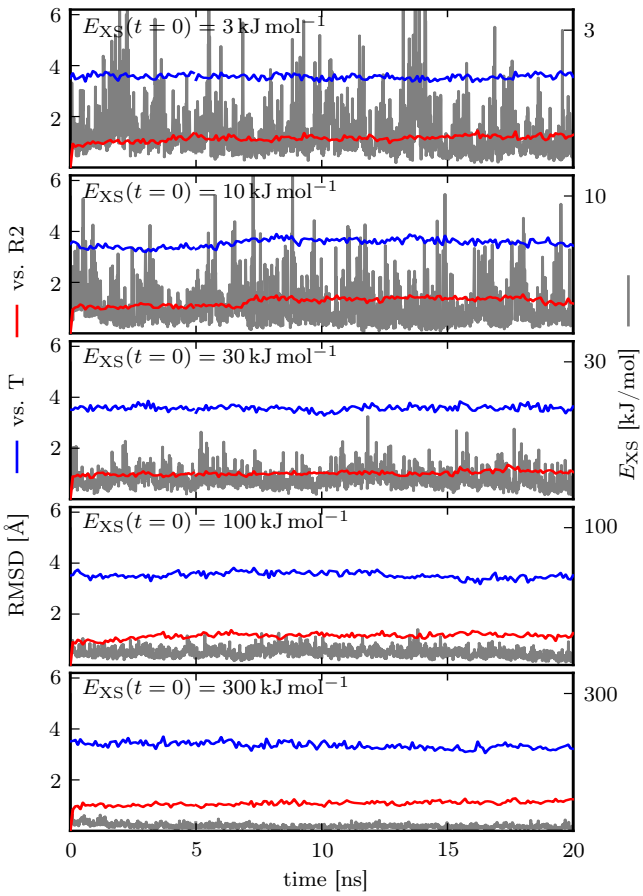


Figure 4.4: Guided MD trials for the Haemoglobin R2→T transition.

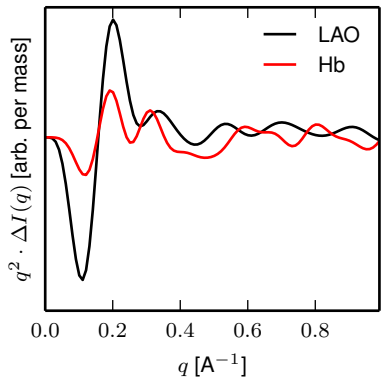


Figure 4.5: Comparison of the theoretical difference curves resulting from the transitions R→T in Hb, and *holo*→*apo* in LAO.

4.6 Uniqueness, sampling, and resolution

In Paper II, the uniqueness problem is rationalised by the argument that low-q features can be more easily assigned structural meaning than high-q features can. The idea is that there is reciprocity between the q and r scales: that small intra-molecular distances scatter at large angles, and vice versa. Because there are fewer long distances than intermediate ones in any near-spherical object, the problem would therefore be less under-determined for low-q signals. This would explain why LAO can be refined with the guided MD method whereas Haemoglobin cannot, as the signal from the former is dominated by low-q features to a greater degree than the latter (Fig. 4.5).

The notion of a direct, reciprocal relation between r and q is borrowed from crystallography [6], and warrants a critical discussion. For crystals, the Bragg law states that repeated planes separated by r will diffract constructively at first order for

$$\lambda = 2r \sin\theta \quad , \tag{4.7}$$

or equivalently,

$$qr = 2\pi \quad . \tag{4.8}$$

Of course, as we have seen, proteins in solution do not follow this condition, since there is no translational symmetry, and therefore no repeated planes to diffract. Instead, the Debye equation states that each *pair* of scatterers adds a term of the form $\sin(qr)/(qr)$ to the intensity, a function illustrated in Fig. 4.6. Therefore, all atom-atom pairs give positive contributions to the overall intensity for $qr < \pi$. At the Bragg condition $qr = 2\pi$, a Debye term contributes exactly zero to the overall intensity. At higher qr, each Debye term oscillates, and since all pairwise distance terms oscillate incoherently, the contributions largely cancel. A SAXS equivalent of Eq. (4.8) might therefore be formulated

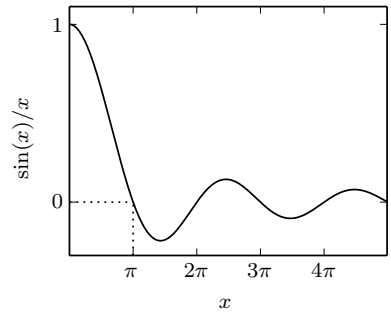


Figure 4.6: The functional form of each Debye term

as

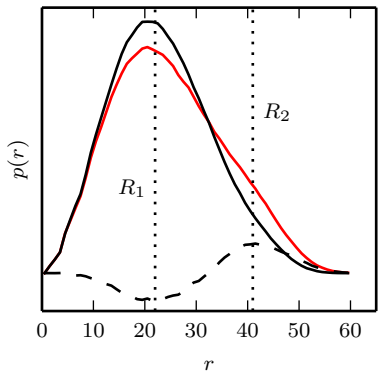
$$qr < 2\pi \quad . \tag{4.9}$$

There is thus no simple map from q to r space as is frequently claimed [6, 144]. The often-repeated rule that low-q features are more easily assigned than high-q features, which overlap to a larger degree, is also not obviously true. In fact, the low-q region is the most crowded, as each and every internal distance contributes there.

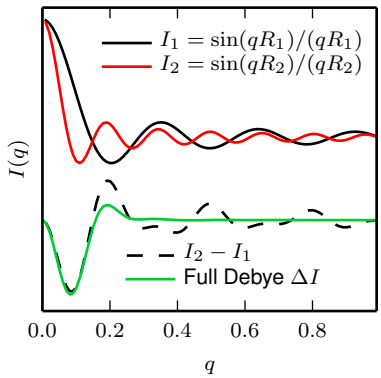
Despite this excursion, it is sometimes true that low-q features *are* more easily interpreted than those at high q, for reasons slightly different than those just described. To explain why, I will consider as examples the LAO and Haemoglobin systems. Fig. 4.7a displays the pair distribution functions of both states of LAO as well as their difference, which shows that lysine unbinding causes a shift in the distribution from R_1 to R_2. The Debye terms corresponding to these two distances are plotted in Fig. 4.7b. Apparently, very crudely approximating the difference scattering by just these two terms produces a curve which is quite similar to the full Debye sum at $q < 2\pi/R_1 \approx 0.3\,\text{Å}^{-1}$ (in the full Debye curve, oscillations at higher q are washed out as noted above).

The Haemoglobin system is analysed in the same way in Fig. 4.8. In contrast to LAO, the Hb pair distribution function is depleted at two distances (R_1 and R_3), and enriched at one distance (R_2) when the ligand departs. The three Debye terms corresponding to R_1, R_2, and R_3 generally reproduce the first features of the difference scattering curve, when weighed appropriately. While slightly more complex, this redistribution involves similar internal distances as for LAO, and the largest difference scattering signals therefore also occur at $q < 0.3\,\text{Å}^{-1}$. Indeed, when viewed as in Fig. 4.7b and 4.8b (without amplifying the y axis by q^2 as in Fig. 4.5), the Hb and LAO signals are rather similar.

This shows that the position of a difference scattering feature on the q axis alone does not decide whether or not that feature can be structurally interpreted. Comparing the Hb and LAO systems, it is obviously the magnitude

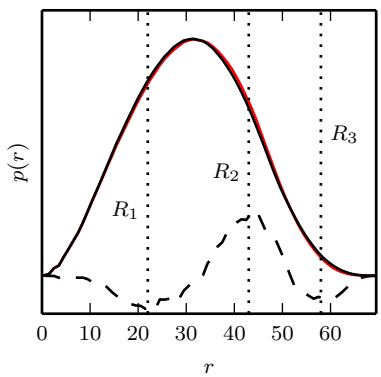


(a) Pair distribution functions of the *apo* (red) and *holo* (black) structures. Dashed curve is the difference.

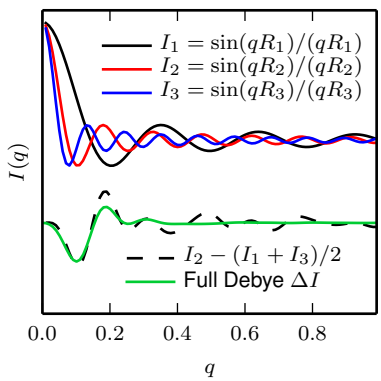


(b) Debye terms corresponding to R_1 and R_2, together with their difference and the full Debye result (both shifted).

Figure 4.7: Analysis of the LAO test system



(a) Pair distribution functions of the T (red) and R states (black). Dashed curve is the difference ×10.



(b) Debye terms corresponding to R_1, R_2, and R_3, together with their weighted difference and the full Debye result (both shifted).

Figure 4.8: Analysis of the Haemoglobin test system

of the change in $p(r)$, and not the distances involved, that is different. But when many distances change in the same way, then their corresponding Debye terms, which have similar decay lengths, will also change in the same way. These terms then add constructively at $qr < 2\pi$ where they are still fairly coherent, but destructively at higher qr, where they oscillate out of phase. This gives rise to difference signals dominated by low-q features.

Thus, the LAO signal is caused by many distances changing in the same way, whereas in Hb, it is the result of more subtle rearrangements. Guided MD easily refines the LAO process, as there are very few ways of producing such a concerted movement (and such a strong change in $p(r)$) without causing unacceptable energies from the force field. It cannot uniquely refine the Hb process, since the subtle change in $p(r)$ which reproduces $\Delta I(q)$ can be accomplished with conformations that are wrong but not sufficiently non-physical to be forbidden in the simulation.

It would be useful to define when, and to what spatial resolution, difference X-ray curves can be uniquely assigned structural meaning. But, as we have seen, position along the q scale is a poor indicator of uniqueness. Similarly, the average magnitude of the structural change also says very little. For example, the transitions in LAO and Hb discussed here correspond to comparable C-alpha RMS displacements of 4.7 Å and 3.5 Å, respectively. A simple resolution limit for unique interpretation in terms of RMSD can therefore not be found. A criterion for uniqueness, which would work as a "detection limit" for structural interpretation, would have to be based on the amount of change in $p(r)$.

As mentioned above, uniqueness and sampling are also entangled, and the tests presented here do not prove that the Hb system cannot be refined correctly. Conceivably, the false solutions found in the simulations (Fig. 4.4) represent local free energy minima. Longer simulations in accurate force fields might (and should, in principle) find the target structure. For these reasons, the results described here and in Paper II do not provide a clear criterion for uniqueness. Unfortunately, *whatever works* is still the state of affairs for structural interpretation of time-resolved solution X-ray scattering experiments.

4.7 Summary

- X-ray scattering data carry a limited amount of information, and complementary knowledge is usually required for structural interpretation.
- Various protein studies have involved *ad hoc* analyses which exploit prior structural or spectroscopic information.
- Molecular dynamics provides a useful way of restricting a molecular model to physically relevant conformations.
- Sampling around known crystal structures or along inferred reaction coordinates is sometimes useful, but relies on prior information about the structural change being studied.

- A more general algorithm of combining MD with experimental X-ray scattering data is described and tested. It is applicable to some systems, but not to others, and the reasons for this are discussed.

Chapter 5

Application to phytochromes

This chapter summarises and discusses Papers III, IV, and V, all of which revolve around a family of light-sensing proteins called phytochromes. These papers build on the methodology described in the preceding chapters. In order to put their results in a biological context, I first briefly introduce cellular signalling and the class of signalling proteins called histidine kinases. I then review the literature on phytochromes, before discussing the connections and overlaps between these two protein families. Finally, the main findings of Papers III, IV, and V are summarised and discussed one at a time.

Cells respond to environmental stimuli of many kinds, and possess complex signalling machinery to gather and manage this information, ultimately causing responses such as changes in shape or metabolism, gene expression, or cell division. A common theme in all kingdoms of life is that information is passed along as phosphate "tags", functional groups which are added and transferred in a controlled manner to and from specific protein sites [145, 146]. The process starts by phosphorylation, the catalysed transfer of a terminal (γ) phosphate from ATP to the receiving site.

In bacteria, the predominant sensing mechanism is two-component signalling, wherein a sensor histidine kinase (HK) interacts with a response regulator (RR) in a way which depends on the state of the sensor [147]. The kinase catalyses the phosphorylation of one of its own histidine residues, so-called autophosphorylation. Depending on the signalling state of the kinase, the phosphate tag is then transferred to the response regulator, which relays the signal further or directly causes the desired biochemical output by binding to, for example, DNA [148]. This chapter mainly deals with a particular bacterial phytochrome, that from *D. radiodurans*, which we have studied in a number of ways throughout Papers III, IV, V, and VI. As we shall see, understanding bacterial phytochromes is made easier by putting them in the context of histidine kinases and two-component signalling, whether they belong there or not.

The primary driving force behind phytochrome research has been to gain

better understanding and control over plants, especially those grown as crops. Indeed, phytochromes were discovered at U.S. federal institutions devoted to agricultural and food research. This aspect of phytochrome research is still valid today, and phytochrome-based modifications can be made to increase the yield in farming, by reallocating the plants' internal energy resources [149]. More recently, modified phytochromes produced in bacterial expression systems have been found to have valuable fluorescent properties, of great potential use in medical and biotechnical applications [150, 151]. They are also considered to be promising tools for optogenetic applications [152].

5.1 Fifty years of plant phytochromes

In 1952, scientists at the Agricultural Research Center in Beltsville, Maryland, discovered that lettuce seeds respond to light in the red and far-red region (600 to 800 nm) [153]. Germination was promoted at certain wavelengths, and inhibited at others, and in an extensive set of wavelength- and power-dependent experiments, they mapped out this dependence. The result, known to plant biologists as an *action spectrum* (Fig. 5.1), is a measure of how light at each wavelength affects organisms' behaviour. In this case, the action spectrum shows that seed germination is promoted by red light (with a maximum around 670 nm) and inhibited by far-red light (most strongly at around 725 nm). A remarkable feature of these seeds, is that germination was reversibly affected by the two shades of red. Thus, alternating red and far-red illumination switched the seeds between states in which they germinated more and less frequently.

The pigment responsible for light-dependent germination was directly detected in 1959, using absorption spectroscopy of living maize shoots [154]. The researchers devised a clever detection scheme which allowed measurement despite the strong scattering in the sample and the overlapping illumination-dependent absorption in chlorophyll. It was found that the absorption spectrum of the entire plant was reversibly switched upon illumination with red and far-red light. A difference absorption spectrum was obtained, the main features of which were coincident with those of the action spectrum of lettuce seeds (Fig. 5.2). With this hallmark study, the phytochrome photoreceptor was identified and given its name. The two stable states of phytochrome became known as the Pr (red-absorbing) and Pfr (far-red-absorbing) states. Usually, the Pr state is the most stable form, and is re-formed by dark relaxation, but some phytochromes have developed a more stable Pfr state, so that dark reversion favours the Pfr state. The Pr- and Pfr-preferring phytochromes are somewhat cryptically referred to as "canonical" and "bathy", respectively. The photochromic behaviour of phytochromes (the way in which their absorption spectra change with illumination) is summarised in Fig. 5.3.

Over the decades, a lot of effort has gone into understanding and manipulating plant phytochromes, and a full review is outside the scope of this introduction. But it is worth noting that it took until 1983 before a native

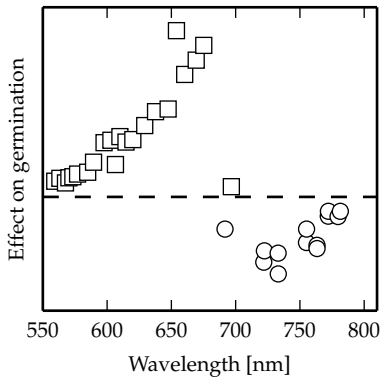


Figure 5.1: Action spectrum of lettuce seed germination, adapted from the original 1952 data [153]. Squares represent promotion, circles represent the weaker inhibition effect, and are multiplied by 10.

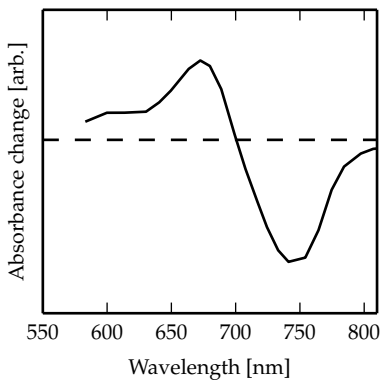
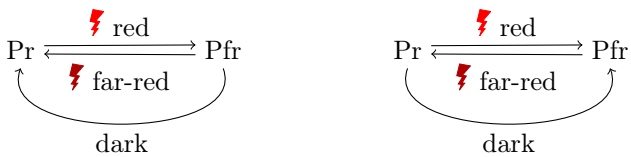


Figure 5.2: Difference spectrum of living maize shoots, displayed as $\Delta A = A(\text{red-illuminated}) - A(\text{far-red-illuminated})$, adapted from the original 1959 data [154].



(a) Canonical **(b)** Bathy

Figure 5.3: Photochromicity and dark reversion in phytochromes.

phytochrome, with photophysical properties similar to those found in living materials, was isolated in vitro, as independently reported by two different groups [155, 156]. The role of plant phytochrome at the organism level was gradually understood [157, 158], while their biochemistry at the molecular level remains unclear to this day [159]. Moreover, because of the difficulty in producing and purifying large amounts of sample, the structural biology of plant phytochromes remained a mystery for many years. To my knowledge, only one plant phytochrome has been crystallised and solved to date, through recombinant expression in bacteria and as late as 2014 [160].

5.2 Bacterial phytochromes: structure and function

A major leap in phytochrome research happened in 1996, when it was discovered that a particular gene in the cyanobacterium* *Fremyella diplosiphon* resembled the genes that encode plant phytochromes [161]. Only the year after, two independent reports showed that a gene from the cyanobacterium *Synechocystis sp. PCC6803* [162] could be over-expressed in *Escherichia coli*, and that the purified protein photoconverts reversibly between red and far-red absorbing states, a behaviour familiar from plant phytochromes [163, 164]. Two years after that, the existence of phytochromes† in the non-photosynthetic bacteria *D. radiodurans* and *Pseudomonas aeruginosa* was discovered [165]. The photoreceptor from the former, awkwardly named *Dr*BphP, is the main subject of study in Papers III, V, and VI, while Paper IV compares *Dr*BphP to some of its bacterial relatives.

The ability to produce large quantities of sample, which is easier to do for bacterial phytochromes, made possible the crystallisation and structural characterisation of these proteins. In part, the surge of structural studies which followed was driven by a desire to understand the elusive plant phytochromes by proxy [159, 166]. Like many other light-sensors, phytochromes (from plants and bacteria alike) have a seemingly modular domain architecture [167–169]. Almost all consist of a photosensory core based on the PAS-GAF-PHY sequence of domains, where the PAS-GAF part binds the chromophore, a linear tetrapyrrole, covalently. For bacterial phytochromes, the chromophore is biliverdin, a metabolite of haem. The output domain, which varies in identity from phytochrome to phytochrome, sits at the C-terminus connected to the PHY subunit. Phytochromes are homodimers both in solution and in crystal structures, where they usually form parallel (head to head) dimer interfaces. This arrangement is illustrated with an example in Fig. 5.4.

*Cyanobacteria, or blue-green algae, are bacteria that perform photosynthesis.

†Phytochrome relatives from bacteria are sometimes called bacterial phytochromes, bacteriophytochromes, or phytochrome-like proteins, to distinguish them from their plant relatives. I use the term phytochrome universally for these proteins, and qualify the term only when needed.

Table 5.1: Selected phytochrome crystal structures. Abbreviations: CBD (chromophore-binding domain), PSM (photosensory module), and FL (full-length protein).

Protein	Construct	PDB	Pr	Pfr	Resolution	Focus	Journal	Year	Ref.
D. radiodurans BPhP	CBD	1ZTU	X		2.5 Å	trefoil knot	Nature	2005	[171]
D. radiodurans BPhP	CBD	2O9C	X		1.5 Å	BV chirality	JBC	2007	[172]
R. palustris BPhP3	CBD	2OOL	X		2.2 Å	photochromicity	PNAS	2007	[173]
P. aeruginosa BPhP	PSM	3C2W		X	2.9 Å	central bundle	PNAS	2008	[174]
Synechocystis 6803 Cph1	PSM	2VEA	X		2.5 Å	arm PRxSF motif	PNAS	2008	[175]
P. aeruginosa BPhP Q188L	PSM	3G6O/IBR	?	X	2.9/3.0Å	BV interactions	PNAS	2009	[176]
P. aeruginosa BPhP	PSM	3NHQ etc.	?	X	2.6 Å	BV rotations	Nature	2011	[177]
D. radiodurans BPhp D207H	CBD	3S7O etc.	X		1.2 Å	fluorescence	JBC	2012	[151]
R. palustris BPhP1	FL	4GW9		X	2.9 Å	signaling complex	Structure	2012	[178]
R. palustris BPhP2	CBD	4E04	X		1.8 Å	dimerization	Acta Cryst. D	2012	[179]
Synechocystis 6803 Cph2	PSM	4BWI	X		2.6 Å	tryptophan switch	JBC	2013	[180]
D. radiodurans BPhP	PSM	4O01/P	X	X	3.2/3.8 Å	arm refolding	Nature	2014	Paper III
A. thaliana PhyB	PSM	4OUR	X		3.4 Å	overview	PNAS	2014	[160]
D. radiodurans BPhP	CBD/PSM	4Q0H/J	X		1.2/2.8 Å	arm interactions	JBC	2014	[170]
R. palustris BPhP2/P3	PSM	4R6L/R70	X		3.4/2.9 Å	arm, central bundle	Structure	2015	[181]

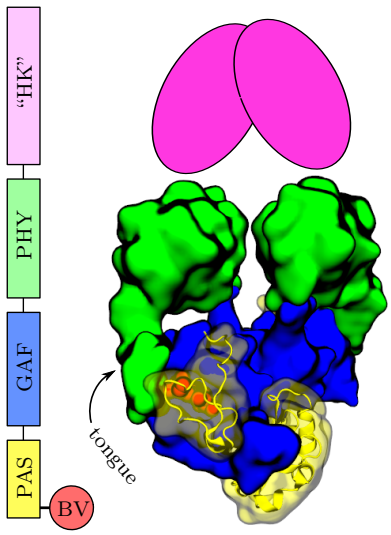


Figure 5.4: Sequence and structure of bacterial phytochromes, as illustrated by a crystal structure of the *Dr*BphP photosensory core (PDB code 4Q0J [170]). The putative histidine kinase output domain (discussed in the text) lacks a crystallographic model, and is shown as blobs. The biliverdin (BV) chromophore is indicated. Reproduced from Paper V.

The overarching goal of phytochrome structural biology has been to establish how the signal, generated close to the N-terminus by the absorption of a photon in the chromophore, is propagated through the molecule towards the output domains. In this quest, a number of crystal structures have been reported, as summarised in Table 5.1. The most straightforward way to characterise structural signalling would of course be to determine crystal models of both the Pr and Pfr states, and work out what has changed. Prior to the study reported in Paper III, this had not been accomplished. Instead, only the ground state (be it Pr or Pfr) of each homologue had been resolved, even if two studies claim to approximate the Pr state of the Pfr-stable phytochrome from *P. aeruginosa* using mutation and trapping at low temperature [176, 177].

In lieu of direct structural data on photoconversion, a number of structural and sequential motifs were identified and postulated to be important for photoconversion, as also listed in Table 5.1. The first phytochrome crystal structure focused on an unusual knot, in which the N-terminus is tied together with a loop in the GAF domain [171]. While peculiar, the meaning of this so-called "light-sensing" knot is unclear. Eventually, two structures were published which contained not just the PAS-GAF, but the whole photosensory PAS-GAF-PHY module for a Pr- and a Pfr-preferring phytochrome [174, 175]. Both noted that a protrusion from the PHY domain, named the "arm" or the "tongue", stretched back onto the chromophore binding pocket, as can be

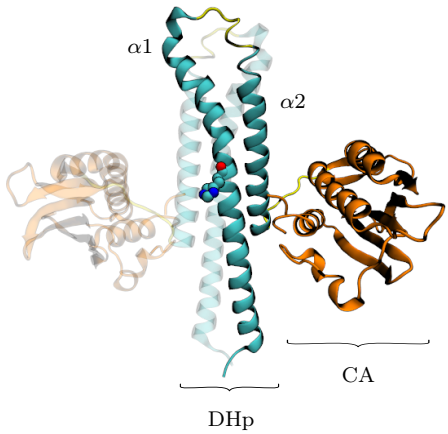
seen in Fig. 5.4. In one, the long helices at the center of the dimer were suggested to carry the signal to the N-terminus [174]. In the other, a well-conserved PRxSF (proline-arginine-X-serine-phenylalanine) sequence motif in the arm was found to interact closely with a well-conserved aspartate residue in the binding pocket, and this interaction was suggested to have a role in signal propagation [175]. Later, the same group suggested that a "tryptophan switch", involving a rotation of the arm about its own axis, was a general device in signalling [180]. Other crystallographic studies have focused on the detailed structure of the chromophore and its environment [172, 176, 177], fluorescence [151], interactions with downstream signalling proteins [178], or dimerisation [179]. Paper III, discussed below, is the first to provide direct structural data on the Pr→Pfr conversion process.

For many plants, phytochrome function has been determined at the organism level, whereas the biochemical details have been rather difficult to understand [182–184]. Conversely, many bacterial phytochromes have unknown functions, but for most, it is known or assumed that they biochemically operate as histidine kinases [185]. Autophosphorylation assays and tests for the transfer of phosphate tags to the response regulator can be performed under red or far-red illumination, to measure activity and to learn about the gating logic. In that way, *Synechocystis* Cph1 [164] and *Agrobacter tumefaciens* Apg1 [186] have been found to signal more actively in the Pr ground state than in Pfr. Other phytochromes, such as *A. tumefaciens* Agp2, have been reported to be more active in the Pfr form [187] (although this is disputed, as discussed in Section 5.5). We now turn to a more detailed discussion of the histidine kinase enzymes, and to how our phytochrome of interest, *Dr*BphP, fits into that context.

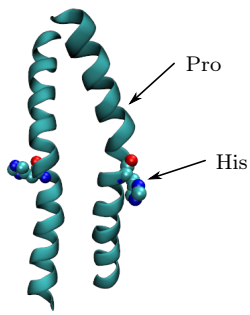
5.3 *Dr*BphP – a Histidine Kinase?

Histidine kinases and two-component systems have received increasing attention during recent years, as more and more structures have become available. At the time of writing, the PDB website's "molecule of the month" section features such a system [188], indicating the broad interest in these proteins. Fig. 5.5a shows the crystal structure of a fragment of the histidine kinase HK853 from *Thermotoga maritima* as an example [189, 190]. Prototypical HK:s are extracellular sensors, connected to intracellular kinase domains via transmembrane segments and signal transducing domains, which eventually connect to the so-called DHp (dimerisation and histidine phosphorylation) domain's α1 helix, marked in the figure. This helix holds the phosphorylatable histidine, as also indicated. A linker of variable length leads to α2, and the homodimeric structure forms a central four-helix bundle. Finally, another linker connects to the CA (catalytic and ATP-binding) domain, which binds ATP in a well-defined pocket and catalyses autophosphorylation.

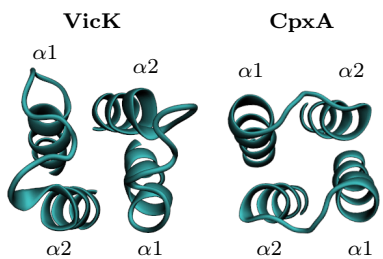
The structural details of histidine kinase autophosphorylation are elusive.



(a) The intracellular domains of the kinase HK853 (PDB code 2C2A) [189]. The phosphorylation site is on $\alpha 1$, shown as spheres.



(b) Asymmetry in a histidine kinase, exemplified by the kink in one of VicK's $\alpha 1$ helices [191] (PDB code 4I5S).



(c) Handedness and linker length in the DHp bundle, exemplified by VicK [191] and CpxA [192] (PDB codes 4I5S and 4BIU).

Figure 5.5: Structural illustrations of histidine kinases.

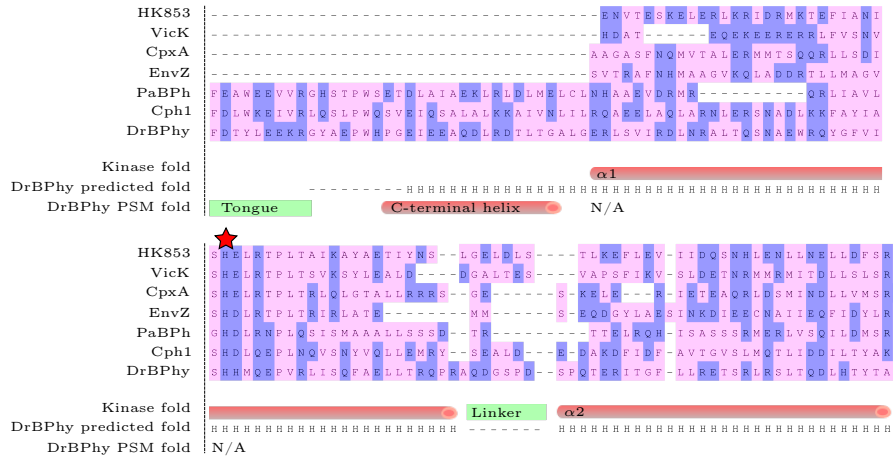


Figure 5.6: Sequence alignment of the histidine kinases HK853, VicK, CpxA, EnvZ, the *P. aeruginosa* bacterial phytochrome PaBphP, the cyanobacterial Cph1, as well as the phytochrome studied here, *Dr*BphP. Also shown is the fold for the kinases, the fold predicted for *Dr*BphP with the program JPRED [196], and the crystallographic secondary structure found for the photosensory module (PSM) of *Dr*BphP [90]. The phosphorylation site is marked with a star. Blue and pink correspond to residues which are frequently charged and neutral, respectively.

This seems to be because the high mobility of the CA domain (which is only loosely linked to DHp) makes these proteins difficult to crystallise [193]. But despite this, a number of themes relevant for the phytochrome family emerge from the developing literature. A recent crystallographic study on the kinase VicK from *Streptococcus mutans* suggests that phosphorylation proceeds via an asymmetric structure, where the $\alpha 1$ helix to be phosphorylated is bent inwards [191], as illustrated in Fig. 5.5b. This bending exposes the phosphate-accepting histidine residue (marked in the figure), and occurs at a highly conserved proline residue found 5 sequence positions towards the C-terminus. Proline is, of course, known to disturb helical structures, and its high conservation (see Fig. 5.6) is circumstantial evidence that this mechanism is important and general. In fact, asymmetry in the DHp domain is a common motif in the recent HK literature, and can also be found in other kinases [192, 194]. A fairly general hypothesis is presented by Bhate *et al.*, in which a symmetric structural signal in the sensory domains is converted to an asymmetric conformational change when it reaches the DHp bundle [195]. Based on a review of recent crystal structures, they argue that this induced asymmetry is a general feature for histidine kinases.

Both the kinases VicK and HK853 discussed above are thought to phosphorylate in the *cis* geometry, such that ATP bound to the CA domain reacts

with the phosphorylatable histidine of the same peptide chain in the dimer. Many HK:s, however, autophosphorylate in a *trans* geometry. For example, the *E. coli* kinases EnvZ [197] and CpxA [192] are known to phosphorylate in a *trans* manner. Interestingly, the *cis*/*trans* autophosphorylation behaviour correlates with sequential and structural traits, in particular with the α1-α2 linker length and with the handedness of the DHp helix bundle. The former is visualised as a sequence alignment in Fig. 5.6, where the region marked "Linker" is clearly shorter for the *trans*-active CpxA and EnvZ homologues, than it is for the *cis*-active HK853 and VicK. The handedness distinction can be understood by building a DHp bundle in a thought experiment. After placing the first α1 helix, the α2 helix on the same peptide chain can be positioned on either side of it, and the two choices are not equivalent. Therefore, the resulting bundle is handed, meaning that it is not identical to its mirror image. Again EnvZ and CpxA represent one class (say, right-handed) while HK853 and VicK fall in the other (then left-handed), as illustrated in Fig. 5.5c. The two possibilities are suggested to have biological meaning, essentially opening up two parallel channels for cellular communication if the handedness is recognised by the response regulator [193]. For phytochromes, the dichotomy represents a choice which must be made when modelling proteins for which no crystal structures exist.

As with the other bacterial phytochromes, sequence homology immediately suggests that the main protein under study in this thesis, the *D. radiodurans* phytochrome *Dr*BphP, is a light-gated histidine kinase [198]. Apart from overall sequence similarity on the C-terminal side of the PHY domain (residues no. 500 and up), the phosphorylatable histidine site and the nearby proline (Fig. 5.5b and 5.6) are conserved too. Many kinases have GAF domains as their signal transducing elements which connect the sensor and DHp domains [195], which in bacterial phytochromes is matched by the very similar fold of the PHY domain. Taken together, sequence similarity, homodimeric assembly, and the sensory–transducing–output domain architecture, make the similarity with kinases striking.

*Dr*BphP also fits into the dichotomy discussed above for histidine kinases, where those with long α1-α2 linkers tend to assemble in left-handed helical bundles and phosphorylate in a *cis* manner, and those with short linkers tend to do the opposite [193]. Fig. 5.6 shows the *Dr*BphP sequence aligned with four histidine kinases, as well as a secondary structure prediction. Both predict that the linker length in *Dr*BphP is very similar to the *cis*-type kinases VicK and HK853, indicating a left-handed DHp bundle.

Although satisfying, viewing *Dr*BphP as a histidine kinase is complicated by that fact that, as far as I am aware, there is no published evidence of kinase activity. The group which discovered both the *Pseudomonas syringae* and *D. radiodurans* phytochromes [165] only show kinase activity for the former [198], with no comment as to why. In fact, there is no clear evidence of any functional or biochemical role for the *D. radiodurans* homologue. In this light, it is interesting that the conserved aspartate or glutamate residue adjacent to

the phosphorylatable histidine is missing in *Dr*BphP, as Fig. 5.6 shows. This residue is thought to activate the histidine toward the reaction with ATP, via an acid-base mechanism, and has been found by mutational studies to be important for autophosphorylation activity [193]. This vaguely suggests that *Dr*BphP might have some other, as yet unknown, biochemical function.

5.4 Photoactivation is structural change

As already mentioned, the overarching goal in phytochrome structural biology is to understand the structural nature of signal propagation. Paper III is a major contribution to this campaign, as it is the first to provide direct information on the Pr and Pfr structures for one and the same phytochrome. It focuses on the photosensory core module (the PAS-GAF-PHY fragment) of *Dr*-BphP, and builds on both crystallographic and solution data. My contribution was in the latter, and I will focus on this aspect here.

The study takes its starting point in time-resolved solution experiments. The first trial was a rapid-readout experiment (Section 3.2) in which two lasers were used to convert a sample back and forth between the Pr and Pfr states. X-rays were collected on a Pilatus 300K detector, with a time resolution of 10 ms. The purpose was, first of all, to see if a highly concentrated solution of *Dr*BphP could withstand repeated laser-induced switching and high-flux X-ray exposure. What we found was not only that the protein is rather stable under these circumstances and that the experiment is feasible, but also that a large $\Delta I(q, \tau)$ signal appeared upon illumination with red light. Fig. 5.7 shows these initial results for the PAS-GAF-PHY fragment.

At the outset of the experiment, the sample is in the Pr form, the most stable state of this canonical (Pr-preferring) phytochrome. As the first red (671 nm) laser pulse hits, a strong difference scattering pattern appears. The curve has a feature at $0.1\,\text{Å}^{-1}$ which completely dominates the signal (note that the data are amplified by a factor q^2), indicating domain movements as discussed in Chapter 4. While the time-resolved data become engulfed in noise at $q > 0.6\,\text{Å}^{-1}$, averaging together all Pr→Pfr difference curves gives entirely consistent data up to $q = 1\,\text{Å}^{-1}$ and beyond. The data directly establish that both the Pr→Pfr and the Pfr→Pr structural conversions are mostly completed in 10 ms, so the actual rise of the signal cannot be resolved in this experiment. Comparing the behaviour of the PAS-GAF, PAS-GAF-PHY and PAS-GAF-PHY-"HK" (full-length) samples in Fig. 5.9, 5.7, and 5.8, respectively, one concludes that the two longer constructs behave alike, indicating that the important structural transformations of photoconversion are described well by the PAS-GAF-PHY fragment. Contrarily, the PAS-GAF fragment gives a response which does not resemble that of the others. Its difference pattern is smaller, and there is no dominating low-q feature. It is also not quantitatively switched back to Pr, probably because its absorption in the far-red region is too weak [90].

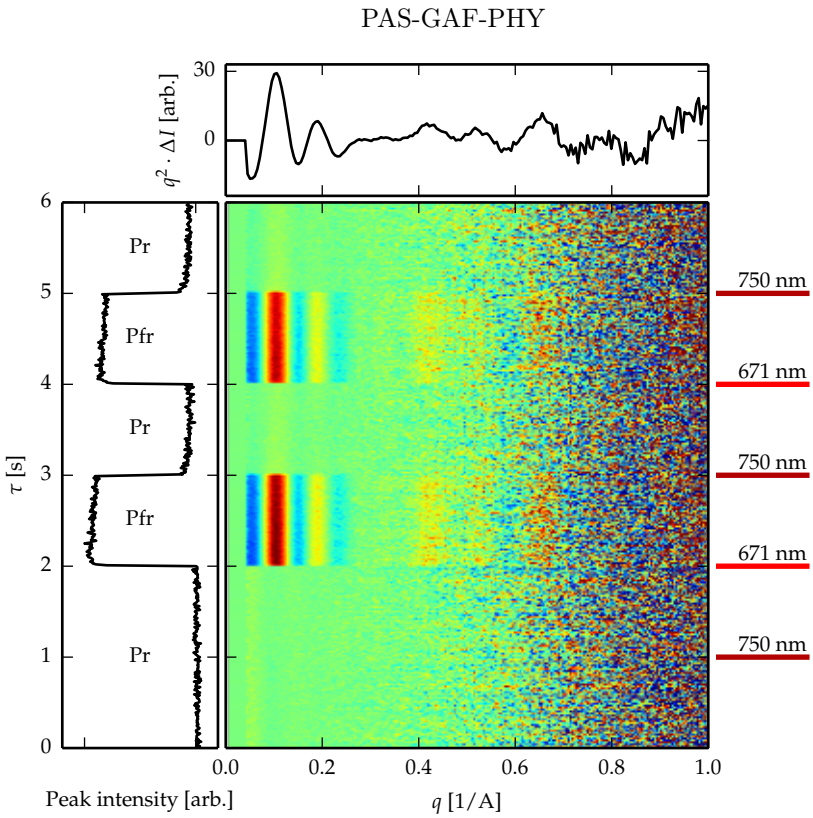


Figure 5.7: A rapid readout experiment of *Dr*BphP PAS-GAF-PHY at 10 ms time resolution. The 2D plot shows $q^2 \cdot \Delta I(q, \tau)$. The kinetic curve on the left shows average intensity of the main peak, the scattering curve at the top shows the average Pr→Pfr difference. Far-red (750 nm) laser pulses were issued at 1, 3, and 5 s, whereas the red (671 nm) laser was fired at 2 and 4 s.

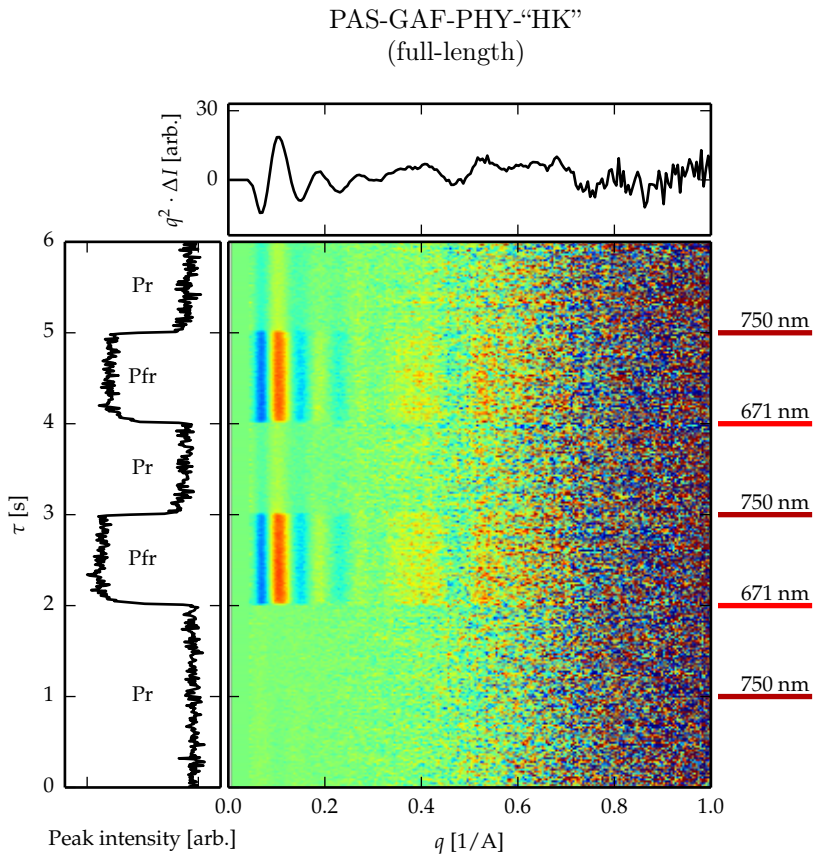


Figure 5.8: Rapid readout data on full-length *Dr*BphP. See Fig. 5.7 for details. The ΔI scale is the same as in Fig. 5.7.

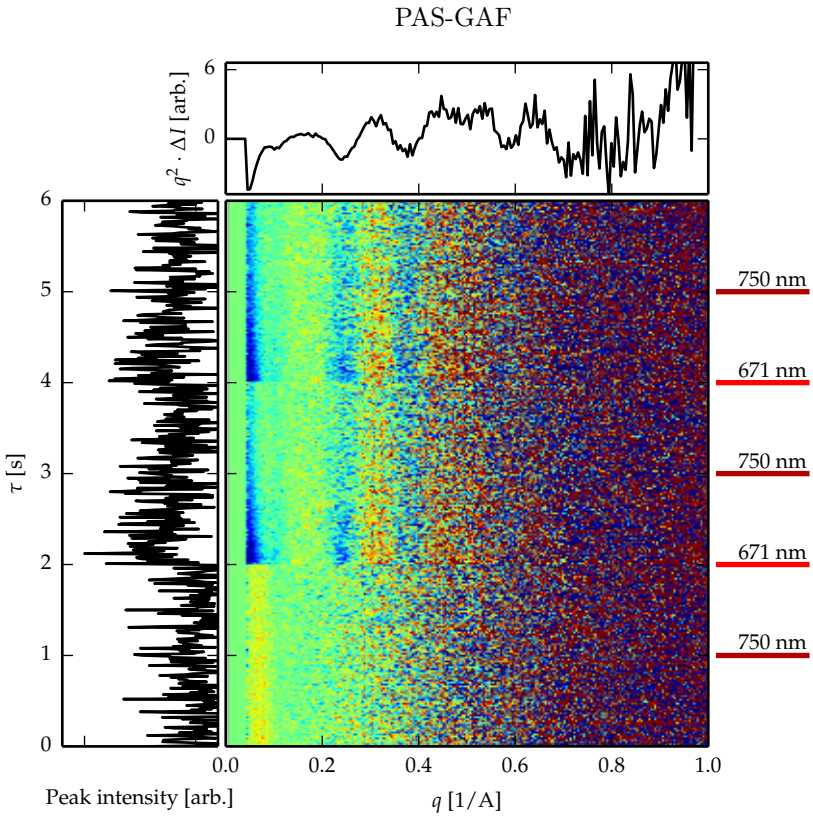


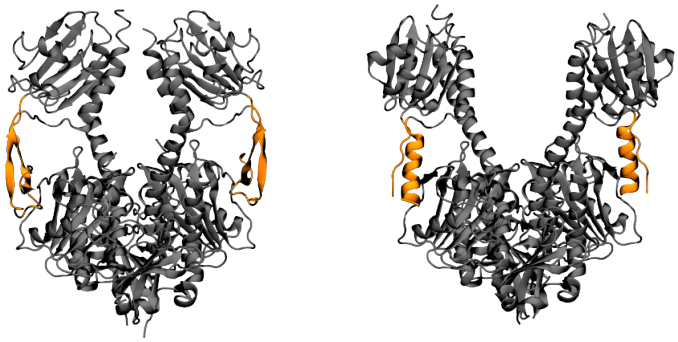
Figure 5.9: Rapid readout data on *Dr*BphP PAS-GAF. See Fig. 5.7 for details. The left-hand panel follows the $q = 0.3\,\text{Å}^{-1}$ peak. The ΔI scale is 5 times smaller than in Fig. 5.7 and 5.8.

Paper III provides a detailed analysis of the structural change in the photosensory (PAS-GAF-PHY) module, to which we now turn. As discussed in Chapter 4, analysis of difference scattering data usually requires prior structural information. Although the structures of other bacterial PAS-GAF-PHY fragments had been published before (see Table 5.1), there were no structures of this fragment from the *D. radiodurans* receptor. Therefore, crystallisation studies were initiated. Our collaboration partners managed to determine the structure of *Dr*BphP crystals grown in the dark, a necessary starting point for any of the protein approaches discussed in the previous chapter. Amazingly, they also managed to crystallise the protein under intermittent red illumination, and the resulting structures exhibit clear differences, as shown in Fig. 5.10. This groundbreaking result (to which I did not contribute) is the first of its kind, as no other crystallographic studies have directly reported on light-induced change in a phytochrome. Two aspects of these structures are especially important. Firstly, illumination causes the opposing PHY domains in the homodimer to move apart, forming a Y-shaped, open arrangement. Interestingly, such an arrangement has been found before by SAXS for a different bacterial phytochrome [199], but was interpreted the other way around, as an opening of the PAS/GAF interface with intact PHY-PHY proximity. This illustrates the uniqueness problem in interpreting SAXS data.

Secondly, the "arm" or "tongue" region discussed above shows different secondary structures in the two crystal models, and is seen as a β-sheet in Pr but as an α-helix in Pfr (orange parts in Fig. 5.10). This provides a clue as to what drives the opening of the PHY-PHY distance. Apparently, changed interactions around the chromophore pocket (for instance, through the motifs previously identified, Table 5.1) causes a shift in the free energy balance between sheet and helix, and the tongue refolds. This affects the effective length of the arm's peptide chain, effectively pulling on the outside of the rigid PHY domains. The central helix bundle, previously thought to actively carry the conformational signal [174], then acts as a mechanical lever, with the end result that the PHY domains are moved apart. Note that this model is not tested through simulation or energetic analysis, but is presented as a mechanistic hypothesis which is in agreement with all the structural observations.

These results are valuable on their own merit, but they also provide a starting point for analysing the solution X-ray data presented above. We did this in the way already described in Fig. 4.1b, with the reaction coordinate given by the internal PHY-PHY distance of the dimer. Reducing the problem to a one-dimensional coordinate obviously comes at the risk of over-simplification, but given the striking difference between the dark and light crystal structures this was warranted, at least as a first-order approach.

Molecular dynamics simulations were started from both crystal models, amended with modelled loops and missing atoms, as described in detail in the supplement to Paper III. In addition to unrestrained simulations, the PHY-PHY distance was artificially sampled by applying a harmonic potential between the two domains, the equilibrium distance of which was slowly



(a) Dark (Pr) structure. **(b)** Light (Pfr-like) structure.

Figure 5.10: Crystallographic models of *Dr*BphP, from crystals formed in the dark or under periodic illumination at 655 nm (PDB codes 4O0P and 4O01, respectively).

scanned. This gave two pools of trial structures, one containing Pr candidates and the other containing Pfr candidates. For each trial structure, the theoretical scattering curve was calculated using SASTBX and Zernicke expansion (see Section 2.2), and all pairwise Pfr-Pr differences were taken. This situation is illustrated schematically in Fig. 5.11 (see also Section 4.4), where 15 randomly chosen frames from each pool of candidates are drawn. As that figure shows, the pairwise difference contained many near-perfect fits, as well as completely orthogonal curves. We chose the 100 best-fitting curves, and considered the participating snapshots (83 Pr and 9 Pfr candidates) as collectively describing the solution structures of the Pr and Pfr forms. Fig. 5.12 shows that all pairwise differences within these sets, in total 747 difference curves, agree very well with experimental data, which means that the sets of solution structures are consistent, and that no ludicrous over-fitting was done when picking structures. The chosen solution structures are shown in Fig. 5.13.

Fig. 5.12 also shows how the model is improved from the starting crystal structures, which do not explain the solution scattering curves very well. It is interesting to compare this situation to that in the first time-resolved protein WAXS study, that of Haemoglobin photolysis. In that case, the subtle structural change was well-described by available crystallographic models. For the phytochrome system, where the structural change is much greater, crystal packing apparently affects the splayed dimeric arrangement, rendering structures which cannot fully explain solution X-ray data. Therefore, the simpler approach of using MD to locally sample around crystal models (Fig. 4.1a), or to "relax" the crystal model, would not be viable in this context as prohibitively long simulations would be required for sufficient sampling.

The difference scattering analysis above generated two sets of structures which were internally consistent, but the analysis is rather improvised. To make sure that frame selection purely based on difference scattering was not

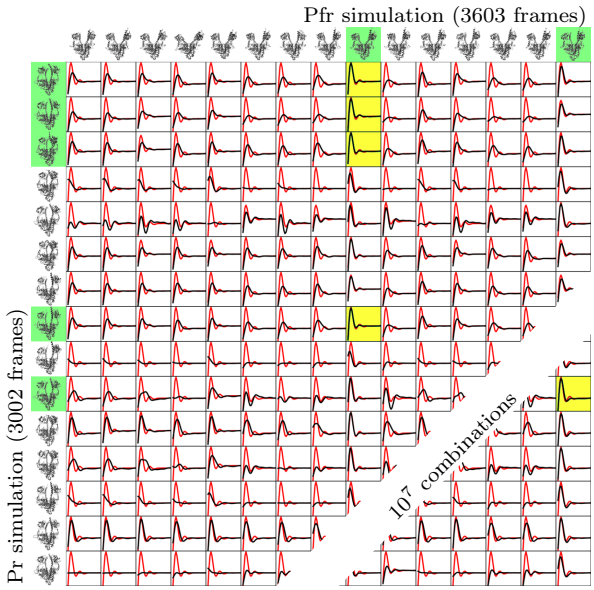


Figure 5.11: Illustration of the difference scattering analysis. Black curves show experimental data, and red curves are theoretical differences. Yellow and green highlights represent best-fitting difference curves and participating structures, respectively.

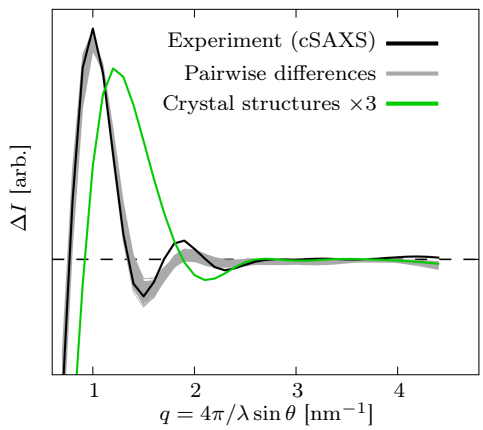
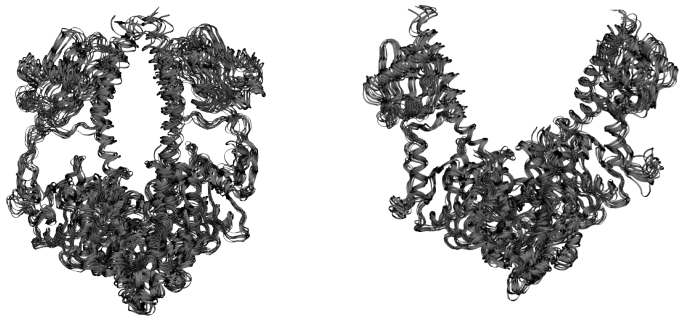


Figure 5.12: All possible pairwise difference curves between the refined Pfr and Pr solution structures, together with experimental data. Also plotted is the theoretical difference curve corresponding to the dark and light crystal structures. Reproduced from paper III.



(a) Pr. **(b)** Pfr.

Figure 5.13: Solution structures, as resulting from the difference scattering analysis.

imprudent, we verified the result using conventional SAXS. In an experiment much less sensitive to the details of structural conversion, we recorded the equilibrium scattering from samples of *Dr*BphP previously illuminated with LED:s. After correcting for the relative Pr and Pfr populations, we then compared the theoretical scattering curves from each of the original candidate structures to data. The result is shown in Fig. 5.14, where a strong correlation between PHY-PHY distance and agreement with absolute SAXS data can be seen. The structures which were picked out in the difference scattering analysis are shown as squares, and they are (i) among the best matches with absolute SAXS data, and (ii) clustered together along both axes. This corroborates the above structural analysis.

Paper III proposes representative sets of solution structures based on time-resolved X-ray experiments, and provides another example of where such data are useful, an example added to the others in Table 5.1. The refined solution structures differ from their crystal models, in that the PHY-PHY distance is larger in solution. This is not surprising, considering the crystal contacts (described in the supplement of Paper III). But what is the value of these solution structures, given that the crystal models already lead to a hypothesis on molecular signal propagation? They are not very valuable in and of themselves. In fact, as discussed in the supplement to Paper III, the exact degree of opening is a rather ill-posed problem, and in reality the distribution of PHY-PHY distances probably has significant width. Rather, the solution data serve to prove that the crystal models represent relevant structures, and that the differences between them has biological meaning. This type of validation is rare in structural biology, and should probably be sought more often to make sure that observed crystal structures actually represent solution conformations [200].

It is interesting to speculate around how the Pr→Pfr structural conversion might be received by the output domain. If *Dr*BphP indeed is a histidine kinase then, as we have seen, the signal might be converted into a kinking of

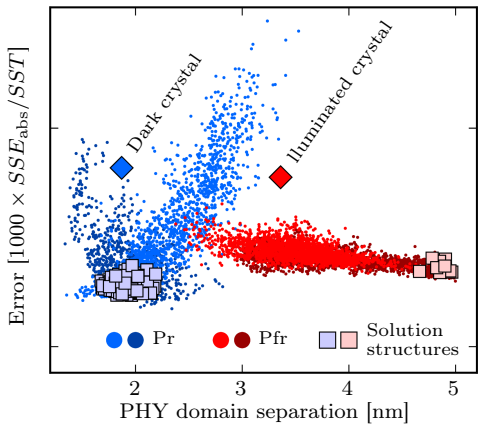


Figure 5.14: Verification of the difference scattering analysis by comparison with conventional SAXS data. See text for explanation. Reproduced from Paper III.

one or more helices in the DHp bundle. It is fully conceivable that the motion of the PHY domains is transferred into an induced asymmetry downstream, with the DHp bundle essentially intact but with its symmetry broken. A different interpretation is made in the commentary which accompanied Paper III [200]. There, it was speculated that the splaying of the PAS-GAF-PHY dimer might fully occur even in the presence of a C-terminal DHp bundle. Then, the dimeric interface of the kinase domain itself would be broken, which would be a rather violent structural signal compared to those discussed in the literature on histidine kinases [193]. The authors argued that such a transformation would hinder *trans*-autophosphorylation, thus gating activity. This type of regulation appears less likely in the light that *Dr*BphP resembles a *cis*-phosphorylating kinase, but not impossible since this assignment is far from definitive. In Paper VI (a "related" article not formally included in this thesis), we also argue that a mechanism where the DHp dimer interface is broken should be considered. However, the data presented here do not clearly support it. A detailed structural analysis of the full-length data in Fig. 5.8 would be needed, which in turn would need a believable homology-based model for the full-length starting structure. This remains to be done and would constitute a major step in the full understanding of phytochrome structural signalling.

5.5 *All phytochromes are equal*: a comparative study

Paper IV also deals with photoconversion in bacterial phytochromes, but takes a different approach. Rather than interpreting solution scattering data struc-

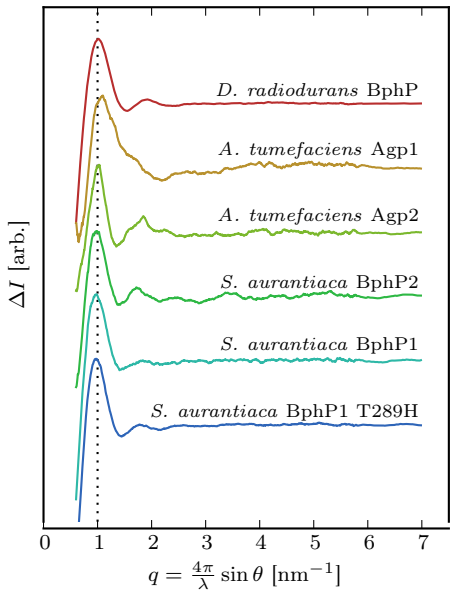


Figure 5.15: Pr→Pfr difference scattering patterns for phytochromes from the bacteria *D. radiodurans*, *A. tumefaciens*, and *S. aurantiaca*, all shifted for clarity. Reproduced from Paper IV.

turally, it establishes that several phytochromes show almost identical response in the time-resolved experiment. The study therefore benefits from the great detection power of the difference scattering method, without suffering from the interpretation problems discussed at length in the previous chapter.

Fig. 5.15 shows the scattering fingerprints of Pr→Pfr photoconversion from the photosensory cores (PAS-GAF-PHY fragments) of five wild-type bacterial phytochromes. Aside from the familiar *D. radiodurans* phytochrome signal, those from the recently discovered phytochromes from *Stigmatella aurantiaca* (*Sa*BphP1 and *Sa*BphP2, including a point mutant of the former), and from *A. tumefaciens* (Agp1 and Agp2) are also shown. The experimental details are described in the supplement to Paper IV, but the experiment is essentially a rapid-readout exercise done at 40 ms time resolution. These five phytochromes give rise to very similar difference scattering patterns. We draw the conclusion that this reflects similar molecular movements upon illumination. Note that this is not a rigorous conclusion; there are, of course, a large number of ways in which to reproduce a given scattering pattern. But given the similarity of all known phytochrome crystal structures (Table 5.1), it appears reasonable that similar signals reflect similar motions.

Kinetic traces of these experiments are shown in Fig. 5.16. Agp2 is not included, as it was illuminated with continuous LED:s for practical reasons. All samples can be seen to switch reversibly, much as in the original *Dr*BphP

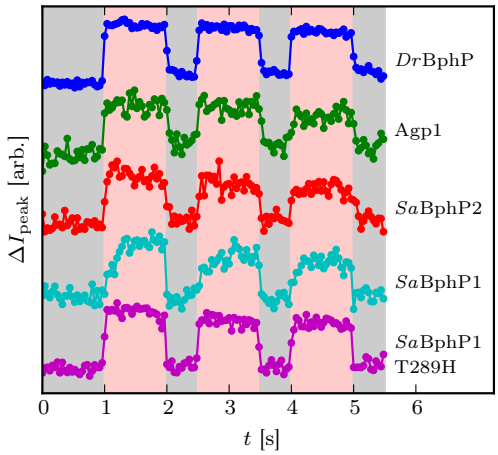


Figure 5.16: Shifted kinetic traces corresponding to Fig. 5.15. Far-red (785 nm) laser pulses were administered at 0.5, 2, 3.5, and 5 ms. Red laser (671 nm) pulses arrived at 1, 2.5, and 4 ms. Reproduced from Paper IV.

experiment (Fig. 5.7). Taken together, these results suggest that the mechanism of conformational signalling has been conserved throughout the evolution of these protein homologs. In light of the differences otherwise found within the phytochrome family, this result is rather significant, and has especially far-reaching consequences for the cases of Agp2 and *Sa*BphP1. I now turn to a discussion of each.

5.5.1 Bathy phytochromes: Agp2

The odd one out among the phytochromes in Fig. 5.15 is Agp2, a bathy phytochrome which relaxes to Pfr in the dark. This homologue was discovered in the genome of *A. tumefaciens* around 15 years ago [201, 202], and was expressed and investigated shortly after [186, 187]. The meaning of the apparent inversion of the Agp2 photocycle has never been fully understood. It is unclear whether Agp1 and Agp2 have opposing physiological function and, if so, how this is accomplished structurally.

The data reported in Paper IV and in Fig. 5.15 establish that the directionality of Agp2 photoconversion is the same as in other phytochromes. That is, illuminating Agp2 with red light causes the same type of structural change as it does for its sibling Agp1, and for its cousins *Dr*BphP, *Sa*BphP1, and *Sa*BphP2. This result is extremely important for understanding the evolution and signalling mechanisms of these apparently opposing phytochromes, but of course says nothing about the biochemical or physiological function. Karniol *et al.* [187] have reported that Agp2 displays inverted autophosphorylation activity, that is, phosphorylates more actively in the Pfr form. To confuse things, Zienicke shows the opposite, that the phosphorylation logic of Agp2

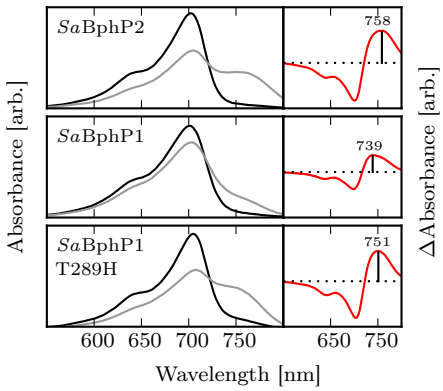


Figure 5.17: Absorption spectra of *S. aurantiaca* phytochromes 1 and 2, obtained at steady state under continuous illumination from red and far-red LED:s (gray and black lines, respectively). Reproduced from Paper IV.

is the same as for all other described systems, with higher activity in the Pr form [203]. Our data does not settle this dispute, but we can at least say, that whatever the logic of the kinase is, it is gated by a structural signal with prototypical directionality. Interestingly, Karniol *et al.* have identified the output domain of Agp2 as the founding member of a new class of histidine kinases called the HWE family. Supposing that Agp2 is biochemically inverted, then its inversion is a result of a substituted kinase domain, not of a change in the structural signalling. This would be a manifestation of modular domain architecture. In fact, phytochromes and other light receptors are known to be modular, and chimeric kinases can be constructed by cutting and pasting together naturally occurring sensory and output domains [168, 204, 205].

5.5.2 Photochromicity and structure: *Sa*BphP1

The above results are rather surprising in another case, that of *S. aurantiaca* BphP1. To see why, consider the absorption spectra in Fig. 5.17. The top panel shows the photochromic response of *Sa*BphP2. Just like most other wild-type bacterial phytochromes, red illumination causes a significant change in the absorption spectrum, which adopts a spectrum typical of the Pfr state. The middle panel shows the corresponding data of *Sa*BphP1. Unlike its sibling, it does not fully adopt the Pfr spectrum, but only develops a shoulder on the red side of the main peak. It is not just a matter of populations, as the difference spectra on the right-hand side show. Sequence alignment (supplement to Paper IV) reveals that a conserved histidine, which for other phytochromes lies close to the chromophore, is replaced by a threonine in *Sa*BphP1. Therefore, our collaborators cloned and produced a point mutant (T289H), where the "missing" histidine is reintroduced. As the bottom panel of Fig. 5.17 shows, this restores photochromicity.

Now, as already mentioned, *Sa*BphP1 structurally behaves like its relatives, and the point mutation T289H does not change this (Fig. 5.15). This unexpected result has both phytochrome-specific and general, methodological implications. For phytochromes, it implies that the conserved histidine (which corresponds to His290 in *Dr*BphP) is not indispensable for structural signalling. Systematic mutagenesis studies have shown that many conserved residues affect the photocycle of phytochromes [160, 206, 207]. With photochromicity and Resonance Raman spectroscopy as the only probes for photoconversion, Wagner *et al.* assigned His290 of *Dr*BphP a "central role" in the light-induced isomerisation of the chromophore. Although these authors acknowledge the need for more data to pin down the residue's exact task, this conclusion appears to be hasty. At least for *Sa*BphP1, structural photoconversion does not depend on this residue.

So what is the role of this conserved histidine? It may be that its effect is exactly what is shown in Fig. 5.17, an alteration of photochromicity itself. Under ambient light conditions, a lower absorbance in the far-red region will affect the Pr and Pfr populations, decreasing the Pr/Pfr ratio if all other factors are kept constant. It may also be to adjust the equilibrium properties (such as the relative stabilities) of the Pr and Pfr forms, or to tune the light or dark conversion dynamics between the two, all of which affect the light-sensing properties of the receptor. In fact, Fig. 5.16 shows that the Pr→Pfr (but not the Pfr→Pr) conversion kinetics of *Sa*BphP1 are in fact much slower than for the other phytochromes, as the rise is seen even at 40 ms time resolution. Just like with photochromicity, the "restoring" mutation T289H recovers prototypical switching behaviour. This is likely a useful clue for those interested in understanding the importance of this particular residue.

From a methodological standpoint, these results demonstrate that quarternary-structural change cannot be probed with spectroscopic techniques which interrogate only the chromophore and its local environment. Photochromicity alone is shown to be a poor indicator of global signal propagation. In lieu of crystallographic information, phytochrome researchers have largely relied on local spectroscopies for understanding signalling. Paper IV shows that this is insufficient, and establishes difference X-ray scattering as a much-needed complementary technique in the field.

5.6 Phytochrome structural dynamics[‡]

While Papers III and IV contain some kinetic considerations, a full time-resolved study based on both fast (pump-probe) and slower (rapid readout) experiments for the PAS-GAF, PAS-GAF-PHY, and full-length *Dr*BphP constructs was deferred to Paper V. Over the decades of plant phytochrome research, much attention has been paid to the kinetics of transformation of the

[‡]This section describes unpublished work, and the primary data are therefore not reproduced. Instead, the text refers to figures in the attached manuscript, Paper V.

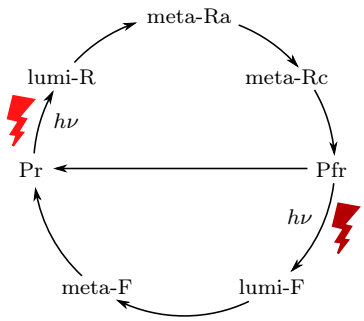


Figure 5.18: Typical representation of a (canonical) phytochrome photocycle.

absorption spectrum in the Pr→Pfr process [208–212]. It soon became clear that the first reaction step is light-dependent, while a series of "dark" reactions lead, through a number of intermediates, to the ultimate formation of the Pfr state [209].

The canonical photocycle is usually represented as in Fig. 5.18. With nomenclature borrowed from the literature on Rhodopsin, the primary photoproduct is labelled lumi-R or lumi-F, when formed from illumination of Pr of Pfr, respectively. The intermediates subsequently formed are respectively termed meta-R or meta-F, with letters a, b, c etc. indicating subsequent intermediates. Thus, according to Kendrick et al. [209], phytochrome from oat, upon illumination as Pr, passes through the intermediates lumi-R, meta-Ra and meta-Rb before forming Pfr.

It is rather difficult to piece together a coherent description of the conversion kinetics from published data, and the overall mechanism has been described as "a series of ill-defined intermediates that appear on the picosecond to millisecond timescales." [159] Indeed, complexity and non-exponential behaviour has been proposed to be intrinsic to this protein family [211]. Time-resolved absorption spectroscopy data are often analysed in terms of a kinetic model, or decomposed using numerical techniques such as Singular Value Decomposition (SVD) or Principal Component Analysis (PCA). The outcome is a series of time constants, which loosely describe the set of processes which constitute the overall reaction. Fig. 5.19 provides a summary of such time constants for various phytochromes.

The spread in the time constants of Fig. 5.19 precludes direct comparison. The numerical analyses used are sensitive to rather subjective choices, especially with regard to the number of components extracted. Looking at the original literature, it is also clear that the component spectra found from said analysis are similarly hard to compare, since they are both featureless and discordant. It is therefore not surprising that a varying number of processes have been found both for the *Avena sativa* (oat) phytochrome A, and

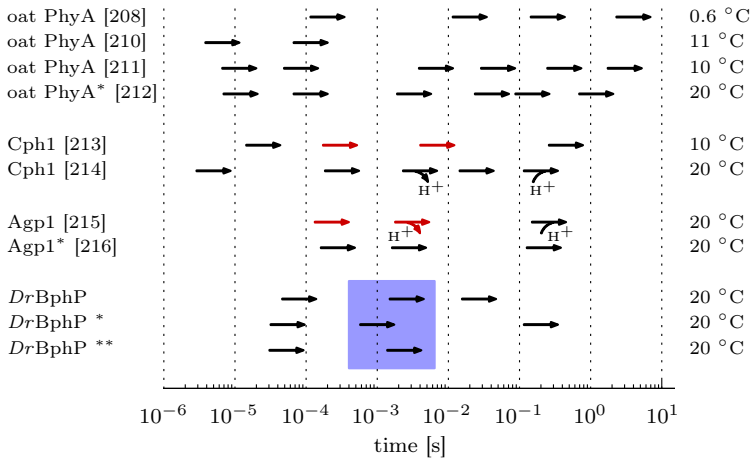


Figure 5.19: Selected literature time constants of phytochrome photoconversion, measured by absorption spectroscopy in the range from microseconds to seconds. Proton release and uptake are indicated, and red arrows denote a kinetic H/D isotope effect larger than 3/2. Samples are full-length unless indicated (* = PAS-GAF-PHY, ** = PAS-GAF). The *Dr*BphP results are those reported in Paper V, as discussed below, and the blue box represents the major structural transformation. Reproduced from Paper V.

for *Synechocystis* Cph1. Still, it appears that compared to the others, the plant phytochrome undergoes more reaction steps, and that these processes span a wider time range. Apparently, a trend in complexity of the phototransformation can be seen, where phytochromes from non-photosynthetic bacteria (Agp1) are the simplest, cyanobacterial Cph1 are intermediate, and the plant phytochrome shows the most complex kinetics. Considering the small number of studies included, this could also be a mere coincidence.

There is almost no structural knowledge about the reactions in Fig. 5.19. Clever pH-measurements and isotope studies have revealed that proton transfers from protein to solution, and back, occur in both Agp1 and Cph1. Both have also been shown to involve rate-determining proton transfers steps, as indicated by red arrows in the figure. But aside from that, the only specific information available pertains to the very early steps of the reaction, where the tetrapyrrole chromophore is isomerised, usually in a matter of picoseconds [212, 217–224].

In Paper V, we present conventional time-resolved absorption-spectroscopic data for three constructs of *Dr*BphP. The novelty of our study is that we also present time-resolved solution X-ray experiments, by which the spectroscopic evolution can be directly related to structural change. Figure 3 of Paper V shows the data. As discussed in Section 5.4, it is immediately clear that the PAS-GAF fragment behaves very differently compared to the PAS-GAF-PHY

and full-length variants. All three show an induced absorption at around 730 nm after 10 µs, which decays over the subsequent 100 µs. For PAS-GAF, this decay leads to a mostly bleached product, while in the larger constructs a different process starts at around 1 ms and leads to the familiar Pfr spectrum (cf Figures 5.2 and 5.17).

Unlike the spectroscopic dynamics, the X-ray difference signals of all three constructs remain essentially featureless over the whole microsecond time scale. This is rather a surprising result. On the one hand, structural dynamics are clearly going on, as the evolution of the absorption spectra shows. On the other hand, these experiments were carried out under similar conditions as the studies referenced in Chapter 4, where tiny (but global) rearrangements were straightforwardly detected. Our conclusion is that until the onset of the millisecond process, structural change is localised to the chromophore environment. In other words, the intermediates of photoconversion are reactant-like.

We decomposed the spectroscopic data in order to very roughly estimate the time constants present in the complex reaction. This was done by fitting a sum of exponentials, as detailed in Paper V, and gave the results shown in Table 5.2. The results are also included in Fig. 5.19 above, and reveal kinetics somewhat similar to the bacterial Agp1 and cyanobacterial Cph1. This establishes comparative behaviour, and allows careful extrapolation of the other results found for *Dr*BphP.

Figure 5 of Paper V directly compares the spectroscopic and X-ray transients. The figure shows the absorbance at 754 nm as a representative of the spectroscopic process. The structural process is represented by the in-growth of the steady-state signal. This is extracted by decomposing the overall signal into a set of N basis curves, each with an associated time evolution. Thus, $\Delta I(q, \tau)$ is expressed as a matrix D, which is factored into a matrix S of basis curves, and a matrix T of time evolutions,

$$\underset{(N_q \times N_\tau)}{\mathrm{D}} = \underset{(N_q \times N)}{\mathrm{S}} \cdot \underset{(N \times N_\tau)}{\mathrm{T}} \quad . \tag{5.1}$$

If either S or T is known, the other can be extracted from the data by solving (5.1) in the least-squares sense. This is almost trivial in many high-level programming languages, and is implemented in the `libwaxs` code described in Section 3.4. For this particular case, two components were used ($N = 2$). The first was a constant offset, found to be almost insignificant, and the other was the steady-state signal, as approximated by the difference scattering over $500\,\mathrm{ms} < \tau < 1\,\mathrm{s}$. The half-times of the main structural process in the three protein constructs are summarised in Table 5.2. If each repeat (each individual detector image) is considered as its own observation, with no averaging

Table 5.2: Spectroscopic time constants (τ_i) and half-times of the main X-ray signal.

Construct	τ_1	τ_2	τ_3	X-ray
PAS-GAF	53 µs	3.5 ms	–	1 ms
PAS-GAF-PHY	55 µs	1.0 ms	200 ms	2 ms
Full-length	80 µs	2.6 ms	27 ms	6 ms

done, then the spread in T for each delay time τ can be used to estimate the uncertainty. Thus, Figure 5 of Paper V also includes 95% confidence intervals.

The two main results of Paper V follow from this analysis. The first is that the main structural change is coincident with the second-to-last spectroscopic transition. This is the step which, for Agp1 and Cph1, involves the transfer of a proton from the protein to the surrounding solution. In the terminology of Fig. 5.18, it is the formation of the meta-Rc intermediate. Extrapolation to plant phytochromes is much harder, as there is little similarity in the sets of time constants. Further time-resolved X-ray studies would perhaps show that one of the many transitions described for oat PhyA is similarly associated with global structural change.

The second result is that the kinetics (Figure 5 of Paper V) are similar for both the PAS-GAF, the PAS-GAF-PHY, and the full-length samples. All three exhibit a main signal which rises in a few milliseconds. This suggests, but in no way proves, that the rate-limiting step is a process which is contained within the PAS-GAF moiety. Apparently, the addition of downstream PHY domains slows the process down only slightly, from 1 ms to 2 ms. The further addition of the output domain retards it slightly more, to a half-time of 6 ms, which seems to represent a frictional retardation rather than an entirely new rate-limiting step. The process which limits the rate of structural conversion within the PAS-GAF domain could, for example, be the transfer of a proton from the chromophore or from chromophore-adjacent residues. This would be consistent with the fact that no structural intermediates can be detected on the microsecond time scale.

5.7 Summary

The following is a bullet-point summary of the phytochrome-specific outcomes of Papers III, IV, and V.

- Photoconversion in solution causes only minor structural change in PAS-GAF, but concerted domain movements in PAS-GAF-PHY and in the full-length protein.
- For PAS-GAF-PHY, the structural signal is expressed as a dimeric open-

ing, wherein the opposing PHY domains are separated. For the other constructs, there is no structural interpretation.

- Several other bacterial phytochromes show similar X-ray signals, indicating that structural signalling is conserved.
- The bathy (Pfr-preferring) phytochrome Agp2 does not have an inverse structural photocycle, it is merely the direction of dark reversion which is reversed.
- The much-investigated His290 (in *D. radiodurans* numbering) is not critical to signalling as previously suggested, but it does affect the kinetics of the Pr→Pfr transition.
- In bacterial phytochromes, the early photoconversion steps do not cause much structural change. The so-called meta-Rc state is the first intermediate which is structurally similar to the Pfr product.
- It seems that local changes in the chromophore-binding PAS-GAF module control the rate of photoconversion.

Chapter 6

Conclusion

This thesis summarises four years of work, which started out with the ambition to generally "do something useful with time-resolved X-ray scattering". The project quickly branched out in two directions. The first was an investigation of how to successfully carry out the experiments themselves, and how to reliably interpret their results. These questions had been very actively researched at our Department for a number of years, and our efforts were directed at improving existing solutions and finding novel applications. As Chapters 2-4 show, this work ranged from calculating X-ray scattering intensities in new ways, to constructing new and improved sample environments and data acquisition schemes, all the way to designing and implementing new software for data interpretation. It was at times extremely stimulating work, the results of which are described in the fine print and supplements of all five included papers. They are in the focal points of Papers I and II.

The second branch was formed when our Finnish collaborators brought the fascinating phytochrome system to our attention. This light-sensing protein was heavily researched yet relatively poorly understood. It turned out that phytochromes were an excellent target for novel scattering measurements, as they are quite stable and their structural signalling is easily controlled with light. Chapter 5 describes a number of papers (III, IV, and V) where different aspects of these light-sensing proteins were studied. These papers are significant contributions to the understanding of phytochrome signalling. They also demonstrate how bringing a new experimental method into an old field of research can quickly advance understanding, highlighting the potential value of basic methodological research and proof-of-principle projects.

6.1 Results summary

The aggregate results of these investigations show that structural change can be studied with time-resolved X-ray scattering at two levels. Firstly, if the initial structure is roughly known and if the conformational transitions studied

are of a suitable type, then the method can be used to deduce likely movements, as in Paper III. The conditions for when this can be done were roughly drawn up in Chapter 4, but the precise limits for structural interpretation will need more work and experience to become clear. Secondly, the detection power and time-resolution of the difference scattering technique are sometimes useful in themselves. For example, and this is already demonstrated in the literature [5, 6], the sequential interconversion between structural intermediates can be followed by numerical decomposition of the data, even if each intermediate scattering curve cannot be interpreted. In fact, this is how time-resolved absorption spectroscopy is often used in practice. Similarly, comparative behaviour between fragments of a protein or between protein family members can be established by comparing the raw data. Papers IV and V ran along these lines.

In summary, then, Paper I shows that structural models of varying resolution can be used for analysing difference scattering data, and that the level of resolution should be chosen rationally for each case. It also shows that difference scattering is easier to calculate than are absolute SAXS data, because of error cancellation to do with the solvation layer. These findings are exploited in Paper II which describes an MD-based interpretation method and discusses its limits. We suggest that the method will serve as a starting point for further developments, and note that two competing projects were published just after ours, which is at least an indication that the work is of interest to the structural biology community.

Paper III proposes a solution-structural change for the *D. radiodurans* phytochrome. The results are based on crystallography, time-resolved X-ray scattering, and conventional SAXS. The crystallographic model leads to a detailed molecular hypothesis of signal propagation, and the main role of the solution data are to lend credibility to this model, showing that domain movements do occur in solution.

Paper IV extends this result by showing that similar domain movements occur in several other bacterial phytochromes. This suggests that a core actuator mechanism (a way of translating light into movement) has been conserved in evolution. It becomes even more interesting when the peculiarities of these phytochromes are considered. In one, the spectroscopic photocycle is apparently reversed, whereas the structural one is not. In another, the photochromicity so typical of phytochromes is barely seen, yet the full structural movement is produced upon illumination. These details show how the core structural apparatus has been adapted and regulated to fill various cellular needs.

Lastly, Paper V extends the result of Paper III to the full-length protein, and dissects the time dependence of molecular signal propagation. Surprisingly, the rich spectroscopic evolution of intermediates seen over the submillisecond range is not reflected in the time-resolved X-ray scattering signal. Considering the sensitivity of the time-resolved technique, this negative result indicates that the early intermediates are structurally very similar to the Pr

starting state. Contrarily, the last intermediate seems structurally very similar to the Pfr state. We conclude that the spectroscopic evolution seen at early times reflect local rearrangements around the chromophore.

6.2 Outlook

The phytochrome results reported here could be furthered by more experiments. In structural biology, slicing up the protein of interest into different fragments or investigating point mutants can sometimes give valuable insights. Our studies already exploit several fragments of the *D. radiodurans* phytochrome, but systematic investigations of point mutants would most likely reveal which residues (or combinations thereof) are crucial to global structural signalling.

To me, the more interesting question is what the future holds for time-resolved X-ray scattering as a method. The technique has been up-and-coming for some time now, but has yet to take off. Its applicability hinges on triggering the dynamic processes of interest, as light-activation is a very rare property for proteins. In that regard, most time-resolved X-ray studies are in a sense "drunkard's searches", attempts to find a wallet left in the park at night by looking for it under a street lamp a few blocks away, where the light is better. This is quite all right, there are plenty of interesting questions which pertain to light-sensitive proteins, and the technique no doubt has a role to fill. In the street lamp metaphor, one could say that the ground is littered with wallets, and the ones under the street lamp are no worse than those in the park. But in the long run, the method will need to be applied to other systems. Chapter 3 outlined ways to extend the experimental utility, and the development of microfluidic devices which allow rapid mixing without requiring excessive sample volumes seems to be a promising way forward. Another interesting route could be to directly and selectively excite low-frequency vibrational modes using low-energy radiation [225], and to study these movements in solution with an X-ray probe.

Disregarding the practical problem of reaction triggers, the ultimate bottleneck in time-resolved X-ray scattering is in data interpretation. The results presented here do not suggest an obvious way forward. Personally, I think that more attention must be paid to the equilibrium fluctuations of macromolecules. It should be systematically investigated whether inherent dynamics encode functional movements, as seems to be the case for some systems [226, 227]. If so, then these meaningful, collective fluctuations could define a lower-dimensional space in which to run, for example, X-ray guided MD simulations. Equilibrium fluctuations can be found and analysed using MD simulations of systems for which good starting structures exist. In this way, exploiting resting-state *dynamics* as well as resting-state *structure* would add to the complementary information so desperately needed for data interpretation.

Bibliography

[1] Lange, O.F., Lakomek, N.A., Farès, C. *et al.* "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." *Science* **320**, 1471–1475 (2008).

[2] Hensen, U., Meyer, T., Haas, J. *et al.* "Exploring Protein Dynamics Space: The Dynasome as the Missing Link between Protein Structure and Function." *PLoS ONE* **7**, e33931 (2012).

[3] Ramanathan, A., Savol, A., Burger, V. *et al.* "Protein conformational populations and functionally relevant substates." *Accounts of Chemical Research* **47**, 149–156 (2014).

[4] Westenhoff, S. & Nazarenko, E. "Time-resolved structural studies of protein reaction dynamics: a smorgasbord of X-ray approaches." *Acta Crystallographica Section A* **66**, 207–219 (2010).

[5] Cammarata, M., Levantino, M., Schotte, F. *et al.* "Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering." *Nature Methods* **5**, 881–886 (2008).

[6] Andersson, M., Malmerberg, E., Westenhoff, S. *et al.* "Structural dynamics of light-driven proton pumps." *Structure* **17**, 1265–1275 (2009).

[7] Minitti, M.P., Budarz, J.M., Kirrander, A. *et al.* "Imaging Molecular Motion: Femtosecond X-Ray Scattering of an Electrocyclic Chemical Reaction." *Physical Review Letters* **114**, 255501 (2015).

[8] Editorial. "Should scientists tell stories?" *Nature Methods* **10**, 1037–1037 (2013).

[9] Guss, J.M., Merritt, E.A., Phizackerley, R.P. *et al.* "Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers." *Science* **241**, 806–811 (1988).

[10] Pollack, L. "SAXS studies of ion-nucleic acid interactions." *Annual Review of Biophysics* **40**, 225–242 (2011).

[11] Lages, S., Goerigk, G. & Huber, K. "SAXS and ASAXS on Dilute Sodium Polyacrylate Chains Decorated with Lead Ions." *Macromolecules* **46**, 3570–3580 (2013).

[12] Warren, B.E. *X-Ray diffraction.* Dover Publications, New York (1990).

[13] Doyle, P. & Turner, P. "Relativistic Hartree-Fock X-ray and electron scattering factors." *Acta Crystallographica A* **24**, 390–397 (1968).

[14] Cromer, D. & Mann, J. "X-ray scattering factors computed from numerical Hartree-Fock wave functions." *Acta Crystallographica A* **24**, 321–324 (1968).

[15] Sorenson, J.M., Hura, G., Glaeser, R.M. *et al.* "What can x-ray scattering tell us about the radial distribution functions of water?" *The Journal of Chemical Physics* **113**, 9149 (2000).

[16] Merzel, F. & Smith, J.C. "Is the first hydration shell of lysozyme of higher density than bulk water?" *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5378–5383 (2002).

[17] Svergun, D., Barberato, C. & Koch, M. "CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates." *Journal of Applied Crystallography* **28**, 768–773 (1995).

[18] Fraser, R.D.B., MacRae, T.P. & Suzuki, E. "An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules." *Journal of Applied Crystallography* **11**, 693–694 (1978).

[19] Grishaev, A., Guo, L., Irving, T. *et al.* "Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling." *Journal of the American Chemical Society* **132**, 15484–15486 (2010).

[20] Chen, P.c. & Hub, J.S. "Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data." *Biophysical Journal* **107**, 435–447 (2014).

[21] Bardhan, J., Park, S. & Makowski, L. "SoftWAXS: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules." *Journal of Applied Crystallography* **42**, 932–943 (2009).

[22] Merzel, F. & Smith, J.C. "SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins." *Acta Crystallographica Section D* **58**, 242–249 (2002).

[23] Tjioe, E. & Heller, W.T. "ORNL_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes." *Journal of Applied Crystallography* **40**, 782–785 (2007).

[24] Park, S. & Bardhan, J. "Simulated x-ray scattering of protein solutions using explicit-solvent models." *The Journal of Chemical Physics* **130**, 134114 (2009).

[25] Yang, S., Park, S., Makowski, L. *et al.* "A rapid coarse residue-based computational method for X-ray solution scattering." *Biophysical Journal* **96**, 4449–4463 (2009).

[26] Ravikumar, K.M., Huang, W. & Yang, S. "Fast-SAXS-Pro: A Unified Approach to Computing SAXS Profiles of DNA, RNA, Protein, and their Complexes." *Biophysical Journal* **104**, 183a (2013).

[27] Pavlov, M.Y. & Fedorov, B.A. "Improved technique for calculating X-ray scattering intensity of biopolymers in solution: Evaluation of the form, volume, and surface of a particle." *Biopolymers* **22**, 1507–1522 (1983).

[28] Schneidman-Duhovny, D., Hammel, M. & Sali, A. "FoXS: a web server for rapid computation and fitting of SAXS profiles." *Nucleic Acids Research* **38**, W540–W544 (2010).

[29] Schneidman-Duhovny, D., Hammel, M., Tainer, J.A. *et al.* "Accurate SAXS profile computation and its assessment by contrast variation experiments." *Biophysical Journal* **105**, 962–974 (2013).

[30] Poitevin, F., Orland, H., Doniach, S. *et al.* "AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models." *Nucleic Acids Research* **39**, W184–W189 (2011).

[31] Virtanen, J.J., Makowski, L., Sosnick, T.R. *et al.* "Modeling the hydration layer around proteins: applications to small- and wide-angle x-ray scattering." *Biophysical Journal* **101**, 2061–2069 (2011).

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

[33] Liu, H., Morris, R.J., Hexemer, A. *et al.* "Computation of small-angle scattering profiles with three-dimensional Zernike polynomials." *Acta Crystallographica Section A* **68**, 278–285 (2012).

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

[35] Aranda, R., Levin, E.J., Schotte, F. *et al.* "Time-dependent atomic coordinates for

the dissociation of carbon monoxide from myoglobin." *Acta Crystallographica Section D* **62**, 776–783 (2006).

[36] Fermi, G., Perutz, M., Shaanan, B. *et al.* "The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution." *Journal of Molecular Biology* **175**, 159–174 (1984).

[37] Silva, M.M., Rogers, P.H. & Arnone, A. "A third quaternary structure of human hemoglobin A at 1.7 Å resolution." *Journal of Biological Chemistry* **267**, 17248–17256 (1992).

[38] Heller, W.T. "ELLSTAT: shape modeling for solution small-angle scattering of proteins and protein complexes with automated statistical characterization." *Journal of Applied Crystallography* **39**, 671–675 (2006).

[39] Gumerov, N.A., Berlin, K., Fushman, D. *et al.* "A hierarchical algorithm for fast Debye summation with applications to small angle scattering." *Journal of Computational Chemistry* **33**, 1981–1996 (2012).

[40] Pantos, E. & Bordas, J. "Supercomputer simulation of small angle X-ray scattering, electron micrographs and X-ray diffraction patterns of macromolecular structures." *Pure and Applied Chemistry* **66**, 77–82 (1994).

[41] Stovgaard, K., Andreetta, C., Ferkinghoff-Borg, J. *et al.* "Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models." *BMC Bioinformatics* **11**, 429 (2010).

[42] Zheng, W. & Tekpinar, M. "Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell." *Biophysical Journal* **101**, 2981–2991 (2011).

[43] Daily, M.D., Makowski, L., Phillips, G.N. *et al.* "Large-scale motions in the adenylate kinase solution ensemble: coarse-grained simulations and comparison with solution X-ray scattering." *Chemical Physics* **396**, 84–91 (2012).

[44] Mortenson, D.E., Satyshur, K.A., Guzei, I.A. *et al.* "Quasiracemic crystallization as a tool to assess the accommodation of noncanonical residues in nativelike protein conformations." *Journal of the American Chemical Society* **134**, 2473–6 (2012).

[45] Marrink, S. & Risselada, H. "The MARTINI force field: coarse grained model for biomolecular simulations." *The Journal of Physical Chemistry B* **111**, 7812–7824 (2007).

[46] Monticelli, L., Kandasamy, S.K., Periole, X. *et al.* "The MARTINI coarse-grained force field: extension to proteins." *Journal of Chemical Theory and Computation* **4**, 819–834 (2008).

[47] Periole, X., Cavalli, M., Marrink, S.j. *et al.* "Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition." *Journal of Chemical Theory and Computation* **5**, 2531–2543 (2009).

[48] López, C.A., Rzepiela, A.J., de Vries, A.H. *et al.* "Martini Coarse-Grained Force Field: Extension to Carbohydrates." *Journal of Chemical Theory and Computation* **5**, 3195–3210 (2009).

[49] Yesylevskyy, S.O., Schäfer, L.V., Sengupta, D. *et al.* "Polarizable water model for the coarse-grained MARTINI force field." *PLoS Computational Biology* **6**, e1000810 (2010).

[50] de Jong, D.H., Singh, G., Bennett, W.F.D. *et al.* "Improved Parameters for the Martini Coarse-Grained Protein Force Field." *Journal of Chemical Theory and Computation* **9**, 687–697 (2013).

[51] Marrink, S.J. & Tieleman, D.P. "Perspective on the Martini model." *Chemical Society Reviews* **42**, 6801–6822 (2013).

[52] Louhivuori, M., Risselada, H.J., van der Giessen, E. *et al.* "Release of content through mechano-sensitive gates in pressurized liposomes." *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19856–19860 (2010).

[53] Schäfer, L.V., de Jong, D.H., Holt, A. *et al.* "Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes." *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1343–1348 (2011).

[54] van den Bogaart, G., Meyenberg, K., Risselada, H.J. *et al.* "Membrane protein sequestering by ionic protein-lipid interactions." *Nature* **479**, 552–555 (2011).

[55] Oberg, K.A., Ruysschaert, J.M. & Goormaghtigh, E. "Rationally selected basis proteins: a new approach to selecting proteins for spectroscopic secondary structure analysis." *Protein Science* **12**, 2015–2031 (2003).

[56] Forman, P. "The discovery of the diffraction of X-rays by crystals; a critique of the myths." *Archive for History of Exact Sciences* **6**, 38–71 (1969).

[57] Eckert, M. "Max von Laue and the discovery of X-ray diffraction in 1912." *Annalen der Physik* **524**, A83–A85 (2012).

[58] Bragg, W.L. "The diffraction of short electromagnetic waves by a crystal." *Proceedings of the Cambridge Philosophical Society* **17**, 43 (1913).

[59] Bragg, W.H. & Bragg, W.L. "The Reflection of X-rays by Crystals." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **88**, 428–438 (1913).

[60] Tomboulian, D.H. & Hartman, P.L. "Spectral and Angular Distribution of Ultraviolet Radiation from the 300-MeV Cornell Synchrotron." *Physical Review* **102**, 1423–1447 (1956).

[61] Rosenbaum, G., Holmes, K.C. & Witz, J. "Synchrotron Radiation as a Source for X-ray Diffraction." *Nature* **230**, 434–437 (1971).

[62] Helliwell, J.R., Greenhough, T.J., Carr, P.D. *et al.* "Central data collection facility for protein crystallography, small angle diffraction and scattering at the Daresbury Laboratory Synchrotron Radiation Source (SRS), England." *Journal of Physics E: Scientific Instruments* **15**, 1363–1372 (1982).

[63] Lindley, P.F. "The use of synchrotron radiation in protein crystallography." *Radiation Physics and Chemistry* **45**, 367–383 (1995).

[64] Graber, T., Anderson, S., Brewer, H. *et al.* "BioCARS: a synchrotron resource for time-resolved X-ray science." *Journal of Synchrotron Radiation* **18**, 658–670 (2011).

[65] Emma, P., Akre, R., Arthur, J. *et al.* "First lasing and operation of an Ångstrom-wavelength free-electron laser." *Nature Photonics* **4**, 641–647 (2010).

[66] Jamison, S. "X-ray science: X-ray FEL shines brightly." *Nature Photonics* **4**, 589–591 (2010).

[67] Huang, Z. & Lindau, I. "Free-electron lasers: SACLA hard-X-ray compact FEL." *Nature Photonics* **6**, 505–506 (2012).

[68] Arnlund, D. *X-ray free-electron laser based methods for structural and ultrafast dynamics studies of a photosynthetic reaction centre.* Ph.D. thesis, University of Gothenburg (2014).

[69] Neutze, R., Wouts, R., van der Spoel, D. *et al.* "Potential for biomolecular imaging with femtosecond X-ray pulses." *Nature* **406**, 752–757 (2000).

[70] Barty, A., Caleman, C., Aquila, A. *et al.* "Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements." *Nature Photonics* **6**, 35–40 (2012).

[71] Henrich, B., Bergamaschi, A., Broennimann, C. *et al.* "PILATUS: A single photon counting pixel detector for X-ray applications." *Nuclear Instruments and Methods in Physics Research A* **607**, 247–249 (2009).

[72] Johnson, I., Bergamaschi, A., Billich, H. *et al.* "Eiger: a single-photon counting x-ray detector." *J. Instrum.* **9**, C05032–C05032 (2014).

[73] Westenhoff, S., Malmerberg, E., Arnlund, D. *et al.* "Rapid readout detector captures protein time-resolved WAXS." *Nature Methods* **7**, 775–776 (2010).

[74] Malmerberg, E. *Conformational dynamics of rhodopsins visualized by time-resolved wide angle x-ray scattering.* Ph.D. thesis, University of Gothenburg (2011).

[75] Wulff, M., Plech, A., Eybert, L. *et al.* "The realization of sub-nanosecond pump and probe experiments at the ESRF." *Faraday Discussions* **122**, 13–26 (2003).

[76] Cammarata, M., Eybert, L., Ewald, F. *et al.* "Chopper system for time resolved experiments with synchrotron radiation." *Review of Scientific Instruments* **80**, 015101 (2009).

[77] Malmerberg, E., Omran, Z., Hub, J.S. *et al.* "Time-resolved WAXS reveals accelerated conformational changes in iodoretinal-substituted proteorhodopsin." *Biophysical Journal* **101**, 1345–53 (2011).

[78] Malmerberg, E., M Bovee-Geurts, P.H., Katona, G. *et al.* "Conformational activation of visual rhodopsin in native disc membranes." *Science Signaling* **8**, ra26 (2015).

[79] Nasedkin, A., Marcellini, M., Religa, T.L. *et al.* "Deconvoluting Protein (Un)folding Structural Ensembles Using X-Ray Scattering, Nuclear Magnetic Resonance Spectroscopy and Molecular Dynamics Simulation." *PLoS ONE* **10**, e0125662 (2015).

[80] Neutze, R., Wouts, R., Techert, S. *et al.* "Visualizing photochemical dynamics in solution through picosecond x-ray scattering." *Physical Review Letters* **87**, 195508 (2001).

[81] Plech, A., Wulff, M., Bratos, S. *et al.* "Visualizing Chemical Reactions in Solution by Picosecond X-Ray Diffraction." *Physical Review Letters* **92**, 125505 (2004).

[82] Davidsson, J., Poulsen, J., Cammarata, M. *et al.* "Structural Determination of a Transient Isomer of CH2I2 by Picosecond X-Ray Diffraction." *Physical Review Letters* **94**, 245503 (2005).

[83] Ihee, H., Lorenc, M., Kim, T.K. *et al.* "Ultrafast x-ray diffraction of transient molecular structures in solution." *Science* **309**, 1223–1227 (2005).

[84] Cho, H.S., Dashdorj, N., Schotte, F. *et al.* "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* **107**, 7281–7286 (2010).

[85] Kim, K., Muniyappan, S. & Oang, K. "Direct Observation of Cooperative Protein Structural Dynamics of Homodimeric Hemoglobin from 100 ps to 10 ms with Pump–Probe X-ray Solution Scattering." *Journal of the American Chemical Society* **134**, 7001–7008 (2012).

[86] Spilotros, A., Levantino, M., Schirò, G. *et al.* "Probing in cell protein structural changes with time-resolved X-ray scattering." *Soft Matter* **8**, 6434 (2012).

[87] Kim, T.W., Lee, J.H., Choi, J. *et al.* "Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe X-ray solution scattering." *Journal of the American Chemical Society* **134**, 3145–3153 (2012).

[88] Kim, K.H., Lee, J.H., Kim, J. *et al.* "Solvent-Dependent Molecular Structure of Ionic Species Directly Measured by Ultrafast X-Ray Solution Scattering." *Physical Review Letters* **110**, 165505 (2013).

[89] Kim, K.H., Ki, H., Oang, K.Y. *et al.* "Global reaction pathways in the photodissociation of I3 (-) ions in solution at 267 and 400 nm studied by picosecond X-ray liquidography." *ChemPhysChem* **14**, 3687–3697 (2013).

[90] Takala, H., Björling, A., Berntsson, O. *et al.* "Signal amplification and transduction in phytochrome photosensors." *Nature* **509**, 245–248 (2014).

[91] Arnlund, D., Johansson, L.C., Wickstrand, C. *et al.* "Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser." *Nature Methods* **11**, 923–926 (2014).

[92] Kim, K.H., Kim, J.G., Nozawa, S. *et al.* "Direct observation of bond formation in solution with femtosecond X-ray scattering." *Nature* **518**, 385–389 (2015).

[93] Levantino, M., Yorke, B.A., Monteiro, D.C. *et al.* "Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules." *Current Opinion in Structural Biology* **35**, 41–48 (2015).

[94] West, J.M., Xia, J., Tsuruta, H. *et al.* "Time evolution of the quaternary structure of Escherichia coli aspartate transcarbamoylase upon reaction with the natural substrates and a slow, tight-binding inhibitor." *Journal of Molecular Biology* **384**, 206–218 (2008).

[95] Akiyama, S., Takahashi, S., Kimura, T. *et al.* "Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering." *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1329–1334 (2002).

[96] Pollack, L., Tate, M.W., Darnton, N.C. *et al.* "Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering." *Proceedings of the National Academy of Sciences of the United States of America* **96**, 10115–10117 (1999).

[97] Lafleur, J.P., Snakenborg, D., Nielsen, S.S. *et al.* "Automated microfluidic sample-preparation platform for high-throughput structural investigation of proteins by small-angle X-ray scattering." *Journal of Applied Crystallography* **44**, 1090–1099 (2011).

[98] Köster, S. & Pfohl, T. "X-ray studies of biological matter in microfluidic environments." *Modern Physics Letters B* **26**, 1230018 (2012).

[99] Lee, H.M., Larson, D.R. & Lawrence, D.S. "Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds." *ACS Chemical Biology* **4**, 409–427 (2009).

[100] Kubelka, J. "Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics." *Photochemical & Photobiological Sciences* **8**, 499–512 (2009).

[101] Pabst, G., Rappolt, M., Amenitsch, H. *et al.* "X-ray Kinematography of Temperature-Jump Relaxation Probes the Elastic Properties of Fluid Bilayers." *Langmuir* **16**, 8994–9001 (2000).

[102] Renner, W., Mandelkow, E.M., Mandelkow, E. *et al.* "Self-assembly of microtubule protein studied by time-resolved X-ray scattering using temperature jump and stopped flow." *Nuclear Instruments and Methods in Physics Research* **208**, 535–540 (1983).

[103] Shkumatov, A.V., Chinnathambi, S., Mandelkow, E. *et al.* "Structural memory of natively unfolded tau protein detected by small-angle X-ray scattering." *Proteins* **79**, 2122–2131 (2011).

[104] Ihee, H., Wulff, M., Kim, J. *et al.* "Ultrafast X-ray scattering: structural dynamics from diatomic to protein molecules." *International Reviews in Physical Chemistry* **29**, 453–520 (2010).

[105] Stengel, G., Zahn, R. & Höök, F. "DNA-induced programmable fusion of phospholipid vesicles." *Journal of the American Chemical Society* **129**, 9584–9585 (2007).

[106] Stengel, G. & Simonsson, L. "Determinants for membrane fusion induced by cholesterol-modified DNA zippers." *The Journal of Physical Chemistry B* **112**, 8264–8274 (2008).

[107] Jeffries, C.M., Graewert, M.A., Svergun, D.I. *et al.* "Limiting radiation damage for high-brilliance biological solution scattering: practical experience at the EMBL P12 beamline PETRAIII." *Journal of Synchrotron Radiation* **22**, 273–279 (2015).

[108] Jacques, D.A. & Trewhella, J. "Small-angle scattering for structural biology – expanding the frontier while avoiding the pitfalls." *Protein Science* **19**, 642–657 (2010).

[109] Skinner, L.B., Benmore, C.J. & Parise, J.B. "Area detector corrections for high quality synchrotron X-ray structure factor measurements." *Nuclear Instruments and Methods in Physics Research A* **662**, 61–70 (2012).

[110] Shannon, C.E. "A Mathematical Theory of Communication." *Bell System Technical*

Journal **27**, 379–423 (1948).

[111] Shannon, C.E. & Weaver, W. *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois (1949).

[112] Moore, P.B. "Small-angle scattering. Information content and error analysis." *Journal of Applied Crystallography* **13**, 168–175 (1980).

[113] Kirian, R.A., Schmidt, K.E., Wang, X. *et al.* "Signal, noise, and resolution in correlated fluctuations from snapshot small-angle x-ray scattering." *Physical Review E* **84**, 011921 (2011).

[114] Rambo, R.P. & Tainer, J.A. "Accurate assessment of mass, models and resolution by small-angle scattering." *Nature* **496**, 477–481 (2013).

[115] Pedersen, M.C., Hansen, S.L., Markussen, B. *et al.* "Quantification of the information in small-angle scattering data." *Journal of Applied Crystallography* **47**, 2000–2010 (2014).

[116] Brinkmann, L.U.L. & Hub, J.S. "Anisotropic time-resolved solution X-ray scattering patterns from explicit-solvent molecular dynamics." *The Journal of Chemical Physics* **143**, 104108 (2015).

[117] Brünger, A.T., Kuriyan, J. & Karplus, M. "Crystallographic R factor refinement by molecular dynamics." *Science* **235**, 458–460 (1987).

[118] Güntert, P., Mumenthaler, C. & Wüthrich, K. "Torsion angle dynamics for NMR structure calculation with the new program DYANA." *Journal of Molecular Biology* **273**, 283–298 (1997).

[119] Brünger, A.T., Adams, P.D., Clore, G.M. *et al.* "Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination." *Acta Crystallographica Section D* **54**, 905–921 (1998).

[120] Brunger, A.T. & Adams, P.D. "Molecular Dynamics Applied to X-ray Structure Refinement." *Accounts of Chemical Research* **35**, 404–412 (2002).

[121] Schwieters, C.D., Kuszewski, J.J., Tjandra, N. *et al.* "The Xplor-NIH NMR molecular structure determination package." *Journal of Magnetic Resonance* **160**, 65–73 (2003).

[122] Brünger, A.T. "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**, 472–475 (1992).

[123] Brünger, A.T. *X-PLOR Version 3.1: a System for X-ray Crystallography and NMR.* Yale University Press (1992).

[124] Murshudov, G.N., Vagin, A.A. & Dodson, E.J. "Refinement of macromolecular structures by the maximum-likelihood method." *Acta Crystallographica Section D* **53**, 240–255 (1997).

[125] Adams, P.D., Afonine, P.V., Bunkóczi, G. *et al.* "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallographica Section D* **66**, 213–221 (2010).

[126] Hura, G.L., Menon, A.L., Hammel, M. *et al.* "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)." *Nature Methods* **6**, 606–612 (2009).

[127] Petoukhov, M.V., Franke, D., Shkumatov, A.V. *et al.* "New developments in the ATSAS program package for small-angle scattering data analysis." *Journal of Applied Crystallography* **45**, 342–350 (2012).

[128] Putnam, C. & Hammel, M. "Solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." *Quarterly Reviews of Biophysics* **40**, 191–285 (2007).

[129] Franke, D., Jeffries, C.M. & Svergun, D.I. "Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra." *Nature Methods* **12**, 419–422 (2015).

[130] Glatter, O. "A new method for the evaluation of small-angle scattering data."

Journal of Applied Crystallography **10**, 415–421 (1977).

[131] Hansen, S. & Pedersen, J.S. "A comparison of three different methods for analysing small-angle scattering data." *Journal of Applied Crystallography* **24**, 541–548 (1991).

[132] Guallar, V., Jarzecki, A.A., Friesner, R.A. *et al.* "Modeling of Ligation-Induced Helix/Loop Displacements in Myoglobin: Toward an Understanding of Hemoglobin Allostery." *Journal of the American Chemical Society* **128**, 5427–5435 (2006).

[133] Ahn, S., Kim, K.H., Kim, Y. *et al.* "Protein tertiary structural changes visualized by time-resolved x-ray solution scattering." *The Journal of Physical Chemistry B* **113**, 13131–13133 (2009).

[134] Maiorov, V. & Abagyan, R. "Energy strain in three-dimensional protein structures." *Folding and Design* **3**, 259–269 (1998).

[135] Grishaev, A. & Wu, J. "Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data." *Journal of the American Chemical Society* **127**, 16621–16628 (2005).

[136] van der Spoel, D., Lindahl, E., Hess, B. *et al.* "GROMACS: fast, flexible, and free." *Journal of Computational Chemistry* **26**, 1701–1718 (2005).

[137] Hess, B., Kutzner, C., van der Spoel, D. *et al.* "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation." *Journal of Chemical Theory and Computation* pages 435–447 (2008).

[138] Pronk, S., Páll, S., Schulz, R. *et al.* "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit." *Bioinformatics* **29**, 845–854 (2013).

[139] Kimanius, D., Pettersson, I., Schluckebier, G. *et al.* "SAXS-Guided Metadynamics." *Journal of Chemical Theory and Computation* **11**, 3491–3498 (2015).

[140] Chen, P.c. & Hub, J.S. "Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics." *Biophysical Journal* **108**, 2573–2584 (2015).

[141] Hayward, S. "Structural principles governing domain motions in proteins." *Proteins: Structure, Function, and Genetics* **36**, 425–435 (1999).

[142] Oh, B.H., Pandit, J., Kang, C.H. *et al.* "Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand." *Journal of Biological Chemistry* **268**, 11348–11355 (1993).

[143] Yang, L.W., Eyal, E., Chennubhotla, C. *et al.* "Insights into Equilibrium Dynamics of Proteins from Comparison of NMR and X-Ray Data with Computational Predictions." *Structure* **15**, 741–749 (2007).

[144] Hirai, M., Iwase, H., Hayakawa, T. *et al.* "Structural hierarchy of several proteins observed by wide-angle solution scattering." *Journal of Synchrotron Radiation* **9**, 202–205 (2002).

[145] Chang, C. "The Two-Component System . Regulation of Diverse Signaling Pathways in Prokaryotes and Eukaryotes." *Plant Physiology* **117**, 723–731 (1998).

[146] Cohen, P. "The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture." *European Journal of Biochemistry* **268**, 5001–5010 (2001).

[147] Stock, A.M., Robinson, V.L. & Goudreau, P.N. "Two-component signal transduction." *Annual Review of Biochemistry* **69**, 183–215 (2000).

[148] Casino, P., Rubio, V. & Marina, A. "The mechanism of signal transduction by two-component systems." *Current Opinion in Structural Biology* **20**, 763–771 (2010).

[149] Robson, P.R., McCormac, A.C., Irvine, A.S. *et al.* "Genetic engineering of harvest index in tobacco through overexpression of a phytochrome gene." *Nature Biotechnology* **14**, 995–998 (1996).

[150] Fischer, A.J. & Lagarias, J.C. "Harnessing phytochrome's glowing potential." *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17334–17339 (2004).

[151] Auldridge, M.E., Satyshur, K.A., Anstrom, D.M. *et al.* "Structure-guided Engineering Enhances a Phytochrome-based Infrared Fluorescent Protein." *Journal of Biological Chemistry* **287**, 7000–7009 (2012).

[152] Pathak, G.P., Vrana, J.D. & Tucker, C.L. "Optogenetic control of cell function using engineered photoreceptors." *Biology of the Cell* **105**, 59–72 (2013).

[153] Borthwick, H.A., Hendricks, S.B., Parker, M.W. *et al.* "A Reversible Photoreaction Controlling Seed Germination." *Proceedings of the National Academy of Sciences of the United States of America* **38**, 662–666 (1952).

[154] Butler, W.L., Norris, K.H., Siegelman, H.W. *et al.* "Detection, assay, and preliminary purification of the pigment controlling photoresponsive development of plants." *Proceedings of the National Academy of Sciences of the United States of America* **45**, 1703–1708 (1959).

[155] Litts, J.C., Kelly, J.M. & Lagarias, J.C. "Structure-function studies on phytochrome. Preliminary characterization of highly purified phytochrome from Avena sativa enriched in the 124-kilodalton species." *Journal of Biological Chemistry* **258**, 11025–11031 (1983).

[156] Vierstra, R.D. & Quail, P.H. "Purification and initial characterization of 124 kDalton phytochrome from Avena." *Biochemistry* **22**, 2498–2505 (1983).

[157] Quail, P.H. "Phytochrome: a light-activated molecular switch that regulates plant gene expression." *Annual Review of Genetics* **25**, 389–409 (1991).

[158] Franklin, K.A. & Quail, P.H. "Phytochrome functions in Arabidopsis development." *Journal of Experimental Botany* **61**, 11–24 (2010).

[159] Vierstra, R.D. & Zhang, J. "Phytochrome signaling: solving the Gordian knot with microbial relatives." *Trends in Plant Science* **16**, 417–426 (2011).

[160] Burgie, E.S., Bussell, A.N., Walker, J.M. *et al.* "Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome." *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10179–10184 (2014).

[161] Kehoe, D.M. & Grossman, A.R. "Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors." *Science* **273**, 1409–1412 (1996).

[162] Kaneko, T., Sato, S., Kotani, H. *et al.* "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions." *DNA Research* **3**, 109–136 (1996).

[163] Hughes, J., Lamparter, T., Mittmann, F. *et al.* "A prokaryotic phytochrome." *Nature* **386**, 663–663 (1997).

[164] Yeh, K.C., Wu, S.H., Murphy, J.T. *et al.* "A cyanobacterial phytochrome two-component light sensory system." *Science* **277**, 1505–1508 (1997).

[165] Davis, S.J., Vener, A.V. & Vierstra, R.D. "Bacteriophytochromes: phytochrome-Like photoreceptors from nonphotosynthetic eubacteria." *Science* **286**, 2517–2520 (1999).

[166] Vierstra, R.D. & Davis, S.J. "Bacteriophytochromes: new tools for understanding phytochrome signal transduction." *Seminars in Cell & Developmental Biology* **11**, 511–521 (2000).

[167] Rockwell, N.C., Su, Y.S. & Lagarias, J.C. "Phytochrome structure and signalling mechanisms." *Annual Review of Plant Biology* **57**, 837–858 (2006).

[168] Möglich, A., Yang, X., Ayers, R.A. *et al.* "Structure and function of plant photoreceptors." *Annual Review of Plant Biology* **61**, 21–47 (2010).

[169] Nagatani, A. "Phytochrome: structural basis for its functions." *Current Opinion in Plant Biology* **13**, 565–570 (2010).

[170] Burgie, E.S., Wang, T., Bussell, A.N. *et al.* "Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion." *Journal of Biological Chemistry* **289**, 24573–24587 (2014).

[171] Wagner, J.R., Brunzelle, J.S., Forest, K.T. *et al.* "A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome." *Nature* **438**, 325–331 (2005).

[172] Wagner, J., Zhang, J. & Brunzelle, J. "High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution." *Journal of Biological Chemistry* **282**, 12298–12309 (2007).

[173] Yang, X., Stojkovic, E.A., Kuk, J. *et al.* "Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion." *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12571–12576 (2007).

[174] Yang, X., Kuk, J. & Moffat, K. "Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: Photoconversion and signal transduction." *Proceedings of the National Academy of Sciences of the United States of America* **105**, 14715–14720 (2008).

[175] Essen, L., Mailliet, J. & Hughes, J. "The structure of a complete phytochrome sensory module in the Pr ground state." *Proceedings of the National Academy of Sciences of the United States of America* **105**, 14709–14714 (2008).

[176] Yang, X., Kuk, J. & Moffat, K. "Conformational differences between the Pfr and Pr states in Pseudomonas aeruginosa bacteriophytochrome." *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15639–15644 (2009).

[177] Yang, X., Ren, Z., Kuk, J. *et al.* "Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome." *Nature* **479**, 428–432 (2011).

[178] Bellini, D. & Papiz, M.Z. "Structure of a bacteriophytochrome and light-stimulated protomer swapping with a gene repressor." *Structure* **20**, 1436–1346 (2012).

[179] Bellini, D. & Papiz, M.Z. "Dimerization properties of the RpBphP2 chromophore-binding domain crystallized by homologue-directed mutagenesis." *Acta Crystallographica Section D* **68**, 1058–1066 (2012).

[180] Anders, K., Daminelli-Widany, G., Mroginski, M.A. *et al.* "Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling." *Journal of Biological Chemistry* **288**, 35714–35725 (2013).

[181] Yang, X., Stojković, E.A., Ozarowski, W.B. *et al.* "Light Signaling Mechanism of Two Tandem Bacteriophytochromes." *Structure* **23**, 1179–1189 (2015).

[182] Elich, T.D. & Chory, J. "Phytochrome: If It Looks and Smells Like a Histidine Kinase, Is It a Histidine Kinase?" *Cell* **91**, 713–716 (1997).

[183] Yeh, K.C. & Lagarias, J.C. "Eukaryotic phytochromes: Light-regulated serine/threonine protein kinases with histidine kinase ancestry." *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13976–13981 (1998).

[184] Fankhauser, C. "Phytochromes as light-modulated protein kinases." *Seminars in Cell & Developmental Biology* **11**, 467–473 (2000).

[185] Auldridge, M.E. & Forest, K.T. "Bacterial phytochromes: more than meets the light." *Critical Reviews in Biochemistry and Molecular Biology* **46**, 67–88 (2011).

[186] Lamparter, T., Michael, N., Mittmann, F. *et al.* "Phytochrome from Agrobacterium tumefaciens has unusual spectral properties and reveals an N-terminal chromophore attachment site." *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11628–11633 (2002).

[187] Karniol, B. & Vierstra, R.D. "The pair of bacteriophytochromes from Agrobacterium tumefaciens are histidine kinases with opposing photobiological properties." *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2807–2812 (2003).

[188] Goodsell, D. "Two-Component Systems." *doi:10.2210/rcsb_pdb/mom_2015_10* (2015).

[189] Marina, A., Waldburger, C.D. & Hendrickson, W.A. "Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein." *The EMBO Journal* **24**, 4247–4259 (2005).

[190] Casino, P., Rubio, V. & Marina, A. "Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction." *Cell* **139**, 325–336 (2009).

[191] Wang, C., Sang, J., Wang, J. *et al.* "Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains." *PLoS Biology* **11**, e1001493 (2013).

[192] Mechaly, A.E., Sassoon, N., Betton, J.M. *et al.* "Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation." *PLoS Biology* **12**, e1001776 (2014).

[193] Casino, P., Miguel-Romero, L. & Marina, A. "Visualizing autophosphorylation in histidine kinases." *Nature Communications* **5**, 3258 (2014).

[194] Diensthuber, R.P., Bommer, M., Gleichmann, T. *et al.* "Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators." *Structure* **21**, 1127–1136 (2013).

[195] Bhate, M.P., Molnar, K.S., Goulian, M. *et al.* "Signal Transduction in Histidine Kinases: Insights from New Structures." *Structure* **23**, 981–994 (2015).

[196] Drozdetskiy, A., Cole, C., Procter, J. *et al.* "JPred4: a protein secondary structure prediction server." *Nucleic Acids Research* **43**, W389–W394 (2015).

[197] Tomomori, C., Tanaka, T., Dutta, R. *et al.* "Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ." *Nature Structural Biology* **6**, 729–734 (1999).

[198] Bhoo, S.H., Davis, S.J., Walker, J. *et al.* "Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore." *Nature* **414**, 776–779 (2001).

[199] Evans, K., Grossmann, J.G., Fordham-Skelton, A.P. *et al.* "Small-angle X-ray scattering reveals the solution structure of a bacteriophytochrome in the catalytically active Pr state." *Journal of Molecular Biology* **364**, 655–666 (2006).

[200] Baker, A.W. & Forest, K.T. "Structural biology: Action at a distance in a light receptor." *Nature* **509**, 174–175 (2014).

[201] Goodner, B., Hinkle, G., Gattung, S. *et al.* "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58." *Science* **294**, 2323–2328 (2001).

[202] Wood, D.W., Setubal, J.C., Kaul, R. *et al.* "The genome of the natural genetic engineer Agrobacterium tumefaciens C58." *Science* **294**, 2317–2323 (2001).

[203] Zienicke, B. *Fluorescence properties and photoconversion of model phytochromes Agp1 and Agp2 from Agrobacterium tumefaciens.* Ph.D. thesis, KIT Karlsruhe (2011).

[204] Dueber, J.E., Yeh, B.J., Chak, K. *et al.* "Reprogramming control of an allosteric signaling switch through modular recombination." *Science* **301**, 1904–1908 (2003).

[205] Möglich, A., Ayers, R.A. & Moffat, K. "Design and signaling mechanism of light-regulated histidine kinases." *Journal of Molecular Biology* **385**, 1433–1444 (2009).

[206] Wagner, J.R., Zhang, J., von Stetten, D. *et al.* "Mutational analysis of Deinococcus radiodurans bacteriophytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes." *Journal of Biological Chemistry* **283**, 12212–12226 (2008).

[207] Zienicke, B., Molina, I., Glenz, R. *et al.* "Unusual spectral properties of bacteriophytochrome Agp2 result from a deprotonation of the chromophore in the red-absorbing form Pr." *Journal of Biological Chemistry* **288**, 31738–31751 (2013).

[208] Linschitz, H., Kasche, V., Butler, W.L. *et al.* "The Kinetics of Phytochrome Conversion." *J. Biol. Chem.* **241**, 3395–3403 (1966).

[209] Kendrick, R.E. & Spruit, C.J.P. "Phototransformations of Phytochrome." *Photochemistry and Photobiology* **26**, 201–214 (1977).

[210] Aramendia, P.F., Ruzsicska, B.P., Braslavsky, S.E. *et al.* "Laser flash photolysis of 124-kilodalton oat phytochrome in water and deuterium oxide solutions: formation and decay of the I700 intermediates." *Biochemistry* **26**, 1418–1422 (1987).

[211] Schmidt, P., Gertsch, T., Remberg, A. *et al.* "The Complexity of the Pr to Pfr Phototransformation Kinetics Is an Intrinsic Property of Native Phytochrome." *Photochemistry and Photobiology* **68**, 754–761 (1998).

[212] Robben, U., Lindner, I. & Gärtner, W. "New open-chain tetrapyrroles as chromophores in the plant photoreceptor phytochrome." *Journal of the American Chemical Society* **130**, 11303–11311 (2008).

[213] Remberg, A., Lindner, I., Lamparter, T. *et al.* "Raman spectroscopic and light-induced kinetic characterization of a recombinant phytochrome of the cyanobacterium Synechocystis." *Biochemistry* **36**, 13389–13395 (1997).

[214] van Thor, J.J., Borucki, B., Crielaard, W. *et al.* "Light-Induced Proton Release and Proton Uptake Reactions in the Cyanobacterial Phytochrome Cph1." *Biochemistry* **40**, 11460–11471 (2001).

[215] Borucki, B., von Stetten, D., Seibeck, S. *et al.* "Light-induced proton release of phytochrome is coupled to the transient deprotonation of the tetrapyrrole chromophore." *Journal of Biological Chemistry* **280**, 34358–34364 (2005).

[216] Borucki, B., Seibeck, S., Heyn, M.P. *et al.* "Characterization of the covalent and noncovalent adducts of Agp1 phytochrome assembled with biliverdin and phycocyanobilin by circular dichroism and flash photolysis." *Biochemistry* **48**, 6305–17 (2009).

[217] Heyne, K., Herbst, J., Stehlik, D. *et al.* "Ultrafast dynamics of phytochrome from the cyanobacterium synechocystis, reconstituted with phycocyanobilin and phycoerythrobilin." *Biophysical Journal* **82**, 1004–1016 (2002).

[218] van Thor, J.J., Ronayne, K.L. & Towrie, M. "Formation of the early photoproduct lumi-R of cyanobacterial phytochrome cph1 observed by ultrafast mid-infrared spectroscopy." *Journal of the American Chemical Society* **129**, 126–132 (2007).

[219] Dasgupta, J., Frontiera, R.R., Taylor, K.C. *et al.* "Ultrafast excited-state isomerization in phytochrome revealed by femtosecond stimulated Raman spectroscopy." *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1784–1789 (2009).

[220] Toh, K.C., Stojkovic, E.A., van Stokkum, I.H.M. *et al.* "Proton-transfer and hydrogen-bond interactions determine fluorescence quantum yield and photochemical efficiency of bacteriophytochrome." *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9170–9175 (2010).

[221] Andel, F., Lagarias, J.C. & Mathies, R.A. "Resonance raman analysis of chromophore structure in the lumi-R photoproduct of phytochrome." *Biochemistry* **35**, 15997–16008 (1996).

[222] Kneip, C., Hildebrandt, P., Schlamann, W. *et al.* "Protonation State and Structural Changes of the Tetrapyrrole Chromophore during the Pr to Pfr Phototransformation of Phytochrome: A Resonance Raman Spectroscopic Study." *Biochemistry* **38**, 15185–15192 (1999).

[223] Foerstendorf, H., Lamparter, T., Hughes, J. *et al.* "The Photoreactions of Recombinant Phytochrome from the Cyanobacterium Synechocystis: A Low-Temperature UV-Vis and FT-IR Spectroscopic Study." *Photochemistry and Photobiology* **71**, 655–661 (2000).

[224] Mroginski, M.A., Murgida, D.H. & Hildebrandt, P. "The chromophore structural changes during the photocycle of phytochrome: a combined resonance Raman and quantum chemical approach." *Accounts of Chemical Research* **40**, 258–266 (2007).

[225] Lundholm, I. *Terahertz radiation as a pump and probe for studying low frequency*

vibrations in proteins. Ph.D. thesis, University of Gothenburg (2015).

[226] Xu, C., Tobi, D. & Bahar, I. "Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T-R2 Transition." *Journal of Molecular Biology* **333**, 153–168 (2003).

[227] Tobi, D. & Bahar, I. "Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state." *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18908–18913 (2005).