

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

**Isotope Labeling of Proteins for
Spectroscopic Studies**

Emil Gustavsson



UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology
Gothenburg, 2018

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

Isotope Labeling of Proteins for Spectroscopic Studies

Emil Gustavsson

Cover: A bacteriophytochrome in Pr and Pfr state on a [1H,15N]-TROSY spectrum

Copyright ©2018 by Emil Gustavsson

ISBN 978-91-7833-079-9 (Print)

ISBN 978-91-7833-080-5 (PDF)

Available online at <http://hdl.handle.net/2077/57807>

Department of Chemistry and Molecular Biology

Division of Biochemistry and Structural Biology

University of Gothenburg

SE-405 30, Göteborg, Sweden

Printed by BrandFactory AB

Göteborg, Sweden, 2018

Till Farfar
Din brinnande passion för naturen är den största anledningen till att jag tagit mig hit.

Emil Gustavsson

Abstract

Protein structural dynamics have been closely tied to the molecular mechanisms controlling for example enzyme activities. In my work, I have mainly studied the bacteriophytochrome from *Deinococcus radiodurans*. Phytochromes are photoreceptors that monitor the level, intensity, duration and color of environmental light to regulate numerous fundamental photomorphogenic processes in plants, fungi, and bacteria. In order to monitor the dynamic nature of such proteins, they are preferably investigated in solution. NMR and IR spectroscopy are ideal techniques for studying proteins in solution, but can suffer from spectral overlap on larger systems.

To investigate large proteins, such as the phytochromes studied in this thesis, biochemical modifications can be applied. Introduction of isotopes, either uniformly, amino acid-selectively or site-selectively make it possible to study the dynamics of large proteins even down to selected residues.

In this thesis, I developed and applied uniform, amino acid-selective and site-selective methods to isotopically label proteins, primarily a bacteriophytochrome. Solution NMR and IR spectroscopy were then applied to probe the structural dynamics of these proteins. The investigations reveal both how structural heterogeneity couples to the photomodulation in phytochromes, and how isomerization of the chromophore is linked to the global structural rearrangement in the sensory receptor. Further, the bacteriophytochrome knot region was investigated, and the results suggest the knot as a novel signaling pathway in phytochromes.

Site-selective isotope labeling, together with femtosecond IR spectroscopy, could also confirm an intermediate state in GFP fluorescent signaling.

My work has enabled spectroscopic investigations to monitor active sites and structural dynamics of proteins in a highly selective manner. It revealed structural and dynamical information underpinning the function of phytochrome photoreceptors.

Acknowledgements

There are many people with whom I have crossed paths during these years, which have had a large impact on me both professionally and personally, and to whom I owe a great deal of gratitude.

First and foremost my supervisor Sebastian, without whom my road to getting a PhD would have been cut short indeed had he not enrolled me. And for always keeping me grounded while still forcing me to push my boundaries I am forever grateful.

I also owe Linnéa a tremendous thanks for everything. When I have been freaking out over this or that you have remained steady as a rock, always with a "Vi löser detta, inga problem!" close at hand. Working close as we have, I am very grateful for your constant support, and for tolerating my sometimes not completely rational thoughts and behavior.

The Westenhoff group has been a a truly developing place to do my PhD, and I am so grateful for all the people, past and present, with whom I have had a chance to work and interact. All beamtimes, coffee breaks and beer clubs have had a huge impact on me on both a personal and professional level, and you have all been a huge part of that so thank you!

And to all the Lundberg people, you have been fantastic colleagues and I am very glad to have met and gotten to know all of you. This will be a chapter of my life which I will never forget!

This work would not have been possible had it not been for all the fantastic collaborations, and to all the people I have met in my travels I thank you.

Min familj och mina vänner har alltid varit ett fantastiskt stöd genom allt, även vid tillfällen då jag pratat ren grekiska. Förhoppningsvis kan den här avhandlingen ge åtminstone lite klarhet i vad det faktiskt är jag sysslat med de senaste åren.

I also want to thank my examiner Richard, who always tend to conclude with the question: "What was the best part?"

And to that question I can safely say it had nothing to do with science. It was meeting the love of my life, Jenny.

Jenny, you are the best thing that ever happened to me and I am looking forward with excitement to all the years ahead of us. Thank you for always supporting me, no matter what! Ich liebe dich!

Publications

This thesis consists of the following research papers:

- PAPER I:** E. Gustavsson*, L. Isaksson*, C. Persson, U. Brath, M. Mayzel, J. Ihalainen, G. Karlsson, V. Orekhov, and S. Westenhoff, "Solution NMR spectroscopy reveals that the PHY-tongue region of a bacteriophytochrome undergoes a soft-rigid transition when the protein is photoactivated" *Manuscript*, (2018)
- PAPER II:** E. Gustavsson*, L. Isaksson*, C. Persson, U. Brath, G. Karlsson, V. Orekhov, and S. Westenhoff, "The knot region of a bacterial phytochrome transduces the photo-signal towards the output domains" *Manuscript*, (2018)
- PAPER III:** J. A. Ihalainen, E. Gustavsson*, L. Schröder*, S. Donnini, H. Lehtivuori, L. Isaksson, C. Thöing, V. Modi, O. Berntsson, B. Stucki-Buchli, A. Liukkonen, H. Häkkinen, E. Kalenius, S. Westenhoff, and T. Kottke, "Chromophore-Protein Interplay During the Phytochrome Photocycle Revealed by Step-Scan FTIR Spectroscopy" *Journal of American Chemical Society*, Sep. 2018.
- PAPER IV:** S. Peuker, H. Andersson, E. Gustavsson, K. S. Maiti, R. Kania, A. Karim, S. Niebling, A. Pedersen, M. Erdelyi, and S. Westenhoff, "Efficient Isotope Editing of Proteins for Site-Directed Vibrational Spectroscopy" *Journal of American Chemical Society*, vol. 138, no. 7, pp. 2312–2318, Feb. 2016.
- PAPER V:** G. Kovacsova, E. Gustavsson, J. Wang, M. Kreir, S. Peuker, and S. Westenhoff, "Cell-free expression of a functional pore-only sodium channel" *Protein Expression and Purification*, vol. 111, pp. 42–47, Jul. 2015.

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

- PAPER VI:** A. Björling, O. Berntsson, H. Takala, K. D. Gallagher, H. Patel, **E. Gustavsson**, R. St. Peter, P. Duong, A. Nugent, F. Zhang, P. Berntsen, R. Appio, I. Rajkovic, H. Lehtivuori, M. R. Panman, M. Hoernke, S. Niebling, R. Harimoorthy, T. Lamparter, E. A. Stojković, J. A. Ihalainen, and S. Westenhoff, “Ubiquitous Structural Signaling in Bacterial Phytochromes” *Journal of Physical Chemistry Letters*, pp. 379–3383, 2015.
- PAPER VII:** H. Takala, S. Niebling, O. Berntsson, A. Björling, H. Lehtivuori, H. Häkkänen, M. Panman, **E. Gustavsson**, M. Hoernke, G. Newby, F. Zontone, M. Wulff, A. Menzel, J. A. Ihalainen, and S. Westenhoff, “Light-induced structural changes in a monomeric bacteriophytochrome” *Structural Dynamics*, vol. 3, no. 5, 2016.
- PAPER VIII:** O. Berntsson, R. P. Diensthuber, M. R. Panman, A. Björling, **E. Gustavsson**, M. Hoernke, A. J. Hughes, L. Henry, S. Niebling, H. Takala, J. A. Ihalainen, G. Newby, S. Kerruth, J. Heberle, M. Liebi, A. Menzel, R. Henning, I. Kosheleva, A. Möglich, and S. Westenhoff, “Sequential conformational transitions and α -helical supercoiling regulate a sensor histidine kinase” *Nature Communications*, vol. 8, no. 1, pp. 1–7, 2017.

PAPER IX: P. Edlund, H. Takala, E. Claesson, L. Henry, R. Dods, H. Lehtivuori, M. Panman, K. Pande, T. White, T. Nakane, O. Berntsson, **E. Gustavsson**, P. Båth, V. Modi, S. Roy-Chowdhury, J. Zook, P. Berntsen, S. Pandey, I. Poudyal, J. Tenboer, C. Kupitz, A. Barty, P. Fromme, J. D. Koralek, T. Tanaka, J. Spence, M. Liang, M. S. Hunter, S. Boutet, E. Nango, K. Moffat, G. Groenhof, J. Ihalainen, E. A. Stojković, M. Schmidt, and S. Westenhoff, “The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography” *Scientific Reports*, vol. 6, no. 1, p. 35279, 2016.

Contribution report

PAPER I: I, together with my co-supervisor, planned and designed the research. I produced and purified the amino acid-selectively labeled samples. I produced the ^2H , ^{13}C , ^{15}N -labeled protein samples, performed all the NMR experiments and did the major part of the backbone assignment, all together with my co-supervisor. We analyzed the data, and wrote the paper. I created the major part of the figures.

PAPER II: I, together with my co-supervisor, planned and designed the research. I produced the ^2H , ^{13}C , ^{15}N -labeled protein samples together with my co-supervisor. I performed all the NMR experiments together with collaborators. I, together with my co-supervisor, analyzed the data, and wrote the paper. I created the major part of the figures.

PAPER III: I, together with my supervisor, came up with the idea and strategy to isotope label the apoprotein. I produced and purified some of the protein. I, together with my co-supervisor, also produced and purified the cell-free produced amino acid-selectively labeled protein. I, together with collaborators, collected the step-scan FTIR data. I took part in writing the paper, and made several of the figures.

PAPER IV: I took part in the data acquisition, and made optimizations to the isotope labeling protocol. I also took part in writing and preparing figures for the paper.

PAPER V: I optimized the cell-free production protocol, and produced the sample for functional studies. I took part in making the functional studies, and in writing the paper.

Abbreviations

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

BV	BiliVerdin (chromophore)
CFPE	Cell- Free Protein Expression (<i>in vitro</i> protein production method)
CK	Creatine Kinase (enzyme)
Cph2	Cyanobacterial phytochrome-like protein 2 (family of phytochromes)
DrBphP	<i>Deinococcus radiodurans</i> Bacterio phytochrome (photoreceptor protein)
GAF	cGMP-specific phosphodiesterase-Adenylyl cyclase-FhlA (protein domain)
GFP	Green Fluorescent Protein
TROSY	Transverse Relaxation Optimized Spectroscopy (2D-NMR technique)
IPTG	IsoPropyl-β-D-1-ThioGalactopyranoside (molecular mimic of allolactose)
Na_vs	Voltage-gated sodium channel (protein family)
NMR	Nuclear Magnetic Resonance (experimental technique)
PϕB	PhytochromoBilin (chromophore)
PAS	Per Arnt Sim (protein domain)
PCB	PhycoCyanoBilin (chromophore)
PDB	Protein Data Bank (online data bank for protein structures)
pdCpA	5'-O-phosphoryl-2'-deoxyCytidylyl-(3'→5') Adenosine (dinucleotide fragment)
PHY	PHYtochrome specific domain (protein domain)
RDC	Residual Dipolar Coupling
RF1	Release Factor 1 (translation-terminating protein)
SEC	Size Exclusion Chromatography (protein purification technique)

Contents

Acknowledgements	vii
Abbreviations	xv
1 Introduction	1
1.1 Let there be light	1
1.1.1 Light sensitive proteins	2
1.1.2 Proteins: structure — function — dynamics	2
1.1.3 Probing protein dynamics	4
1.2 Phytochromes	5
1.2.1 Discovery and background	5
1.2.2 Structure and function of phytochromes	6
1.3 Green Fluorescent Protein	11
1.3.1 Discovery and background	11
1.3.2 Structure of GFP	11
1.4 Voltage-gated sodium channels	12
1.5 Scope of this thesis	13
2 Protein expression and purification	15
2.1 <i>In vivo</i> recombinant protein expression	16
2.1.1 Recombinant protein expression in <i>E. coli</i>	16
2.2 Cell-free protein expression	18
2.2.1 History of cell-free expression	19
2.2.2 The cell-free system and its components	20
2.2.3 Cell-free versus <i>in vivo</i> expression systems	24
2.3 Protein purification	26

3	Isotope labeling	29
3.1	Uniform isotope labeling <i>in vivo</i>	29
3.1.1	Applications with uniformly isotope labeled proteins	29
3.1.2	Principles of uniform isotope labeling	31
3.2	Amino acid-selective isotope labeling	32
3.2.1	Applications with amino acid-selectively labeled proteins	32
3.2.2	<i>In vivo</i> selective labeling	33
3.2.3	Selective labeling with CFPE	34
3.3	Site-selective isotope labeling	35
3.3.1	Principles of site-selective labeling	36
3.4	Summary	38
4	Spectroscopy	41
4.1	NMR spectroscopy of proteins	41
4.1.1	Short theory of protein NMR	42
4.1.2	Multidimensional protein NMR spectroscopy	44
4.2	Infrared spectroscopy of proteins	48
4.2.1	History of vibrational spectroscopy	48
4.2.2	Short theory of protein IR spectroscopy	48
4.2.3	Probing structural changes by IR spectroscopy	50
4.2.4	Step-scan FTIR of proteins	51
4.2.5	Ultra-fast IR spectroscopy of proteins	52
4.3	Summary	53
5	Concluding remarks	55
	Bibliography	57

Chapter 1

Introduction

1.1 Let there be light

Light is needed for almost all life on Earth, and evolution has made sure to take advantage of the photons released by the sun that reaches our pale blue dot. Light is used by life to drive photosynthesis, vision, signaling, controlling our day-night cycle and is also stored as chemical energy within the cells. Since light comes in different wavelengths, it has been crucial for life to develop ways of absorbing light from a broad spectrum of colors. The various ways organisms take advantage of light are of highest scientific interest as it is a key concept for understanding life itself [1].

Photosynthesis is one of the crucial ways life has found to utilize light energy. The light is absorbed, and through a series of chemical reactions the energy is transformed into chemical energy such as ATP. One of the side products of this reaction sequence is O_2 , for which most higher animals are forever grateful.

In order for photosynthesis to progress, sensing the variation in light conditions has also been of crucial importance. As plants stand deeply rooted in the ground, unable to move, they need to be able to detect whether something grows in front of them and gives shade, thereby hindering the vital photosynthesis. In response to this change in light they need to change how they grow, instead of another leaf perhaps a longer stem is better to get out of the shade. Hence, a way to sense different levels of shade is of utmost importance for plants,

and specific light-sensing biomolecules are responsible for this crucial task. These light-sensing biomolecules have been the main focus in my work.

1.1.1 Light sensitive proteins

As mentioned above, life has found many ways to capitalize on the energy coming from the sun. There are specific types of proteins which can sense light. They are responsible for absorbing the energy of the photon and transforming it to various biochemical signals depending on the proteins structure and function, and the wavelength of the photon. It is not surprising then that the number of proteins involved in capturing or sensing light numbers in the thousands [2].

To optimize the light uptake different organisms have had to develop proteins with different spectral characteristics, e.g. proteins that absorb in various spectral windows, or different wavelengths. Proteins are built up from amino acids, none of which absorb light in the visual range. Therefore certain cofactors absorbing visual light have been incorporated into the proteins throughout evolution. These cofactors are known as **chromophores**. The chromophores are able to absorb light over a broad wavelength range (Figure 1.1), and it depends on the structural details of the protein on how the energy is subsequently transferred and used in signaling.

1.1.2 Proteins: structure — function — dynamics

The name "protein" was first mentioned by the Swedish scientist Jöns Jacob Berzelius [3]. Proteins are responsible for the majority of the biochemical processes taking place in nature. They control everything from vision and neurotransmission to sending flashes of light and telling plants they are currently growing in the shade. The different proteins in our body alone numbers in the tens of thousands, each responsible for a specific function. Malfunctioning proteins is the cause of many diseases. Therefore, proteins have been a major research target since already the 1800's.

1.1. Let there be light

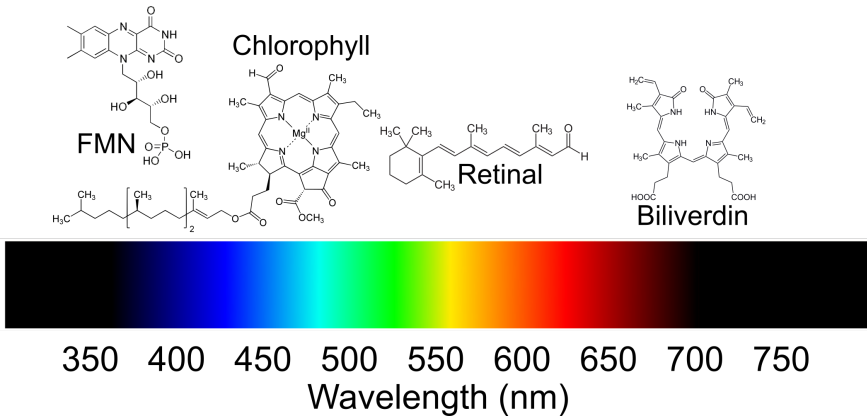


Figure 1.1: Selection of chromophores and their spectral absorption, ranging from UV to near-IR.

The structure of proteins is closely linked to their functions, a fact that has been known for decades. Consequently much effort has been put into determining the three-dimensional structures of these biomolecules. As a result the number of protein structures in the online databank currently numbers in the hundreds of thousands [2].

A clear majority of the structures solved to date were solved using X-ray crystallography. Protein crystals are illuminated with a high-energy X-ray beam, which causes diffraction patterns correlating to their structure [4]. In order to use this method, the protein first has to be crystallized, which is not trivial. The method can determine a protein structure to a very high resolution, but suffers from the very unnatural environment of crystals.

To clarify, when talking about structures special care has to be given to the fact that the word relates to several levels. The primary structure is the amino acid sequence, the secondary structure corresponds to whether the peptide forms for instance α -helices or β -sheets, which is mostly determined by the primary structure. The tertiary structure is the 3D structure of a single-polypeptide protein and what is usually referred to as protein structure. The 3D protein structure is therefore mostly dependent on the individual secondary

structure elements, while the quaternary structure describes the relative alignment of multi-polypeptide protein domains in large proteins.

However, knowing the complete 3D structure of a protein is in many cases not enough to understand how it functions. This could be compared to having a photograph of a car and trying to figure out what the car does and how it works. Filming the car in action is much more informative. Proteins are not rigid structures, they constantly change conformation and various substates exist in an equilibrium which can be perturbed to shift the conformational equilibrium towards one particular state. As with the car example, "filming" the protein in action could reveal much more information as the function of the protein is more closely related to its dynamical nature than the rigid crystal structure [5,6].

The dynamics–function relationship is extra clear when it comes to the group of intrinsically disordered proteins. This group of protein consists of fully unstructured to partially unstructured proteins, but nevertheless serve vital functions in cellular control mechanisms and signaling [7,8].

1.1.3 Probing protein dynamics

Other techniques than X-ray crystallography, either complementary or by themselves, are needed to determine the function of a protein. Preferably the proteins should be studied in as natural environment as possible, which often means in solution. Such techniques as nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy, small- and wide-angle x-ray scattering, or circular dichroism can be used in order to probe the structure and dynamics of proteins in solution in order to elucidate their function and physiological role [5,6].

In order to successfully probe dynamics of active sites or distinguish interesting domains, the proteins themselves may have to be modified. However, special care has to be taken that these modifications does not alter the proteins functionality. Otherwise all obtained results might not reflect the natural function of the protein studied.

Modifications can be as simple as truncation of certain domains, or advanced as incorporation of unnatural amino acids [9, 10].

For some of the above mentioned techniques, isotope labeling of the sample is needed as isotopes usually won't disturb the function. NMR spectroscopy requires a specific magnetically active type of isotope in order to detect the signal, and IR spectroscopy can be greatly helped by the introduction of certain isotopes at interesting sites. However, isotope labeling is in many cases not a trivial task, something that will be further detailed in Chapter 3.

1.2 Phytochromes

1.2.1 Discovery and background

The first indication of a photosensory control in plants was found as early as the 1920's when Garner and Allard reported on how plants responded to variations in the day-night cycle, with regard to growth, flowering and fruition [11]. The next major step was taken by Borthwick et al. in 1952, when they showed how seed germination could be controlled by red light and furthermore could be inhibited by illumination with far-red light. They even determined the absorbance maximum (Figure 1.2 shows the absorbance spectrum for *Deinococcus radiodurans* phytochrome monomer) for the activation and deactivation to 660nm and 710-750nm respectively [12].

Shortly after the biomolecule responsible was isolated and purified, it was named **Phytochrome**. Phytochromes are photosensory proteins and have since their original discovery been found in various other organisms such as fungi, bacteria and algae, but so far not in animals or archaea [13–17]. The well-spread hypothesis that plant chloroplasts originated as a symbiosis with a photosynthetic bacteria which was later incorporated into the cell lead researchers to believe that the phytochrome actually originated from a prokaryote [18, 19].

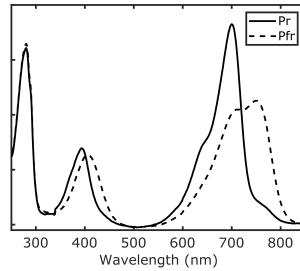


Figure 1.2: Pr and Pfr state of *Deinococcus radiodurans* phytochrome monomer. Resting (Pr) state shows an absorbance maximum at 700nm, while light-activated (Pfr) state shows a maximum at 750nm.

1.2.2 Structure and function of phytochromes

The reason behind the different responses to red and far-red light lies in the two meta-stable forms of phytochromes: one absorbing red light, the other far-red, named Pr and Pfr, respectively (Figure 1.2). The phytochrome family also shares a combination of highly conserved domains, these seem to be modular (see Figure 1.3 for a layout of the various modular compositions of different phytochromes) as various phytochrome families either lack certain domains or contain duplicates. The conserved domains building the photosensory core of the main family are: PAS (Per/Arnt/Sim), GAF (cGMP phosphodiesterase/Adenyl cyclase/Fhl 1) and PHY (Phytochrome specific).

In this protein family we find plant phytochromes, bacteriophytochromes, fungal phytochromes, diatom phytochromes and Cph1 from cyanobacteria. Cyanobacteria also utilize two other families of phytochromes which lack either the PAS domain, or both PAS and PHY. These are called Cph2s (after the first member of the family, *Synechocystis* Cph2) and CBCRs (cyanobacteriochromes), respectively [20,21]. Despite their differences in composition, all known phytochromes exist as homodimeric complexes, and are soluble proteins. In addition to the photosensory core, phytochromes also contain variations of C-terminal output domains, in bacteriophytochromes these are in many cases histidine kinases [22, 23], and may or may not have

1.2. Phytochromes

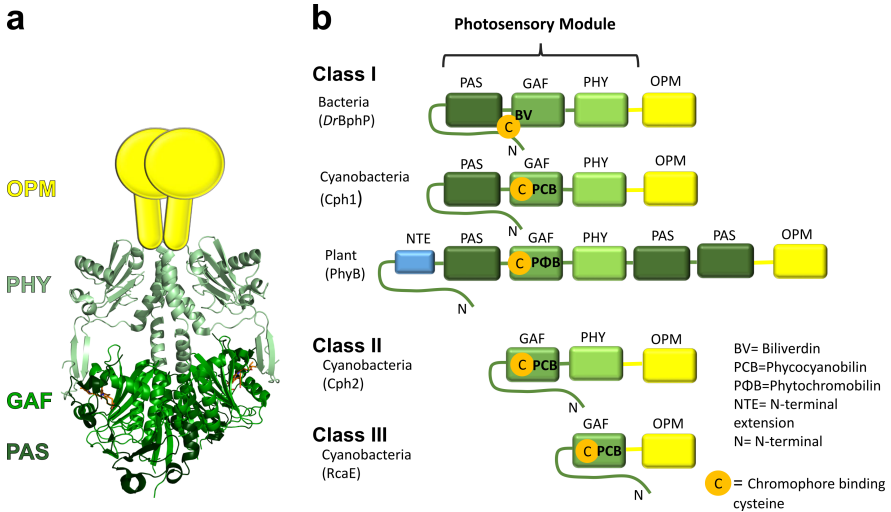


Figure 1.3: a) Homodimeric modular structure of phytochromes, consisting of PAS (dark green), GAF (green), PHY (light green) and output module (yellow). b) Phytochromes are divided in three classes determined by their modular structure.

an N-terminal extension domain thought to be involved in signaling [20,24].

Phytochromes also utilize variations of linear tetrapyrroles, so called bilins, as chromophores. These are obtained by oxidative degradation of heme to biliverdin (BV), a chromophore used by bacteriophytochromes and fungal phytochromes. The BV can then be further metabolized by ferredoxin-dependent bilin reductases into either PCB (phycocyanobilin) or PΦB (phytochromobilin) which are used by cyanobacteria or plant phytochromes, respectively [20, 24]. The chromophore is covalently bound to a cysteine, and although it can be bound to either the PAS (bacteriophytochromes) or GAF (cyanobacterial and plant phytochromes) domain, it resides within the GAF domain.

While the biological function of plant phytochromes is well-studied [25], the biological function of bacteriophytochrome from non-photosynthetic bacteria remains poorly understood. This, together with their simple

modular composition, resemblance to plant phytochromes, less complex photocycle and ease to produce in large quantities, has made them a prime target of investigation. If the complete biochemical mechanism of phytochromes can be elucidated, it would open up a world of possibilities to control and manipulate many biological functions of plants at atomic level.

Bacteriophytochrome from *Deinococcus radiodurans* (*DrBphP*) was the first phytochrome fragment (PAS-GAF domains only) to be crystallized, and therefore the first one to have the structure resolved at an atomic level [26]. Since then, *DrBphP* has been used as a model system for studying phytochromes, and several structures of different modular composition have been resolved [27–32].

The PAS-GAF-PHY fragment seen in Figure 1.4 shows a comparison of crystal structures of the Pr and Pfr states in *DrBphP*. The crystal structures have helped to shed light on many aspects of the phytochromes function. Among them the photochemistry of the chromophore involves an isomerization of the C15=C16 double-bond and thereby a flip of the D-ring (tetrapyrrole rings denoted A-D counting from the cysteine-bond) [27], which had previously been proposed [33,34].

The crystal structures also showed an opening of the PAS-GAF-PHY dimer and rearrangement of a conserved loop region called "tongue" from β -sheet to α -helix upon light illumination, which was proposed to be significant for signal transduction [31]. In **Paper I** we investigate the role of the tongue in transducing the signal and find that the tongue exerts a stabilizing effect in Pfr state by going from a structurally highly dynamic form to more rigid. We also show that the first arm is a β -strand, and not random coil, in Pfr and gets structurally more rigid, while the other changes conformation to α -helical.

Another interesting finding was the region in the PAS-GAF domain which forms a knot, a feature which is rare in protein structures [26]. The significance of this knotted region has still not been completely resolved, although it has been speculated to have a stabilizing effect on the PAS-GAF interface and chromophore binding pocket. In terms of how the signal is transduced from the chromophore to the output domain has not yet been resolved, the knot could play a more

1.2. Phytochromes

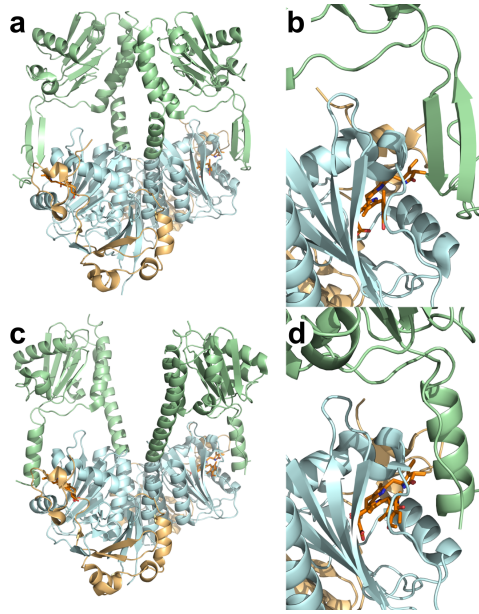


Figure 1.4: *DrBphP* crystal structure of PAS-GAF-PHY fragment. The different domains PAS, GAF and PHY are colored orange, cyan and green respectively. In a) the Pr structure is depicted (PDB ID 4O0P), where b) shows a zoom in of the tongue in β -sheet conformation. c) shows the Pfr structure with the dimer splayed apart, and a zoom in on the tongue in d) shows the change in conformation to α -helical.

significant role than believed. In **Paper II** we investigate the role of the knot, and propose a novel signaling pathway stretching from the propionic acid on the biliverdin B-ring, through the knot, to the helical spine connecting the PAS and GAF domains with the PHY and output domains.

As previously mentioned, the light absorption by the chromophore initiates an isomerization and flip of the D-ring. This encompasses the photocycle of the phytochrome, and includes in addition to the two meta-stable Pr and Pfr states several intermediate states (see Figure 1.5). Upon red-light illumination, the Lumi-R state is formed in picoseconds, whereafter the Meta-R state is formed in a matter of

microseconds. The meta-stable Pfr state is reached within a few milliseconds, a state which is stable for minutes up to hours or until a far-red light initiates the second half of the photocycle (Pfr \rightarrow Lumi-F \rightarrow Meta-F \rightarrow Pr).

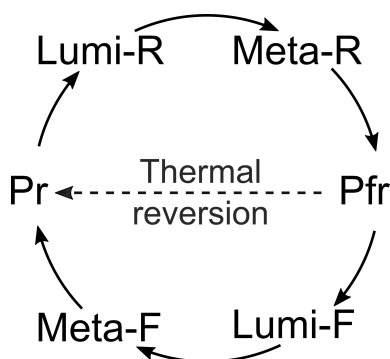


Figure 1.5: *DrBphP* photocycle The two meta-stable states Pr and Pfr, and the intermediate states Lumi-R, Meta-R, Lumi-F and Meta-F.

The light-activated state reverts back to resting state in darkness over a matter of minutes up to hours in a process called thermal or dark reversion. For phytochromes such as *DrBphP*, the resting and light-activated states are Pr and Pfr, respectively. There are however phytochromes which have the opposite thermal reversion and Pfr state as resting state. These phytochromes are categorized as "bathy" phytochromes while the Pr resting state phytochromes are categorized as "canonical".

Even though the structures of the two meta-stable Pr and Pfr states are well-known and the photocycle has been determined, the structural rearrangements during the intermediate steps are still poorly understood. In **Paper III**, we investigate the dynamics of the intermediate states and show how the hydrogen bonding network between the chromophore and protein changes during the course of the photocycle and how this couples to the global structural rearrangement of the phytochrome.

1.3 Green Fluorescent Protein

1.3.1 Discovery and background

Green Fluorescent Protein was first discovered in 1955 by Davenport and Nicol in the jellyfish *Aequorea victoria* when they reported a green fluorescence from the jellyfish upon UV illumination [35]. It was isolated in 1962 by Shimomura et al. [36] and has since then been discovered in other organisms. The green fluorescent protein (GFP) from *Aequorea victoria* remains the most widely used model system though. And although GFP has been known for many decades and is used heavily as a molecular tool, the biological function of this fluorescent protein has still not been determined.

1.3.2 Structure of GFP

The protein itself consists of 11 β -strands which together form a β -barrel type structure, as was discovered when the structure was solved in 1996 by Yang et al. [37]. It also has a self-assembling chromophore formed by the cyclization of serine 65 - tyrosine 66 - glycine 67 [38] and therefore does not need the addition of an external chromophore and can be readily expressed and detected in whole cells.

GFP has for a long time been used as a molecular tool in protein science, and is a highly overexpressing protein. A lot of effort has been put into changing its fluorescent properties (as was done by Crameri in 1996 [39]) to be used as a biochemical tool. It is being used as a co-expression partner, gene expression marker, co-localization marker, as a model system for developing new biochemical methods (as was done in **Paper IV**) and much more [40]. In **Paper IV and V**, GFP was used to screen for optimal expression conditions in a cell-free expression system, where the GFPcyc3 mutant from Crameri et al. was used [39].

Although GFP has been studied for a long time, its special spectral and fluorescent properties are still intensely studied [41]. GFP fluorescence is dependent on the self-assembly of the chromophore, also called maturation. The key amino acid in the chromophore is

tyrosine 66, which absorbs the photon causing the delocalization of an electron in the side-chain ring. This excited state lives for a few picoseconds before the phenolic hydroxyl group gets deprotonated (through excited-state proton transfer) and the proton gets transferred to glutamic acid 222 via a water molecule and serine 205. This mechanism has been suggested earlier [42,43] and was confirmed in **Paper IV**.

1.4 Voltage-gated sodium channels

Voltage-gated ion channels is a superfamily of membrane proteins, which are further classified by their ion selectivity (e.g. sodium, potassium and calcium channels). The different classes are closely related and are responsible for the generation of electrical signals and propagation of action potentials in the neurons and other excitable cells. They have three distinct functional states, which were described by Huxley and Hodgkin in 1952 [44] as activated, inactivated and resting states.

These channels were first discovered in eukaryotes, but it was not until the discovery of a prokaryotic voltage-gated sodium channel (Na_vs) that structural determination became feasible as it enabled higher protein production yields [45,46]. Bacterial and human Na_vs share a 20-25% identity and almost identical hydrophobic profiles and predicted topologies. Due to this, bacterial orthologues are expected to be a good model system for human Na_vs .

Although the structure has been solved for a bacterial Na_v in the open conformation [47], the actual mechanism for ion transport and selectivity has not yet been completely understood. Molecular dynamics simulations have yielded a wide range of proposed mechanisms for ion transport, as can be seen in the case of potassium channels [48,49].

A way to elucidate the mechanism would be to use advanced labeling techniques of residues residing in the pore of the channel, with the help of *in vitro* (or cell-free) protein synthesis. However, *in vitro* production of prokaryotic voltage-gated sodium channels had

not previously been reported. In **Paper V** we show that we can produce milligram quantities of a functional pore-only sodium channel from *Silicibacter pomeroyi* with the help of *in vitro* protein synthesis, which opens up for advanced labeling of Na_vs.

1.5 Scope of this thesis

In this thesis uniform, amino acid-selective and site-selective isotope labeling, together with NMR and IR spectroscopic techniques have been used to study dynamics of proteins, with a focus on the *Deinococcus radiodurans* phytochrome. The thesis is structured as follows:

Chapter 2 will describe the different expression and purification systems used throughout the work listed below.

Chapter 3 will focus on the various isotope labeling techniques used, and how they can be applied.

Chapter 4 will give a brief introduction to the basics of the different NMR experiments, and deal with the selected vibrational spectroscopy techniques I have used in this work.

Chapter 5 will summarize, conclude and give an outlook for future prospects

Paper 1 investigates the role of the loop region called tongue in a bacterial phytochrome. This was accomplished with advanced isotope labeling techniques, including uniform and *in vivo* amino acid-selective labeling. The structure and dynamics of the isotope labeled phytochrome were probed with NMR spectroscopy.

Paper 2 describes a novel signaling pathway in phytochromes, connecting the biliverdin chromophore with the conserved knot and further the helical spine and output domain. Uniform isotope labeling was performed on the bacterial phytochrome, and NMR spectroscopy was used to probe the structure and dynamics.

Paper 3 shows how isomerization of the chromophore is linked to the global structural rearrangement in the bacteriophytochrome sensory receptor. Uniformly and *in vitro* amino acid-selectively labeled phytochromes were subjected to time-resolved FTIR spectroscopy. This revealed how the hydrogen-bonding network in the chromophore binding pocket changes throughout the photocycle.

Paper 4 describes an efficient way to site-selectively isotope label proteins with the help of cell-free protein synthesis. The method was used to site-selectively label tyrosine 66 in GFP which was probed with femtosecond IR spectroscopy. We could confirm an intermediate state between photoexcitation and deprotonation.

Paper 5 details how to synthesize a fully functional membrane spanning pore-only sodium channel with cell-free protein synthesis. The method opens up the possibility to site-selectively monitor the active site with isotope labeling techniques.

Chapter 2

Protein expression and purification

Production of proteins is the basis of all biochemical and biophysical experiments. Different biophysical applications require different amounts of protein, from single particles up to several grams for the really sample-intensive experiments. Many proteins have low expression levels in their native cells. Finding ways to overexpress these proteins is a prerequisite for efficient biochemical studies.

Furthermore, some experimental methods require specific modifications to be made in the proteins, such as labeling with isotopes or unnatural amino acids. Developing and optimizing production protocols to meet the specific needs is therefore crucial, and the production of proteins is far from a trivial task.

In addition to protein production, protein purification is in most cases needed and the purification steps can also pose major challenges. Also here much effort is put into developing and optimizing protocols tailored for a specific application.

In Section 2.1 and 2.2 I will detail how these challenges were overcome. Both the *in vivo* and cell-free protein expression systems developed and used in this work will be explained, and the benefits for using one over the other for our NMR and IR spectroscopic studies. Finally in Section 2.3 I will discuss the purification steps taken and optimized for the proteins in my work.

2.1 *In vivo* recombinant protein expression

There are several overexpression systems in daily use at research laboratories. Both eukaryotic and prokaryotic systems as their respective proteins may differ in post-translational modifications or require different chaperone proteins in order to reach their native fold.

2.1.1 Recombinant protein expression in *E. coli*

The most widely used expression system is based on *Escherichia coli*, and the reasons for that are many: *E. coli* grow extremely fast, and can under optimal conditions double in cell density every 20 minutes [50]. Theoretically, *E. coli* can reach cell densities as high as 200 grams per liter of media, although in practice only a small percentage of that number is reached [50, 51]. Growth media for *E. coli* can easily be prepared and at a very low cost. Transformation of exogenous DNA vectors into *E. coli* can be performed in a matter of minutes or hours [52, 53].

DNA vectors, or plasmids as they are also called, are the circular DNA molecules in which to insert the gene of interest. A large selection of different vectors have been developed depending on which type of expression system used, prokaryotic or eukaryotic, *in vivo* or cell-free. What type of gene is inserted, large or small, and for what purpose the vector will be used, cloning or expression. A commonly used plasmid family for overexpression of proteins is the pET vectors, derived from the pBR322 vector developed already in 1977 [54].

In the pET vectors, gene expression is under control of the T7 promoter region from T7 bacteriophage, a site recognized by the highly specific orthogonal T7 RNA polymerase. The T7 promoter sequence is rarely encountered outside the T7 bacteriophage genome, and as such the overexpression is highly specific for the gene downstream of the T7 promoter sequence. This means we can reliably overexpress the target protein while keeping endogenous protein expression to a minimum.

The T7 RNA polymerase is controlled by the *lac* operon, it is triggered by the addition of Isopropyl β -D-1-thiogalactopyranoside

2.1. *In vivo* recombinant protein expression

(IPTG) which binds to the *lac* repressor and thereby frees the DNA for transcription and following translation. It is, however, not part of the pET vectors and needs to come from another source (see next paragraph). In addition, the pET vectors also contain resistance to ampicillin or kanamycin for efficient selection of cells containing the vector.

For *in vivo* expression, a pET vector was the one selected in most of the work in this thesis. In **Paper IV** an expression vector optimized for cell-free protein synthesis (pIVEX2.4d) was chosen, for simplicity and as reference for the cell-free synthesized proteins [55]. For **Paper V** the protein was overexpressed in the plasmid which was gifted (pHM3C-LIC), and had previously been shown to successfully work as expression vector for Na_vs [56].

Transformation of the vector into host cells has been a standard practice in biochemical work for decades, and it is therefore not surprising that the number of different cell lines for recombinant protein expression is huge. For *E. coli* a good starting candidate and also the most widely used would be the BL21(DE3) strain, a strain which can be traced back through scientific articles at least 100 years [57]. During those years, it has been subjected to numerous modifications to improve recombinant protein expression, among others certain genes coding for proteases which target foreign and extracellular proteins have been removed. BL21(DE3) also contains the gene for T7 RNA polymerase in its chromosomal DNA, and is essential for overexpression of proteins controlled by the T7 promoter region [52].

Note that even within the BL21(DE3) cell line there are strain variations. Rosetta(DE3) is for instance optimized for prokaryotic expression of eukaryotic proteins, and BL21 Star (DE3) increases the stability of mRNA (used in **Paper IV**). It is therefore of utmost importance to choose the expression host with care, as the protein yields can vary greatly. As the work in this thesis is concerned with overexpression of prokaryotic proteins, variations of the BL21(DE3) cell line were used.

Finally, after choosing the correct vector and expression host, additional factors such as growth medium, temperature and expression time will have an effect on the cell growth and in extension protein yields. In case of overexpression without further modifications Lysogeny Broth (or LB medium) is the most commonly used growth medium. It has been in use since the 1950's when Bertani first published the recipe [58].

As was the case in **Paper I, II and III**, where modifications to the protein during expression are needed, another medium such as the M9 minimal medium might be used. These modifications could be various forms of isotope labeling (**Paper I, II and III**), further detailed in Chapter 3.

The M9 minimal medium is based on a phosphate buffer containing sodium, potassium, calcium, magnesium and ammonium (which can be exchanged for isotope labeled ammonium for uniform labeling). Glucose is normally used as a carbon source, which also can be exchanged for various labeled versions for uniform labeling. Additional nutrients such as trace metals, vitamins and amino acids (labeled amino acids for amino acid-selective labeling) can be added if necessary for the specific application. Normal water can also be exchanged for D₂O, in order to deuterate the expressing protein. This medium gives a better control of the conditions, but usually results in slower growth and lower expression yields compared to *E. coli* grown in LB medium.

2.2 Cell-free protein expression

An alternative to using living cells as protein factories is to extract the components necessary for protein synthesis, such as the ribosome and various enzymes (see Figure 2.1), from the cells. The so-called cell-free extract is then mixed in an appropriate vessel with your gene of interest. This technique is called cell-free protein expression (CFPE) or *in vitro* protein expression.

As will be described in following sections and was introduced in the previous section, the choice of expression system is something

2.2. Cell-free protein expression

that requires a lot of consideration. In the work for this thesis I have taken advantage of the strengths of both *in vivo* and cell-free expression systems.

2.2.1 History of cell-free expression

CFPE has been used in biochemical studies since the 1950's when it was discovered that protein expression continued in disrupted cells. It was for instance used by Zamecnik and colleagues when they discovered the tRNA [59] and elucidated the biochemistry behind protein synthesis (for instance they showed that protein synthesis takes place at the ribosomes). CFPE became widely used after Nirenberg and Matthaei demonstrated a way to add an exogenous template in the form of RNA [60], although their work was mainly focused on studying the genetic code. In 1973, Zubay further optimized the CFPE system by digesting endogenous nucleic acids in the preparation of the cell-free extract and thereby reducing the expression of other endogenous proteins [61]. Further developments of the system, such as the addition of the orthogonal T7 RNA polymerase and T7 promoter site, as described in Section 2.1, were performed during the 70's and 80's.

A major advance in cell-free protein expression came in 1988 by Spirin et al. when they developed a continuous expression system (called continuous-flow cell-free) where the substrates were continuously added and the products removed [62]. This completely revolutionized the CFPE field, as up until that point all expression had been performed in so called batch mode, where all components are added in the beginning. The batch mode had several drawbacks as the reaction ran out of substrates, and by-products which acted inhibitory accumulated which resulted in a synthesis time limit of maximum an hour. With this new system, synthesis could be maintained for many hours, which resulted in higher protein yields. The continuous-flow cell-free system works by having a small reaction chamber over which a continuous flow of substrates is kept, while the products are separated away from the reaction chamber via a porous membrane [62].

This continuous flow of supplements can be a disadvantage. In the case of expensive isotope labeled substrates for labeling of proteins, a continuous flow is simply not cost effective. Here, batch mode has a distinct advantage.

Another adaptation of the continuous system is called continuous-exchange cell-free, where the major difference is the reaction taking place in a dialysis chamber and hence the polypeptide product remains in the reaction chamber throughout the reaction, while smaller by-products get replaced with fresh substrates through the dialysis membrane [62, 63]. All three expression systems (batch, continuous-flow and continuous-exchange cell-free) are still being used with smaller modifications.

For the work done in **Paper III, IV and V**, we have chosen the batch mode approach. For **Paper III and IV** the cost of externally supplied components (mainly the isotope labeled amino acids or aminoacyl-tRNAs) would have been unfeasible with the other approaches. For simplicity, the same setup was then used also in **Paper V**.

2.2.2 The cell-free system and its components

The cell-free system is an *in vitro* system, meaning the transcription and translation reactions takes place outside of living cells. The different components needed for the system to work are externally supplied, and can vary with the application. In **Paper V**, detergents were added to the reaction, as a membrane protein was expressed.

The system is most usually built up from *E. coli*, wheat germ or rabbit reticulocytes. Systems from insect cells, yeast and tobacco are other examples being used and technically systems from any kind of organism should work as long as they provide the necessary translational machineries [64, 65]. The various systems have different strengths and weaknesses. For example the *E. coli* system, used in this work, is by far the most robust but has huge problems with post-translational modifications. The choice of system needs to be based on the protein of interest and applications.

Cell-free extracts consist of the major components from a cell lysate. The major components needed for protein expression are ribosomes

2.2. Cell-free protein expression

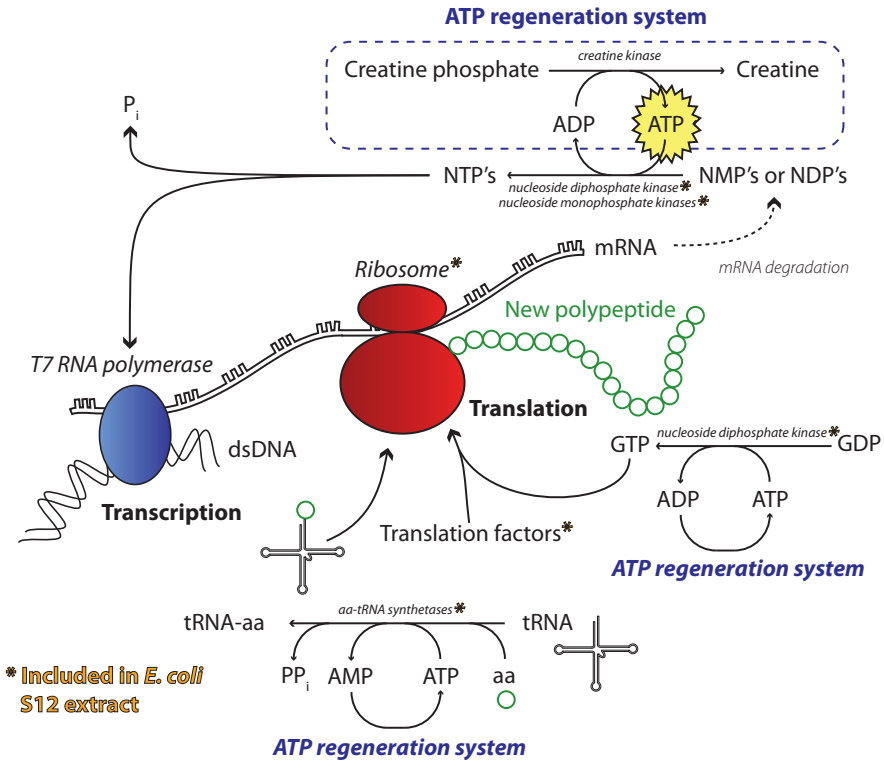


Figure 2.1: The CFPE system A schematic view of the CFPE system developed by Pedersen et al. Printed with permission from Dr. Pedersen.

for the actual protein synthesis, translation factors (initiation, elongation and release factors), aminoacyl-tRNA synthetases to load the tRNA with amino acids and various kinases to convert the NMP's and NDP's to NTP's. A schematic view of the components in a CFPE system can be seen in Figure 2.1.

Endogenous nucleic acids are degraded during an incubation step and removed either during centrifugation or dialysis in the extract preparation. There are variations in the preparation of this cell-free extract and the most common one is based on an extract called S30 extract created by Nirenberg in 1963 with small modifications [61,66, 67].

The S30 extract is a so called cleared extract, meaning lipids and membrane components have been removed during an ultracentrifugation step (at 30000g and hence the name S30). However, the preparation of this lysate is both tedious and costly (about 30% of the total cost of CFPE with S30 extract comes from the extract preparation) and in 2006 Kim et al developed a simplified extract they called S12 [68]. They noted that the *E. coli* lysate worked in its non-treated form.

Based on this, they added a few simple preparation steps such as a low-speed centrifugation, short incubation and dialysis, which resulted in a 60-80% reduction of preparation cost and significantly higher expression yields [68]. There is the drawback of slightly higher background expression of endogenous proteins, and it does not seem to work well for all cell lines. It also still contains lipids and other membrane components, and is thus a cloudy extract, which could influence the expression of especially membrane proteins.

In addition to the cell-free extract, other components are needed for the CFPE to work: template DNA containing the gene of interest, RNA polymerase, amino acids, tRNA and an energy source. As template DNA purified plasmid DNA or PCR products can be used. Amino acids and tRNA can be easily purchased, and unnatural or isotope labeled amino acids can be added (as was done in **Paper I, III and IV**).

Regarding the RNA polymerase, the T7 phage system can be used with great results (as described in Section 2.1). When it comes to energy source a lot of work has been put into developing an efficient energy regeneration system. The cell uses high-energy phosphate compounds to generate ATP (adenosine triphosphate).

As certain by-products such as inorganic phosphates act inhibitory on the expression, these need to be removed. In continuous-flow and continuous-exchange cell-free this is taken care of by continuous exchange, but for batch mode the accumulation has long been a problem. Several different regeneration systems have been developed. For instance the Cytomim tries to mimic the cytoplasmic conditions of the host cells and utilizes pyruvate as energy source and thus does not generate inorganic phosphates as a by-product [69].

2.2. Cell-free protein expression

Other systems utilize phosphoenolpyruvate, acetyl phosphate or creatine phosphate together with pyruvate kinase, acetate kinase or creatine kinase, respectively, to generate ATP from ADP.

Additional additives such as various cofactors, detergents (**Paper V**), chaperone proteins and other interaction partners can also be used with the *E. coli* CFPE system, depending on the need of the protein of interest. A schematic view of a CFPE system is shown in Figure 2.1.

Optimizations and modifications of the cell-free systems is still ongoing, and there are now many specialized strains for CFPE, some created with the help of the new revolutionary CRISPR/CAS9 system [70–74]. The expression yields have reached up to around two milligrams of protein per milliliter of reaction, and further developments will make CFPE more and more feasible for large-scale applications [73].

Recently, commercial CFPE-kits have also been developed as educational tools for molecular biology and its central dogma utilizing a freeze-dried version of the cell-free extract. Besides showing that freeze-dried extracts, which are later rehydrated, are still active, this will be of huge help in the coming years teaching a new generation of scientists [75,76].

A CFPE system was developed by Pedersen et al in 2011 [77], and is schematically shown in Figure 2.1. This system is an *E. coli* based system from the BL21(DE3) Rosetta strain with a modified S12 extract [68,77]. It is a batch system specifically optimized for expression of proteins labeled for NMR studies. Further, the concentration of creatine phosphate and key amino acids were optimized to improve protein expression yields. Additional additives increasing the yields include 2-oxoglutaric acid, succinic acid and malic acid. As a batch system the effective reaction time is limited but the advantage being cost effective isotope labeling, easy miniaturization, parallelization and hence fast screening of optimal reaction conditions [78].

As was mentioned previously and will be described in Chapter 3 it was used for a number of isotope labeling techniques (**Paper III and Paper IV**) and for these applications batch mode is more cost-effective. Some advantages and disadvantages of using *in vivo* or

cell-free expression, as well as our home-made CFPE system versus commercial ones will be discussed in the next subsection.

2.2.3 Cell-free versus *in vivo* expression systems

When choosing which expression system to use, there are a lot of considerations one has to take into account. As such, a comparison between cell-free and *in vivo* expression systems is not completely easy to do and is not meant to decide which system is better. They both have their strengths and weaknesses for separate applications. In the following subsection I will try to go through a few of the differences, and give a number of examples where one might choose one expression system over the other.

Depending on what the final goal is, either a CFPE or *in vivo* system might be chosen. The main advantage of CFPE is the open expression system, which for example allows for advanced labeling techniques (as was done in **Paper III and Paper IV** and will be further discussed in Chapter 3) or expression of proteins which might be toxic for the cells (or in concentrations toxic for the cell).

In addition, the time until ready protein sample is much shorter compared to *in vivo* methods, and an NMR spectrum has been reached within 90 minutes of CFPE reaction start [79]. This can be a huge advantage for screening purposes. More generally time-scales of 2-3 hours from start to purification step is the standard, and a schematic comparison can be seen in Figure 2.2.

The possibility to add detergents, lipids, nanodiscs or chaperone proteins have also opened up the possibility of expressing membrane proteins (a fact I took advantage of in **Paper V** when I expressed a pore-only sodium channel with CFPE), which *in vivo* can cause cytotoxicity or be non-functional or even aggregate due to misfolding [80,81].

The ease of high throughput screening with CFPE is also valuable to find optimum expression conditions. In addition the environment of the reaction is open for changes such as in pH, redox potential and temperature, as long as it does not hinder the transcription and translation reactions [64,77].

2.2. Cell-free protein expression

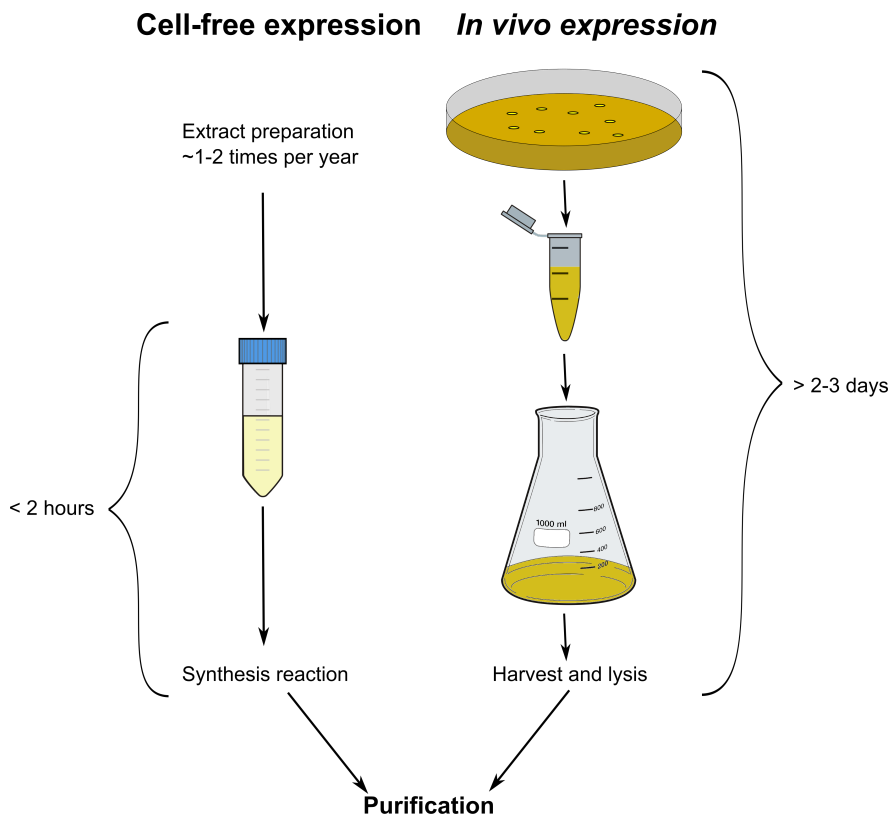


Figure 2.2: Comparison of CFPE and *in vivo* reaction times A schematic comparison of CFPE versus *in vivo* expression systems. The time for extract preparation is around 2-3 days, and one 20 liter batch is enough for roughly 1 liter of CFPE reaction. When the extract has been prepared, the synthesis reaction takes less than two hours, in comparison to *in vivo* expression which normally lasts over a period of days.

A few drawbacks with the CFPE is the problems with expressing large or highly complex proteins, which comes from either incomplete translation or insufficient post-translational modifications (in prokaryotic systems) or low yields (which is the general norm for eukaryotic systems) [65]. These problems are consistently being addressed, and major improvements have been made over the last few

years [65,82–84].

In general, there is also the problem of choosing between continuous-flow and continuous-exchange cell-free or batch mode. One will either have long reaction times and higher yields but laborious setup or simple setup but shorter reaction times and lower yields, respectively.

There are commercial systems available, but the cost of commercially available CFPE systems runs into the hundreds of dollars per milliliter of reaction. This is also one of the drawbacks, but something that can be rectified by setting up and maintaining a home-grown cell-free system as previously described [77]. By performing large-scale cultures and preparations of the cell-free extract (20 liter culture is enough for roughly 1 liter of CFPE reaction), the time spent on this most laborious part (2-3 days) of using a home-made system can be significantly reduced.

2.3 Protein purification

When expression system has been chosen and developed, the next consideration will be regarding the purification of the protein. First question would be if the protein at all needs to be purified? Some applications can be performed on the crude lysate or even in whole cells, while others require a completely pure protein sample.

For the applications in this thesis, some purification was necessary, and the steps taken will be described in this section. An example of unexpected difficulties one can encounter will be described in the end.

Generally, protein purification starts with disrupting the cells, for which there are several approaches. During this work, the cells have been lysed by abruptly changing the external pressure while passing them through a needle valve which causes the cells to disrupt. This was achieved with the help of either a French pressure cell (named after the American Charles Stacy French), or an Emulsiflex (developed by Avestin) which utilizes the same principle. Cell debris is removed by centrifugation.

2.3. Protein purification

After centrifugation the lysate is normally further purified with a combination of chromatographic techniques. These normally include affinity, ion exchange (not used in this work), and/or size exclusion chromatography.

In the case of *DrBphP*, biliverdin is added in excess under mildly reducing conditions to the centrifuged cell lysate before chromatographic purification, which spontaneously incorporates and covalently binds the protein matrix.

For the use of affinity chromatography, a polyhistidine-tag has been inserted either N- or C-terminally of the protein, and in this work both versions have been utilized.

Affinity chromatography is usually the first purification step as it can separate an affinity-tagged protein from a very crude mixture of proteins. Normally a second chromatographic method is used to yield a highly pure protein sample, and for this work SEC (size exclusion chromatography) was the method of choice.

There are a few differences in the purification of CFPE and *in vivo* expressed proteins, but in general the protocol do not change much. As CFPE proteins are produced in an open system, there is no need for an initial cell lysis step. There is however still major cell components from the extract present, which need to be removed by centrifugation. The purification then progress in the same fashion as for *in vivo* expressed proteins, with affinity chromatography and following SEC. Protein purity is then evaluated with gel electrophoresis.

A complication in regards to the CFPE of *DrBphP* was discovered. The incorporation of biliverdin caused some difficulties as the CFPE milieu was too strongly reducing. To solve this, the protein was purified as an apoprotein following the steps described in the previous paragraph and a reducing agent was added in small molar excess to the purified protein after which the biliverdin was added also in excess. Excess biliverdin and reducing agent was later washed away during a final buffer exchange step.

Chapter 3

Isotope labeling

Isotope labeling has been key to the work presented here. I have used different labeling techniques in order to follow protein dynamics in a site-specific way. With this we have been able to reveal new signaling pathways and how changes in structural dynamics are used to modulate signaling in phytochromes (**Paper I and II**). We have also shown how isomerization of the chromophore is linked to the global structural rearrangement in the same phytochrome (**Paper III**). And with the help of site-specific isotope labeling of GFP we could confirm the existence of an intermediate state between photoexcitation and proton transfer (**Paper IV**).

It is safe to say isotope labeling has been of crucial importance to my work. This chapter aims at explaining the various isotope labeling techniques used throughout this thesis. The first section will detail the uniform labeling of *in vivo* expressed proteins. Section 3.2 will describe amino acid-selective labeling, and the final Section 3.3, site-selective isotope labeling.

3.1 Uniform isotope labeling *in vivo*

3.1.1 Applications with uniformly isotope labeled proteins

For several applications, two of them being used in this work (NMR and vibrational spectroscopy), there is either a need or advantage to uniformly isotope label proteins. For NMR spectroscopy (detailed in Chapter 4), the sample needs to contain magnetically active nuclei

in order to give detectable signals. In vibrational spectroscopy (also described in Chapter 4), the isotopes will cause a shift of the protein-associated peaks, making it possible to distinguish them from other compounds which can give rise to overlapping peaks. In both cases, isotope labeling will make it possible to directly follow protein structural dynamics in solution.

For the use of NMR spectroscopy, the nuclei need to be magnetically active (further explained in Chapter 4). The most common magnetically active nuclei for protein NMR purposes are ^1H , ^{13}C , ^{15}N , ^{31}P , of which only ^1H and ^{31}P have a naturally high abundance. For ^{13}C (1.1% natural abundance) and ^{15}N (0.4% natural abundance) other means of enrichment are needed [85].

In small particles, where the concentration can be sufficiently high, this is usually not a problem as it then is possible to measure on the natural abundance. For large molecules with low concentrations, such as proteins, this is not feasible, and the magnetically active nuclei need to be enriched in some fashion.

Uniform labeling for vibrational, or infrared (IR), spectroscopy serves a slightly different purpose from labeling for protein NMR spectroscopy. The inclusion of isotopes shifts the chemical bond vibrations to other frequencies, making it possible to identify the protein vibrational modes (discussed in Chapter 4).

Protein IR spectroscopy mostly suffers from spectral congestion and difficulties in assigning the peaks to specific amino acids [86]. At a first glance, the uniform isotope labeling, which would shift all peaks, would look fairly useless as a tool to help with this problem.

In **Paper III**, we show that the combination of uniformly labeled apoprotein coupled with non-labeled chromophore gives an insight into the molecular mechanism of signal transduction in *DrBphP*. This is achieved by separating shifting protein peaks from non-shifting peaks from the biliverdin chromophore. Another way of describing this type of uniform labeling of the protein would be selective **un**labeling of the chromophore.

Uniform labeling has also been used in investigating protein-protein and protein-ligand interactions, where one protein or ligand is left

3.1. Uniform isotope labeling *in vivo*

unlabeled and therefore certain regions, such as the so called amide I bands, do not overlap between the two interaction partners [87–89].

Although uniform labeling with isotopes is used in vibrational spectroscopy, other isotope labeling techniques outlined in Section 3.2 and 3.3 are considered more powerful as they give more site-specific information [86].

3.1.2 Principles of uniform isotope labeling

In order to isotopically label proteins, one has to systematically consider all alternatives listed in the previous chapter (Chapter 2).

Isotopically labeled precursors, such as [U- ^{13}C]-glucose and $^{15}\text{NH}_4\text{Cl}$, can be purchased from several vendors and can be used in substitution for the regular carbon and nitrogen sources. However, in order to ensure a protein fully labeled with ^{13}C and ^{15}N , the regular source also has to be removed and normal LB medium can therefore not be used when expressing these proteins. Instead, various minimal media such as the M9 minimal medium, which was described in Section 2.1, are utilized.

As large amounts of sample might be required for structural and dynamical studies, the use of isotopically labeled media could become very costly and protocols have therefore been developed where the cells are first grown in unlabeled rich medium to high cell density to later be transferred to a labeled medium for the expression phase. This modification requires a few extra steps during cell culturing, but reduces the costs associated with isotope labeling dramatically [90].

Full deuteration of the protein should also be considered for certain applications. Deuterating the protein could help increasing resolution in protein NMR spectroscopy. Deuteration of the protein means the hydrogens in the protein are not visible in the NMR spectrum as ^2H has a very different resonant frequency. In complex spectrum with many peaks this is helpful as the amount of ^1H contributes significantly to peak broadening in larger proteins (e.g. >20kDa), which congests the spectrum (further explained in Chapter 4) [91, 92].

There is, however, still a need for the amide hydrogen to be ^1H as the sensitivity for proton experiments is much higher than for deuterium. Amide hydrogens are easily back-exchanged if sufficiently exposed to the solvent. The most common approach is to grow the cells and express the protein in a fully deuterated environment.

The host cells are therefore acclimatized to grow in D_2O over several cycles with increasing D_2O concentration. In addition, some of the precursors also need to be deuterated (for example [$\text{U-}^2\text{H}, ^{13}\text{C}$ -glucose]). Normally the cell growth is slower in D_2O -based medium, but the protein yields are typically only slightly affected [92].

The protein is then expressed and purified as previously described in normal aqueous buffers and the back-exchange occurs spontaneously for all solvent-exposed amides. For highly stable proteins with low numbers of solvent-exposed amides, it could be beneficial to investigate a potential de- and renaturation protocol of the protein to fully back-exchange all amide- ^2H for ^1H [92].

Deuteration further helps to increase the signal-to-noise ratio and decrease signal overlap in the triple resonance experiments most commonly used for backbone assignment (more on this in Chapter 4) [92]. These types of uniform labeling are a necessity for the backbone assignment of large proteins, as was done in **Papers I and II**.

3.2 Amino acid-selective isotope labeling

3.2.1 Applications with amino acid-selectively labeled proteins

Amino acid-selective labeling is a method used to gain more selective information. This could be certain amino acids involved in active sites, or to follow the dynamics of specific protein parts in a time-resolved manner.

Amino acid-specific labeling encompasses both labeling with isotope variants of standard amino acids and unnatural amino acids, and both techniques have been used for decades [10, 93, 94]. The incorporation of unnatural amino acids has the drawback of changing

3.2. Amino acid-selective isotope labeling

the chemical nature of the side chain, which could alter the nature of the mechanism being investigated.

There are numerous examples of how the amino acid-selective labeling technique have been used to gain a better understanding of biochemical mechanisms. In **Paper I and III** we show two different applications where amino acid-selective isotope labeling have been utilized.

In **Paper I**, I selectively ^{13}C , ^{15}N -label all lysines in two *DrBphP* monomer samples containing either all lysines or with one lysine mutated to alanine. With this we were able to unambiguously assign and show changes in dynamics in the *DrBphP* tongue between Pr and Pfr state.

In **Paper III**, we ^{13}C -labeled all aspartic acids in an attempt to elucidate how the isomerization of biliverdin couples to the global structural rearrangement in *DrBphP*.

These two papers also detail two different ways to incorporate isotope-labeled amino acids, and will also be explained in the following subsections.

Selective isotope labeling could also be used for systematic backbone assignment in NMR, by creating 19 different samples (as proline does not contain an amide proton) each containing a different labeled amino acid. This is a labor-intensive and possibly expensive way to decrowd the NMR spectrum [95].

There is also a number of combinatorial labeling methods, which can be used to efficiently decrowd the spectrum [96–98].

Amino acid-selective labeling can in favorable cases also give site-specific information. If the site of interest contain a unique sequence which can be labeled, as was done by Professor Yokoyama and colleagues [99]. This labeling technique was explored within the work of this thesis, but was put aside in favor of other techniques.

3.2.2 *In vivo* selective labeling

In vivo isotope labeling takes advantage of *E. coli*'s highly efficient ability to take up exogenous compounds, in combination with growth

in minimal medium. The procedure to express *in vivo* labeled proteins follows the same principles as detailed in Chapter 2.

The minimal medium is supplied with all amino acids, exchanging unlabeled with labeled equivalent (which are commercially available) at ones convenience. The addition of unlabeled amino acids is important for the purpose of reducing isotopic scrambling. Isotopic scrambling is an effect from the bacterias efficient metabolic pathways. It can utilize amino acids supplied in excess to produce other amino acids in shortage and the isotope label can therefore end up in sites not intended. The addition of excess unlabeled amino acids tends to inhibit the metabolic pathways involved to some extent [95,100].

In order to completely avoid scrambling, genetic manipulation are required to provide strains auxotrophic for the enzymes involved in the amino acid biosynthesis, and the need for special strains to label certain difficult target amino acids is one of the drawbacks with the *in vivo* labeling technique [101, 102]. However, once a suitable host strain has been found or developed, *in vivo* labeling does have the advantage over CFPE in the production of large and complex proteins, and normally higher expression yields [103].

There are also *in vivo* labeling techniques taking advantage of the metabolic pathways by introducing selectively labeled precursors to label for instance methyl groups of methyl group-containing amino acids. Or to label hydrophobic residues, employing the fact that these amino acids are at the end point of their respective biochemical pathways. This is very useful in NMR structure and dynamics studies [93,104,105].

3.2.3 Selective labeling with CFPE

The problem with isotope scrambling is not as pronounced with CFPE due to lower metabolic activity, and the system can be highly optimized to produce amino acid-selectively labeled protein [106]. The general CFPE method is detailed in Chapter 2.

One labeling technique taking advantage of the much lower degree of scrambling from CFPE is Stereo-Array Isotope Labeling which

3.3. Site-selective isotope labeling

stereospecifically isotope label proteins for structure determination [107].

The enzymes involved in the metabolism of the labeled amino acids are present in the cell-free extract, but most of them do not get regenerated and therefore the problem is not as severe. The amino acids most subjected to scrambling in the CFPE are asparagine, aspartic acid, glutamine, glutamic acid, serine and alanine. To deactivate the metabolic enzymes it is possible to heat-treat the extract or supplementing the reaction with inhibitors targeting these enzymes [101]. Albeit with a slight loss in protein yields [108]. This was done in **Paper III** where aminooxyacetic acid was added to avoid the degradation of aspartic acid to glutamic acid.

The incorporation level of amino acids is generally low and with the addition of isotopically labeled amino acids this tend to dramatically increase the cost. Both for batch but especially for continuous-flow and continuous-exchange cell-free systems where a continuous supply of supplements is required [80, 106].

This can be mitigated by adjusting the concentrations of the labeled amino acids in relation to their occurrence in the primary sequence without affecting the labeled protein yields [79, 109]. Such stoichiometric amino acid mixtures can be used to further reduce the costs of labeling [77, 110].

3.3 Site-selective isotope labeling

Site-selective labeling of a single amino acid in the protein gives direct insight into highly specific areas of the protein structure and its dynamics. Not surprisingly, site-selective labeling has received a lot of attention over the years.

The most straight-forward method for site-selective labeling is peptide synthesis, and therefore the most widely used. The efficiency of the synthesis however makes production of higher yields of large proteins unfeasible. Even with 98% ligation efficiency the yield would drop to 13% for a 100 amino acid long protein (or 0.004% for a 500 amino acid protein like *DrBphP* PAS-GAF-PHY).

For some larger proteins it could be possible to ligate a synthesized peptide to a truncated *in vivo* produced version of the protein using native chemical ligation [111, 112]. This reaction was discovered in 1953 and developed for peptide ligation in 1994 [113, 114].

Another way to label proteins site-specifically is to introduce unnatural amino acids, using a chemically or enzymatically aminoacylated suppressor tRNA, which recognizes the amber stop codon (UAG), and CFPE. However, this is not feasible if the aim is to study amino acids in the active site of the protein as they are likely to alter the chemical nature of the amino acid in question [115, 116].

There is also a problem with the competing RF1 (Release Factor 1), which recognizes the amber stop codon and terminates translation, which dramatically lowers the yields of full-length proteins. This was addressed by Otting and colleagues in 2012, when they replaced the wild-type RF1 with a Chitin Binding Domain-tagged RF1, allowing for affinity chromatographic removal of RF1 from the cell-free extract [117].

3.3.1 Principles of site-selective labeling

CFPE is the most widely used technique for site-selective labeling, as it has the advantages of being an open system and thus easy to modify. As site-selective labeling requires the loading of the suppressor tRNA with an aminoacyl group by chemical ligation, the open system is a huge benefit.

There is a number of strategies regarding this: either utilizing a modified aminoacyl tRNA synthetase (aaRS) which recognizes the specified amino acid, or with the T4 RNA ligase 1 reaction, as was first reported by Hecht et al. in 1978 [118], and is further explained below.

A truncated suppressor tRNA is transcribed in a so-called "run-off transcription" where transcription is performed with an enzymatically (Fok1) digested and thus linearized DNA plasmid resulting in the tRNA lacking the last two nucleotides (-CA) [119]. Instead a dinucleotide (pdCpA) fragment is chemically aminoacylated with the

3.3. Site-selective isotope labeling

amino acid of choice. An efficient way to aminoacylate the pdCpA fragment has previously been reported [120].

Amino acid amides and functional side chain groups are accessible for side reactions and must be chemically protected. For amino acids without side chain functionality, the 4-pentenoyl group, which is labile in aqueous iodine can be used. While for amino acids with side chain functionality, 6-nitroveratryloxycarbonyl and 2-nitrobenzyl, which are both susceptible to photolysis, can be used to protect the amide groups and side chains, respectively [120, 121]. Various other protection groups have been explored but are not mentioned here.

The aa-pdCpA is then ligated to the truncated tRNA(-CA) with the help of T4 RNA ligase 1, deprotected with either iodine solution or light exposure and flash frozen. After deprotection of the amide group, the aminoacyl-tRNA ester linkage is susceptible to hydrolysis and start degrading to full-length tRNA and free amino acid. This is also one of the reasons to include protection groups from the start, as the hydrolysis reaction is much faster than the ligation reaction [122].

This method of introducing site-selective labels can be used with both unnatural and isotope-labeled amino acids. It does not require the aaRS, which has highly specific recognition sites for both the specific amino acid and its respective tRNA and would require structural modifications to allow for unnatural amino acids. The T4 RNA ligase method is therefore a general method, as compared to the aaRS. [120].

The above described method was used for **Paper IV** to site-specifically isotope label tyrosine 66 in GFP. A schematic view of the site-selective labeling used in **Paper IV** can be seen in Figure 3.1.

In vivo site-selective labeling has been quite heavily studied and developed over the years. A lot of work has gone into developing aaRS to accommodate a huge variety of unnatural amino acids and ligating these to amber suppressor tRNA or other modified tRNAs [117, 123].

A big advantage of the *in vivo* site-selective labeling is that it realizes the possibility to site-specifically monitor a protein in living cells. Unfortunately the problem of a competing RF1 still remains and can drastically reduce the yields of full-length protein. There has

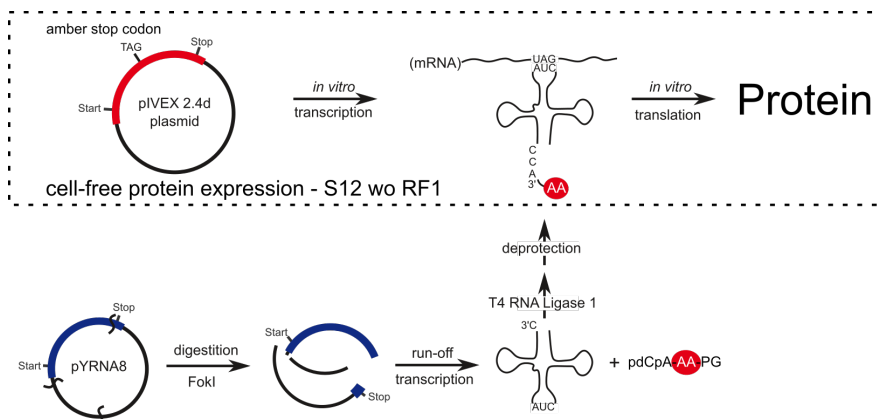


Figure 3.1: Site-selective labeling of proteins with CFPE A schematic view of the experimental principle applied to produce site-selectively labeled protein. Abbreviations: AA (amino acid), RF1 (release factor 1), PG (protection group), pdCpA (5'-O-phosphoryl-2'-deoxycytidylyl-(3'→5')adenosine). Figure adapted from **Paper IV**.

been attempts to remedy this, although to no greater success as the complete removal of RF1 is normally lethal to the host cell [124].

Further complications include the problems of incorporating isotope-labeled instead of unnatural amino acids, as the native aaRS cannot distinguish between the native and isotope-labeled amino acids. In very specific cases it has been possible to introduce a labeled photoreactive unnatural amino acid, which upon radiation and after incorporation decays an isotope-labeled natural amino acid [125–127]. Since unnatural amino acids tend to have altered chemical properties, the use in probing functional sites is limited.

3.4 Summary

A large number of techniques for isotope-labeling exist, many developed to solve a specific challenge. I have used a combination of uniform, amino acid-selective and site-selective labeling in this work.

3.4. Summary

By uniformly isotope labeling a *DrBphP* monomer, we have in **Paper I** been able to perform backbone assignment with NMR spectroscopy. This, together with amino acid-selective labeling, has revealed that the tongue arms form an α -helix and β -sheet in Pfr form, while being highly dynamic in Pr. This indicates that changes in structural dynamics of the tongue region are coupled to the photomodulation of phytochromes.

Uniform labeling has also made it possible to investigate the knot region in **Paper II**. We detect major changes between Pr and Pfr and suggests this as a novel signaling route in phytochromes.

In addition we uniformly isotope labeled the *DrBphP* apoprotein in **Paper III** to detail how the isomerization of the chromophore couples to global structural rearrangements in the photosensory receptor.

By using amino acid-selective isotope-labeling we could assign a band to the chromophore D-ring in Meta-R state by excluding a protonated aspartic acid 207 as the origin, as described in **Paper III**.

In **Paper IV**, we site-selectively isotope labeled tyrosine 66 in the chromophore of GFP, and were able to unambiguously confirm the existence of an intermediate state between photoexcitation and deprotonation of tyrosine 66, as had previously been proposed [42, 43].

Chapter 4

Spectroscopy

This chapter will give a short description on the different spectroscopic techniques which have been used during the work on this thesis. Section 4.1 will give a brief introduction to NMR spectroscopy and list the various NMR experiments used. Section 4.2 will be the equivalent for infrared spectroscopy, and Section 4.3 will summarize this chapter.

4.1 NMR spectroscopy of proteins

NMR spectroscopy was independently developed by two different groups lead by Dr. Purcell at Harvard University and Dr. Bloch at Stanford University, for which they shared the Nobel Prize in physics in 1952 [128, 129]. They in turn had based their research on the findings of Dr. Rabi who is credited with discovering Nuclear Magnetic Resonance in the 1930's, and who received the Nobel Prize in physics in 1944.

Over the following 10 years, NMR spectroscopy continued to develop and in 1957 the first NMR spectrum of a protein was recorded [130]. This was shortly before the first crystal structure of a protein was revealed, and the dream of using NMR in solving protein structures was born [4]. But it would take until 1985 before the first 3D-structure was solved by Wüthrich et al [131], who later would receive the Nobel Prize for his work. Ever since then, NMR has been a powerful tool to solve structures of proteins in solution and as of 2017,

over 12000 structures by NMR have been deposited in the protein data bank [2].

4.1.1 Short theory of protein NMR

NMR spectroscopy uses the intrinsic spin (or spin angular momentum) of a nucleus. This is a quantum mechanical property of certain elementary particles and as nuclei consist of these elementary particles they too can have spins. The spin of a nucleus can also have different states, which have the same energy in the absence of an external magnetic field.

Classically, the spin of a nucleus can be viewed as the nucleus spinning around its own axis. Not all nuclei are magnetically active though, and thus can not be detected by NMR. Spinning charged particles create a magnetic field around them, and as such have a magnetic moment, which is closely linked to their spin.

An NMR spectrometer is in essence an extremely strong magnet, equipped to send and receive radiofrequency (RF) waves, with magnetic fields tens of thousands times stronger than the Earth's magnetic field. The sample is immersed in this strong field, and the nuclei magnetic moments align with the magnetic field direction of the spectrometer.

This causes the energy levels of the nucleus spin states to shift, which creates an energy gap. The energy required to transition between these states is dependent on the magnetic field strength experienced by the nucleus. In a molecule this depends not only on the magnetic field of the spectrometer but also the local electron distribution in the molecule contribute. Therefore, two different hydrogens in a molecule can have slightly different RF resonant energies (or resonant frequencies).

This effect is called chemical shift, and is one of the reasons NMR have become so widely used in protein science. The chemical shift makes it possible to follow events in proteins at atomic resolution. The chemical shift is defined as the relation of the resonant frequency of the nucleus of interest (ν) to a reference nucleus (ν_{ref}) which varies

4.1. NMR spectroscopy of proteins

between the different measured nuclei, and is given as parts per million (ppm).

$$\delta = 10^6 * \frac{(v - v_{ref})}{v_{ref}}$$

Analysis of the carbon chemical shifts are often used to reveal the secondary structure. By subtracting a so-called random coil shift value, it is possible to distinguish between regions having a propensity for α -helical or β -strand structure [7, 132]. This analysis, called secondary chemical shift analysis, can be used to determine secondary structure, or changes in secondary structure between two protein states.

Further analysis from chemical shifts useful for protein NMR spectroscopy is chemical shift perturbation. The amides of proteins are very sensitive to changes in local magnetic environment, giving large chemical shifts. Chemical shift perturbation follows the changes in chemical shifts occurring from ligand binding or changes in structure between different states [133].

Spin-spin coupling is another great source of information in protein NMR, which arises from the different spin states of a nucleus and how those different spin states are "felt" by other nearby nuclei. This gives rise to splitting of peaks, which increases the amount of information but also makes a more complex spectrum which could present a difficulty in the already crowded spectra of protein molecules.

In protein NMR spectroscopy, many experiments are therefore performed in a decoupled manner. Decoupling the ^1H - ^{13}C -coupling, as an example, is performed by irradiating the sample with ^1H resonance frequency during the recording of the ^{13}C signal and thus removes the splitting of the ^{13}C signal.

The major contribution to the spin-spin coupling is called the J -coupling (or scalar coupling) and is for a specific coupling always the same. However, the information contained in the splitting of peaks can be very useful for structure determination as it also gives information about the dipolar couplings of two nuclei. The dipolar coupling of two nuclei depends on the orientation and magnitude of

their magnetic moments, and the vector which describes their relative positions in relation to the magnetic field.

As the molecules in an isotropic solution tumble, the dipolar coupling contribution gets averaged out and cannot be measured. This can be solved with the addition of alignment medium to the protein solution which partially aligns the molecules and creates an anisotropic non-zero signal that is measurable and called residual dipolar coupling (RDC). [134–137].

As RDCs give information on the relative vector of chemical bonds, they can be used to refine crystal or NMR structures, investigate protein dynamics or even solve structures *de novo* [136–138].

Spin relaxation is the most common way to measure dynamics in protein NMR spectroscopy. The energy difference between a nucleus spin states provides the possibility to "flip" a spin from one low energy state to a high energy state with the help of a RF pulse. As the system returns to thermal equilibrium, the energy from high energy spin states dissipates out into the surroundings (or "lattice"). The lattice can be lower energy spin states, vibrational and rotational motion. The relaxation speed is very much dependent on changes to the local environment and T_1 relaxation (also called spin-lattice relaxation) experiments measure these relaxation rates.

As the RF pulse is delivered and the spin states flipped, the spins are in phase with each other. Depending on the local magnetic environment, i.e. other local spin systems, the spins return to equilibrium with different decay times and therefore ends up out of phase. T_2 relaxation (or spin-spin relaxation) experiments measure these dephasing (or phase randomization) times.

This is a short summary of different NMR experiments used in this work. For a more thorough explanation of the theory behind NMR spectroscopy, see Levitt [139] and Keeler [140].

4.1.2 Multidimensional protein NMR spectroscopy

One of the most frequently used experiments in NMR spectroscopy on large protein is TROSY (transverse relaxation optimized spectroscopy).

4.1. NMR spectroscopy of proteins

It is a two-dimensional experiment with one axis presenting the proton frequency and the other axis either carbon or nitrogen frequency (See Figure 4.1 for an example of a [^1H , ^{15}N]-TROSY).

As can be imagined, the addition of another dimension helps tremendously with signal overlap. It has become widely used as a "fingerprint" of the protein to evaluate whether more advanced and time-consuming experiments are feasible.

2D-experiments take advantage of the fact that spin polarization (or magnetization) can be transferred from one spin to another coupled spin, either through bonds or space. A heteronuclear experiment couples protons to other types of nuclei such as ^{13}C or ^{15}N . [^1H , ^{15}N]-TROSY is particularly useful in protein NMR as all amino acids besides proline have one proton coupled to a nitrogen in the amide group. The number of peaks in relation to the size of the protein, and the peak distribution and line width of the peaks gives a clue as to the possibilities of recording higher dimensional spectra.

Resonance assignment of amino acids is necessary in order to extract residue-specific information from the NMR data. For large proteins, two dimensions is not enough to fully resolve all the different peaks, and some way to link sequential residues is needed.

Multidimensional triple resonance experiments require samples uniformly labeled with ^{13}C and ^{15}N , as they transfer the magnetization from the amide proton to the amide nitrogen as the first detection step. In the case of a HNCA experiment, the magnetization is further transferred to the α -carbon as the second detection step, and last back to the amide proton for the third and final dimension.

As can be seen in Figure 4.2, the magnetization transfers to both residue i and residue $i-1$ and gives both internal and sequential information. This combination is the basis of backbone assignment, as it gives the possibility to connect a sequence of residues.

Triple resonance experiments often used also include detection of β -carbon (HNCACB) or carbonyl-carbon (HNCO), and experiments specifically designed to detect sequential information (HN(CO)CA, HN(CO)CACB and HNCACO) [141,142]. These experiments are normally used in tandem or in pairs (HNCA and HN(CO)CA) for backbone assignment of proteins.

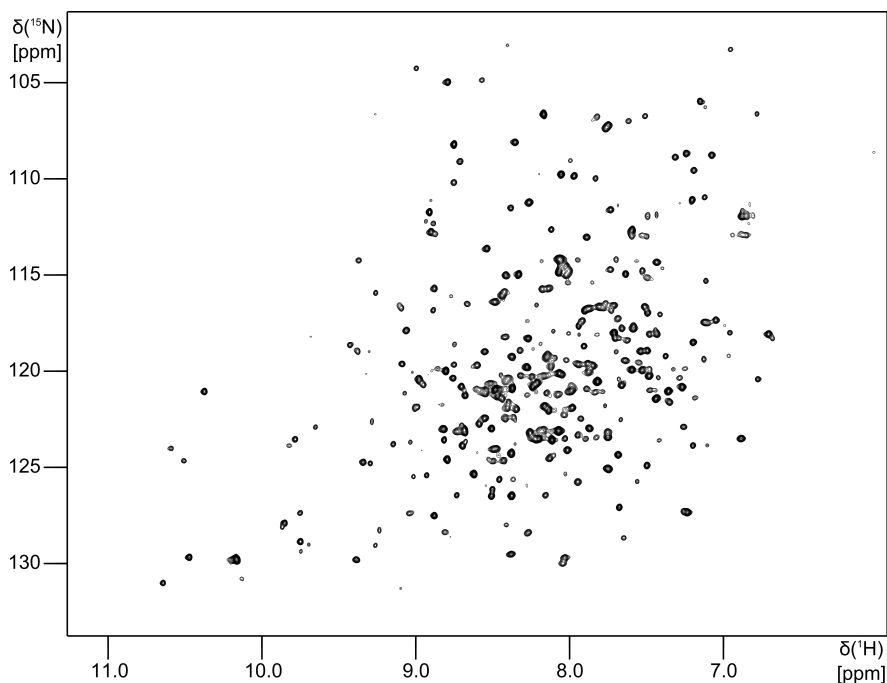


Figure 4.1: $[^1\text{H}, ^{15}\text{N}]$ -TROSY of *DrBphP* monomer in Pr state, each peak representing one amide group. Here the proton frequency is represented on the vertical axis and nitrogen frequency on the horizontal axis.

The backbone assignment can to some extent be done automatically with programs such as FLYA or AutoAssign [143,144]. For large proteins there is still a need to manually go through the assignment, linking each internal signal with the sequential signal.

For **Paper I and II** we recorded the following triple-resonance experiments: HNCA, HN(CO)CA, HNCACB, HN(CO)CB, HNCO, HNCACO and $[^1\text{H}, ^{15}\text{N}]$ -NOESY, for a 509 residue *DrBphP* PAS-GAF-PHY monomer in both Pr and Pfr states. Backbone assignment was performed both using Targeted Acquisition, an automated non-uniform sampling and assignment program [145–147], and manual assignment.

From the secondary chemical shift analysis, we could in **Paper I**

4.1. NMR spectroscopy of proteins

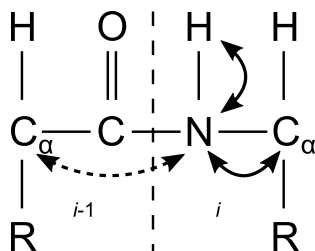


Figure 4.2: HNCA magnetization transfer. As can be seen, the experiment also recognizes the residue $i-1$ α -carbon (dashed line), and thus gives sequential information in addition to internal couplings. This is used for linking sequential residues in the protein.

show that while the second arm of the *DrBphP* tongue folds as an α -helix, the first arm is a stable β -strand in Pfr state.

The tongue is highly dynamic in Pr state as evident from the data collected on *DrBphP* with isotope-labeled lysines. The disappearance of a signal for the tongue lysine (K476) strongly suggests micro- to millisecond structural dynamics.

While secondary chemical shift analysis showed no change in secondary structure for the knot region (**Paper II**), chemical shift perturbation revealed major changes in the knot and helical spine. The residues showing the largest changes are located between the biliverdin and helical spine, and could manifest a new signaling pathway in phytochromes.

Further, RDC measurements on both states were made using bacteriophage Pf1 as alignment media. These showed good correlation between Pr crystal structure and observed Pr state data, but poor correlation to the Pfr structure. This can be interpreted that the crystal structure of Pfr state is a poor approximation of the solution state structure.

4.2 Infrared spectroscopy of proteins

4.2.1 History of vibrational spectroscopy

Infrared (one type of vibrational) spectroscopy dates back to the discovery of IR (infrared) radiation in 1800 by Sir William Herschel [148]. Following Sir Isaac Newton's first spectroscopic experiments in the 1600's [149], in the late 1800's and early 1900's IR spectra of hundreds of compounds were recorded.

It was among others noted that methyl groups absorb IR at characteristic wavelengths, and in the following hundred years IR spectroscopy developed into an extremely powerful tool for analyzing small molecules. The infrared spectrum is very information rich, and has therefore also been applied to larger molecules such as proteins.

IR spectroscopy is just one of a few different spectroscopic techniques that qualify as vibrational, others are for instance Raman and photoacoustic spectroscopy [86]. However, these will not be detailed further in this work.

4.2.2 Short theory of protein IR spectroscopy

Vibrational spectroscopy probes the molecular absorption energy at different resonant frequencies. These resonant frequencies arise from vibrations in chemical bonds, so called vibrational modes, such as stretching and bending of the bond.

In order for IR spectroscopy to pick up the chemical bond vibrations, there need to be a change in dipole moment. Symmetric molecules, such as N_2 , are therefore not "IR active". The resonant frequencies of vibrational modes follows Hooke's law which states that the resonant frequency (f) is proportional to the bond strength (k) and inversely proportional to the reduced mass (m_1 and m_2) of the bonded atoms (see Figure 4.3a).

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad \mu = \frac{m_1 m_2}{m_1 + m_2}$$

4.2. Infrared spectroscopy of proteins

As the environment (inter- and intramolecular effects) influences the frequency and dipole moments, vibrational spectra contain information on the chemical environment of the absorbing group. The absorption bands in an IR spectrum give information on the chemical structure of the molecule [86].

Non-linear molecules, such as proteins, with N number of atoms have $3N - 6$ vibrational modes, and thus the spectra are extremely information rich but gets very crowded and assignment of the individual spectral peaks is a formidable task.

In larger molecules, such as proteins, the spectra are congested but some bands stand out as particularly strong. The main peptide chain amide I (mainly C=O stretching) and amide II (mainly N-H bending and C-N stretching) bands are among these. All residues contribute to these vibrations. Figure 4.3b depicts the origin of the amide I and II bands.

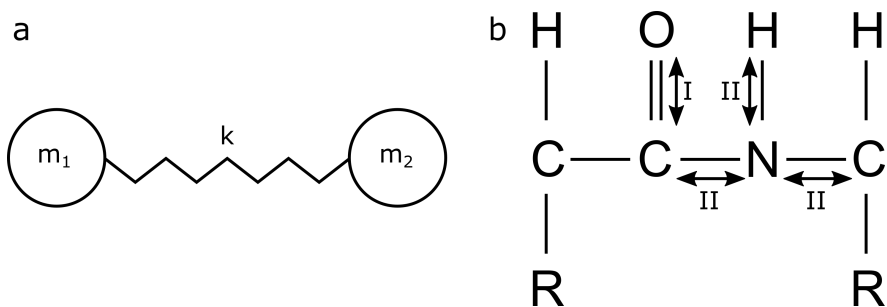


Figure 4.3: Vibrational bands arise from the resonant frequencies which follow Hooke's law regarding bond strength (k) and the mass of the bonded atoms (m_1 and m_2) seen in panel a. In panel b the origin of amide I and II vibrations are shown.

Several variants of IR spectroscopy has been developed. Two methods covered here are time-resolved FTIR and femtosecond infrared spectroscopy (or transient absorption spectroscopy).

FTIR differs from traditional absorption techniques in that it collects the entire spectral range simultaneously as interferograms. The beam is split and diverted to either a fixed mirror or a mirror which is moved discrete distances between each measurement. As the beams

are reflected back and recombined they are out of phase due to the shifted position of the mirror. This creates varying constructive or destructive interference. The beam is then passed through the sample and some frequencies are absorbed. This is repeated for several mirror positions, and the changes in the interferogram are then Fourier transformed (hence FT-IR) into a spectrum.

The two main advantages using FTIR is the higher signal-to-noise ratio and the greater speed at which a spectrum can be recorded. For standard measurements, FTIR has become the method of choice. Traditional absorption techniques find their uses in the time-resolved methods further detailed in the coming subsections.

As almost all molecules give rise to absorption bands in the IR spectrum, isotope labeling may seem a bit redundant. But inspecting Hooke's law and the equation for reduced mass shows that an increase in atomic mass shifts the vibrational band to a lower frequency. This possibility to shift peaks in the spectrum can be extremely useful, as assignment of the peaks in a congested spectrum is so difficult.

By labeling only selected parts of a protein, the absorption bands from the selected region can be resolved from the congested spectra due to their spectral shifts.

The isotopic shift is hard to predict, the contribution to the absorbance from each atom is not clearly defined. They can to some extent be guided by DFT (density functional theory) calculations as was done in **Paper IV**. Several labeling schemes were proposed and the selection was guided by which scheme showed the largest calculated isotopic shift.

4.2.3 Probing structural changes by IR spectroscopy

In IR spectroscopy, proteins give rise to a number of bands from the amide groups (amide A, B and I-VII), originating from the main peptide chain [150]. In addition, all side chains have their particular set of vibrational bands in the spectrum, which also are affected by the local chemical environment. Infrared spectrum for all amino acids has been reported by Hellwig [151] and Barth [86], and even though

4.2. Infrared spectroscopy of proteins

the bands can shift dramatically in the protein context these studies can serve as guidelines when trying to assign protein bands.

When talking about the local chemical environment, there are several factors that contribute to the shifting of bands. The major contributor comes from the chemical structure of the vibrating group. The chemical structure changes with, for example, protonation or redox state of the group. The chemical structure of neighboring groups, hydrogen bonding, electric fields and bond strengths and angles also contribute. Any of these changes will get picked up in the IR spectrum as a change in the absorption bands to a lesser or larger extent.

In order to filter out changes from the heavily congested spectrum it could be useful to measure difference spectra. A difference spectrum would give information only on changes in the vibrational modes. Negative peaks (or bleaches) represent a decrease in the population of a state and positive peaks an increase. Difference measurements present a way to decrowd spectra where changes in structure or local chemical environment occur.

This can be done as so-called steady-state, where a final state spectrum is compared to an initial state spectrum, giving information on changes between two discrete states. It can also be done at discrete time-points along the reaction, which will yield a time-resolved IR experiment, in order to follow the reaction in a time-resolved manner. Two types of time-resolved experiments used in this work will be detailed in the following subsections.

4.2.4 Step-scan FTIR of proteins

Step-scan FTIR spectroscopy is a time-resolved FTIR experiment operating in the tens of nanoseconds to hundreds of milliseconds time regime. It makes use of the higher signal-to-noise and measurement speeds of the FTIR, as changes in absorbance tend to be very signal weak and require several cycles of measurement.

It functions by detecting the decay of signal over time at each mirror position. The time-resolution is thus dependent on the detector response time and how fast the transient recorder can convert the signal to a digital one.

The need for several consecutive measurements at different mirror positions and signal averaging to increase signal-to-noise ratio puts a high demand on the sample to tolerate being pushed back and forth between initial and final states for up to several hours. Alternatively the sample can be exchanged for a spectroscopically identical sample after each cycle [152–155].

In **Paper III** we apply this technique in order to resolve how the isomerization of the chromophore couples to the global structural rearrangement of *DrBphP*. We performed time-resolved step-scan FTIR on both monomeric and dimeric *DrBphP* photosensory module. The protein was pushed between the two meta-stable Pr and Pfr states, and spectra were recorded with a time-resolution of $2.5\mu\text{s}$.

4.2.5 Ultra-fast IR spectroscopy of proteins

Many photochemical reactions take place at timescales shorter than step-scan FTIR is able to detect. For this purpose femtosecond transient IR spectroscopy has become a method of choice. With this technique it is possible to follow reactions on a femtosecond timescale, up to the nanoseconds.

A short pump-pulse is generated, which excites the sample followed by a probe-pulse, and difference in absorption is measured between pump-on and pump-off. The timing between pump and probe pulses (delay time) is varied by changing the optical path length of the pump-pulse. Repeating the measurement for several different delay times gives a change in difference spectra, and thus information on the reaction, over time. Presently, however, the technique is limited to proteins and protein reactions which can be triggered by a laser flash.

In **Paper IV** femtosecond IR spectroscopy was used, in combination with site-selective isotope labeling of tyrosine 66 in GFP. We could unambiguously identify a short-lived (3ps) intermediate state between photoexcitation and deprotonation of tyrosine 66.

The application range for IR spectroscopy is huge and for deeper insights into other vibrational spectroscopy techniques than for biomolecules and the theory behind, see Stuart [156].

4.3 Summary

Various spectroscopic techniques have long been used to probe structural dynamics of proteins in solution.

We used solution NMR spectroscopy to follow structural dynamics in *DrBphP*. In **Paper I** we show how changes in structural dynamics of the tongue region is used in photomodulation of phytochromes.

In **Paper II** we use solution NMR spectroscopy to investigate the knot region as a possible novel signaling pathway of phytochromes.

Step-scan FTIR experiments was used in **Paper III** to show how isomerization of the biliverdin chromophore couples to global structural rearrangements in *DrBphP*.

In **Paper IV** femtosecond IR spectroscopy revealed a previously proposed intermediate state in GFP.

The results given here give an insight into what can be achieved with the vast collection of spectroscopic applications available for protein science.

Chapter 5

Concluding remarks

All the work conducted in the making of this thesis (with exception of **Paper V**) have been probing structural dynamic events in proteins. Most of the work has been carried out using a range of isotope labeling schemes and spectroscopic techniques.

The large number of techniques for labeling proteins with either isotopes or unnatural amino acids, in combination with the various spectroscopic techniques available, result in almost endless possibilities to investigate protein structural dynamics.

Solution NMR spectroscopy has long been used to investigate proteins, and solve their structures. In **Paper I** we investigate the changes in structural dynamics of *DrBphP* tongue region. We show that the tongue is highly dynamic in Pr state and assumes a stable α -helix and β -strand in Pfr state. This suggests photomodulation of phytochromes is coupled to structural heterogeneity and changes in structural dynamics from soft-to-rigid.

These conclusions were aided by the amino acid-selective labeling of lysines in the protein, which revealed structural dynamics in the Pr state tongue region on a micro- to millisecond timescale.

In **Paper II** we use a selection of solution NMR spectroscopy experiments to investigate the knot region in the PAS-GAF bidomain. Changes in amide chemical shifts reveal large changes in structure between Pr and Pfr state. Secondary chemical shifts show that these changes are not coupled to changes in secondary structure.

RDC experiments also reveal a poor correlation between solution and crystal structures of the Pfr state. Further RDC experiments

will make it possible to refine the solution structural change of the *DrBphP* PAS-GAF-PHY.

Two types of time-resolved infrared spectroscopic techniques were also used in this thesis.

Step-scan FTIR revealed how the isomerization of biliverdin couples to global structural rearrangements in *DrBphP* in **Paper III**. Uniform isotope labeling of the apoprotein allowed us to distinguish the protein absorption bands from those of the biliverdin.

Femtosecond IR spectroscopy was used to investigate the initial events after photoexcitation in GFP (**Paper IV**). By site-specifically labeling tyrosine 66, we could identify a previously proposed intermediate state between photoexcitation of tyrosine 66 and proton transfer.

This site-specific labeling protocol can be used for other proteins in order to highly selectively monitor the active site or specific residues role in chemical reactions.

In **Paper V** we could also detail a CFPE system able to produce a functional pore-only sodium channel membrane protein. Site-specific labeling could in the future reveal how sodium transport is conducted over the membrane on a molecular level.

Bibliography

- [1] "The light fantastic," *Nature Chemical Biology*, vol. 10, p. 483, 6 2014.
- [2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 1 2000.
- [3] H. Hartley, "Origin of the Word 'Protein'," *Nature*, vol. 168, p. 244, 8 1951.
- [4] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis," *Nature*, vol. 181, no. 4610, pp. 662–666, 1958.
- [5] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins.," *Nature*, vol. 450, no. 7172, pp. 964–972, 2007.
- [6] U. Hensen, T. Meyer, J. Haas, R. Rex, G. Vriend, and H. Grubmüller, "Exploring protein dynamics space: the dynasome as the missing link between protein structure and function.," *PLoS one*, vol. 7, no. 5, p. e33931, 2012.
- [7] H. J. Dyson and P. E. Wright, "Unfolded Proteins and Protein Folding Studied by NMR," *Chemical Reviews*, vol. 104, pp. 3607–3622, 8 2004.
- [8] P. E. Wright and H. J. Dyson, "Linking folding and binding," *Current opinion in structural biology*, vol. 19, no. 1, pp. 31–38, 2009.
- [9] Y. Ryu and P. G. Schultz, "Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*," *Nature Methods*, vol. 3, p. 263, 3 2006.
- [10] K. Lang and J. W. Chin, "Cellular incorporation of unnatural amino acids and bioorthogonal Labeling of Proteins," *Chemical Reviews*, vol. 114, no. 9, pp. 4764–4806, 2014.
- [11] W. W. Garner Allard, H.A., "Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants," *Journal of Agricultural Research*, vol. 18, pp. 553–606, 1920.
- [12] H. A. Borthwick, S. B. Hendricks, M. W. Parker, E. H. Toole, and V. K. Toole, "A Reversible Photoreaction Controlling Seed Germination," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38, no. 8, pp. 662–666, 1952.
- [13] S. J. Davis, A. V. Vener, and R. D. Vierstra, "Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria," *Science*, vol. 286, no. 5449, pp. 2517–2520, 1999.

- [14] K. C. Yeh and J. C. Lagarias, "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*, vol. 95, no. 23, pp. 13976–13981, 1998.
- [15] C. Fankhauser, "The Phytochromes, a Family of Red/Far-red Absorbing Photoreceptors," *Journal of Biological Chemistry*, vol. 276, no. 15, pp. 11453–11456, 2001.
- [16] N. C. Rockwell, D. Duanmu, S. Shelley, C. Bachy, D. C. Price, D. Bhattacharya, A. Z. Worden, J. C. Lagarias, N. C. Rockwell, D. Duanmu, S. S. Martin, C. Bachy, D. C. Price, and D. Bhattacharya, "Eukaryotic algal phytochromes span the visible spectrum," *Proceedings of the National Academy of Sciences*, vol. 111, no. 10, p. 3871–3876, 2014.
- [17] J. Hughes, T. Lamparter, F. Mittmann, E. Hartmann, W. Gartner, A. Wilde, and T. Borner, "A prokaryotic phytochrome," *Nature*, vol. 386, no. 6626, p. 663, 1997.
- [18] Z. Y. Jiang, L. R. Swem, B. G. Rushing, S. Devanathan, G. Tollin, and C. E. Bauer, "Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes," *Science*, vol. 285, no. 5426, pp. 406–409, 1999.
- [19] H. A. W. Schneider-Poetsch, U. Kolukisaoglu, D. H. Clapham, J. Hughes, and T. Lamparter, "Non-angiosperm phytochromes and the evolution of vascular plants," *Physiologia Plantarum*, vol. 102, no. 4, pp. 612–622, 1998.
- [20] N. C. Rockwell and J. C. Lagarias, "A Brief History of Phytochromes," *Chemphyschem*, vol. 11, no. 6, pp. 1172–1180, 2010.
- [21] M. Ikeuchi and T. Ishizuka, "Cyanobacteriochromes: a new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria," *Photochemical & Photobiological Sciences*, vol. 7, no. 10, pp. 1159–1167, 2008.
- [22] A. Björling, O. Berntsson, H. Lehtivuori, H. Takala, A. J. Hughes, M. Panman, M. Hornke, S. Niebling, L. Henry, R. Henning, I. Kosheleva, V. Chukharev, N. V. Tkachenko, A. Menzel, G. Newby, D. Khakhulin, M. Wulff, J. A. Ihalainen, and S. Westenhoff, "Structural photoactivation of a full-length bacterial phytochrome," *Science Advances*, vol. 2, no. 8, pp. e1600920–e1600920, 2016.
- [23] M. E. Auldridge and K. T. Forest, "Bacterial phytochromes: more than meets the light," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 46, no. 1, pp. 67–88, 2011.
- [24] E. S. Burgie and R. D. Vierstra, "Phytochromes: an atomic perspective on photoactivation and signaling," *Plant Cell*, vol. 26, no. 12, pp. 4568–4583, 2014.
- [25] P. H. Quail, "Phytochrome photosensory signalling networks," *Nature Reviews Molecular Cell Biology*, vol. 3, no. 2, pp. 85–93, 2002.
- [26] J. R. Wagner, J. S. Brunzelle, K. T. Forest, and R. D. Vierstra, "A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome," *Nature*, vol. 438, no. 7066, pp. 325–331, 2005.
- [27] E. S. Burgie, J. Zhang, and R. D. Vierstra, "Crystal Structure of Deinococcus Phytochrome in the Photoactivated State Reveals a Cascade of Structural Rearrangements during Photoconversion.," *Structure*, vol. 24, no. 3, pp. 448–457, 2016.

BIBLIOGRAPHY

- [28] J. R. Wagner, J. Zhang, J. S. Brunzelle, R. D. Vierstra, and K. T. Forest, "High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution," *Journal of Biological Chemistry*, vol. 282, no. 16, pp. 12298–12309, 2007.
- [29] E. S. Burgie, A. N. Bussell, J. M. Walker, K. Dubiel, and R. D. Vierstra, "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*, vol. 111, no. 28, pp. 10179–10184, 2014.
- [30] F. Li, E. S. Burgie, T. Yu, A. Héroux, G. C. Schatz, R. D. Vierstra, and A. M. Orville, "X-ray radiation induces deprotonation of the bilin chromophore in crystalline D. Radiodurans phytochrome," *Journal of the American Chemical Society*, vol. 137, no. 8, pp. 2792–2795, 2015.
- [31] H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Koshelova, R. Henning, A. Menzel, J. A. Ihalainen, and S. Westenhoff, "Signal amplification and transduction in phytochrome photosensors," *Nature*, vol. 509, no. 7499, pp. 245–248, 2014.
- [32] P. Edlund, H. Takala, E. Claesson, L. Henry, R. Dods, H. Lehtivuori, M. Panman, K. Pande, T. White, T. Nakane, O. Berntsson, E. Gustavsson, P. Båth, V. Modi, S. Roy-Chowdhury, J. Zook, P. Berntsen, S. Pandey, I. Poudyal, J. Tenboer, C. Kupitz, A. Barty, P. Fromme, J. D. Koralek, T. Tanaka, J. Spence, M. Liang, M. S. Hunter, S. Boutet, E. Nango, K. Moffat, G. Groenhof, J. Ihalainen, E. A. Stojković, M. Schmidt, and S. Westenhoff, "The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography," *Scientific Reports*, vol. 6, no. 1, p. 35279, 2016.
- [33] N. C. Rockwell, Y.-S. Su, and J. C. Lagarias, "Phytochrome structure and signaling mechanisms," *Annual Review of Plant Biology*, vol. 57, pp. 837–858, 6 2006.
- [34] S. E. Braslavsky, W. Gartner, and K. Schaffner, "Phytochrome photoconversion," *Plant Cell and Environment*, vol. 20, no. 6, pp. 700–706, 1997.
- [35] D. Davenport and J. A. C. Nicol, "Luminescence in Hydromedusae," *Proceedings of the Royal Society of London. Series B - Biological Sciences*, vol. 144, pp. 399 LP – 411, 11 1955.
- [36] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea," *Journal of Cellular and Comparative Physiology*, vol. 59, no. 3, pp. 223–239, 1962.
- [37] F. Yang, L. G. Moss, G. N. Phillips, G. N. Phillips Jr., and G. N. Phillips, "The molecular structure of green fluorescent protein," *Nature Biotechnology*, vol. 14, no. 10, pp. 1246–1251, 1996.
- [38] C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward, "Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein," *Biochemistry*, vol. 32, pp. 1212–1218, 2 1993.
- [39] A. Cramer, E. A. Whitehorn, E. Tate, and W. P. Stemmer, "Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling," *Nature Biotechnology*, vol. 14, no. 3, pp. 315–319, 1996.

- [40] R. Y. Tsien, "The Green Fluorescent Protein," *Annual Review of Biochemistry*, vol. 67, pp. 509–544, 6 1998.
- [41] J. J. van Thor, "Photoreactions and dynamics of the green fluorescent protein," *Chemical Society Reviews*, vol. 38, no. 10, pp. 2935–2950, 2009.
- [42] C. Fang, R. R. Frontiera, R. Tran, and R. A. Mathies, "Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy," *Nature*, vol. 462, no. 7270, pp. 200–204, 2009.
- [43] D. Stoner-Ma, A. A. Jaye, P. Matousek, M. Towrie, S. R. Meech, and P. J. Tonge, "Observation of excited-state proton transfer in green fluorescent protein using ultrafast vibrational spectroscopy," *Journal of the American Chemical Society*, 2005.
- [44] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, pp. 500–544, 8 1952.
- [45] J. Payandeh, T. Scheuer, N. Zheng, and W. A. Catterall, "The crystal structure of a voltage-gated sodium channel," *Nature*, vol. 475, p. 353, 7 2011.
- [46] D. Ren, B. Navarro, H. Xu, L. Yue, Q. Shi, and D. E. Clapham, "A Prokaryotic Voltage-Gated Sodium Channel," vol. 294, no. December, pp. 2372–2375, 2001.
- [47] A. Sula, J. Booker, L. C. Ng, C. E. Naylor, P. G. Decaen, and B. A. Wallace, "The complete structure of an activated open sodium channel," *Nature Communications*, vol. 8, pp. 1–9, 2017.
- [48] S. Y. Noskov, S. Bernéche, and B. Roux, "Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands," *Nature*, vol. 431, no. 7010, pp. 830–834, 2004.
- [49] D. A. Köpfer, C. Song, T. Gruene, G. M. Sheldrick, U. Zachariae, and B. L. de Groot, "Ion permeation in K⁺ channels occurs by direct Coulomb knock-on," *Science*, vol. 346, pp. 352 LP – 355, 10 2014.
- [50] G. Sezonov, D. Joseleau-Petit, and R. D'Ari, "Escherichia coli physiology in Luria-Bertani broth," *Journal of Bacteriology*, vol. 189, no. 23, pp. 8746–8749, 2007.
- [51] S. Y. Lee, "High cell-density culture of Escherichia coli," *Trends in Biotechnology*, vol. 14, no. 3, pp. 98–105, 1996.
- [52] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in Escherichia coli: advances and challenges," *Frontiers in Microbiology*, vol. 5, p. 172, 2014.
- [53] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in microbial systems," *Frontiers in Microbiology*, vol. 5, p. 341, 2014.
- [54] F. Bolivar, R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow, "Construction and characterization of new cloning vehicle. II. A multipurpose cloning system," *Gene*, vol. 2, no. 2, pp. 95–113, 1977.

BIBLIOGRAPHY

- [55] J. Rogé and J.-M. Betton, "Use of pIVEX plasmids for protein overproduction in *Escherichia coli*," *Microbial Cell Factories*, vol. 4, p. 18, 6 2005.
- [56] D. Shaya, M. Kreir, R. A. Robbins, S. Wong, J. Hammon, A. Brüggemann, and D. L. Minor, "Voltage-gated sodium channel (NaV) protein dissection creates a set of functional pore-only proteins," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 12313 LP – 12318, 7 2011.
- [57] P. Daegelen, F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim, "Tracing Ancestors and Relatives of *Escherichia coli* B, and the Derivation of B Strains REL606 and BL21(DE3)," *Journal of Molecular Biology*, vol. 394, no. 4, pp. 634–643, 2009.
- [58] G. Bertani, "STUDIES ON LYSOGENESIS I. : The Mode of Phage Liberation by Lyso-genic *Escherichia coli*," *Journal of Bacteriology*, vol. 62, pp. 293–300, 9 1951.
- [59] M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht, and P. C. Zamecnik, "A soluble ribonucleic acid intermediate in protein synthesis," *Journal of Biological Chemistry*, vol. 231, pp. 241–257, 3 1958.
- [60] M. W. Nirenberg and J. H. Matthaei, "The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, no. 10, pp. 1588–602, 1961.
- [61] G. Zubay, "In Vitro Synthesis of Protein in Microbial Systems," *Annual Review of Genetics*, vol. 7, pp. 267–287, 12 1973.
- [62] A. S. Spirin, V. I. Baranov, L. A. Ryabova, S. Y. Ovodov, and Y. B. Alakhov, "A continuous cell-free translation system capable of producing polypeptides in high yield," *Science*, vol. 242, no. 4882, pp. 1162–1164, 1988.
- [63] M. N. Chekulayeva, O. V. Kurnasov, V. A. Shirokov, and A. S. Spirin, "Continuous-Exchange Cell-Free Protein-Synthesizing System: Synthesis of HIV-1 Antigen Nef," *Biochemical and Biophysical Research Communications*, vol. 280, no. 3, pp. 914–917, 2001.
- [64] F. Katzen, G. Chang, and W. Kudlicki, "The past, present and future of cell-free protein synthesis," *Trends in Biotechnology*, vol. 23, no. 3, pp. 150–156, 2005.
- [65] A. Zemella, L. Thoring, C. Hoffmeister, and S. Kubick, "Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems," *ChemBioChem*, vol. 16, pp. 2420–2431, 10 2015.
- [66] M. W. Nirenberg, "[3] Cell-free protein synthesis directed by messenger RNA," 1963.
- [67] J. Pratt, "Coupled transcription-translation in prokaryotic cell-free systems," *Transcription and Translation: a Practical Approach.*, pp. 179–209, 1984.
- [68] T.-W. Kim, J.-W. Keum, I.-S. Oh, C.-Y. Choi, C.-G. Park, and D.-M. Kim, "Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system," *Journal of Biotechnology*, vol. 126, no. 4, pp. 554–561, 2006.

- [69] M. C. Jewett and J. R. Swartz, "Mimicking the Escherichia coli Cytoplasmic Environment Activates Long-Lived and Efficient Cell-Free Protein Synthesis," *Biotechnology and Bioengineering*, vol. 86, no. 1, pp. 19–26, 2004.
- [70] A. S. Spirin and J. R. Swartz, *Cell-free protein synthesis: methods and protocols*. John Wiley & Sons, 2008.
- [71] G. Rosenblum and B. S. Cooperman, "Engine out of the chassis: Cell-free protein synthesis and its uses," *FEBS Letters*, vol. 588, no. 2, pp. 261–268, 2014.
- [72] Y.-C. Kwon and M. C. Jewett, "High-throughput preparation methods of crude extract for robust cell-free protein synthesis," *Scientific Reports*, vol. 5, p. 8663, 3 2015.
- [73] R. W. Martin, B. J. Des Soye, Y. C. Kwon, J. Kay, R. G. Davis, P. M. Thomas, N. I. Majewska, C. X. Chen, R. D. Marcum, M. G. Weiss, A. E. Stoddart, M. Amiram, A. K. Ranji Charna, J. R. Patel, F. J. Isaacs, N. L. Kelleher, S. H. Hong, and M. C. Jewett, "Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids," *Nature Communications*, vol. 9, no. 1, pp. 1–9, 2018.
- [74] A. Nieß, J. Failmezger, M. Kuschel, M. Siemann-Herzberg, and R. Takors, "Experimentally Validated Model Enables Debottlenecking of in Vitro Protein Synthesis and Identifies a Control Shift under in Vivo Conditions," *ACS Synthetic Biology*, vol. 6, no. 10, pp. 1913–1921, 2017.
- [75] J. C. Stark, A. Huang, P. Q. Nguyen, R. S. Dubner, K. J. Hsu, T. C. Ferrante, M. Anderson, A. Kanapskyte, Q. Mucha, J. S. Packett, P. Patel, R. Patel, D. Qaq, T. Zondor, J. Burke, T. Martinez, A. Miller-Berry, A. Puppala, K. Reichert, M. Schmid, L. Brand, L. R. Hill, J. F. Chellaswamy, N. Faheem, S. Fetherling, E. Gong, E. M. Gonzalzes, T. Granito, J. Koritsaris, B. Nguyen, S. Ottman, C. Palfy, A. Patel, S. Skweres, A. Slaton, T. Woods, N. Donghia, K. Pardee, J. J. Collins, and M. C. Jewett, "BioBits™ Bright: A fluorescent synthetic biology education kit," *Science Advances*, vol. 4, 8 2018.
- [76] A. Huang, P. Q. Nguyen, J. C. Stark, M. K. Takahashi, N. Donghia, T. Ferrante, A. J. Dy, K. J. Hsu, R. S. Dubner, K. Pardee, M. C. Jewett, and J. J. Collins, "BioBits™ Explorer: A modular synthetic biology education kit," *Science Advances*, vol. 4, 8 2018.
- [77] A. Pedersen, K. Hellberg, J. Enberg, and B. G. Karlsson, "Rational improvement of cell-free protein synthesis," *New Biotechnology*, vol. 28, no. 3, pp. 218–224, 2011.
- [78] T. W. Kim, D. M. Kim, and C. Y. Choi, "Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system," *Journal of Biotechnology*, vol. 124, no. 2, pp. 373–380, 2006.
- [79] L. Isaksson, *Intrinsically Disordered Domains of the B Cell Receptor Cell-Free Expression and Characterization by NMR*. PhD thesis, University of Gothenburg, 2014.
- [80] L. Isaksson, J. Enberg, R. Neutze, G. B. Karlsson, and A. Pedersen, "Expression screening of membrane proteins with cell-free protein synthesis," *Protein Expression and Purification*, vol. 82, no. 1, pp. 218–225, 2012.
- [81] C. Klammt, D. Schwarz, F. Löhr, B. Schneider, V. Dötsch, and F. Bernhard, "Cell-free expression as an emerging technique for the large scale production of integral membrane protein," *FEBS Journal*, vol. 273, no. 18, pp. 4141–4153, 2006.

BIBLIOGRAPHY

- [82] L. Thoring, S. K. Dondapati, M. Stech, D. A. Wüstenhagen, and S. Kubick, "High-yield production of "difficult-to-express" proteins in a continuous exchange cell-free system based on CHO cell lysates," *Scientific Reports*, vol. 7, no. 1, pp. 1–15, 2017.
- [83] J. Li, L. Gu, J. Aach, and G. M. Church, "Improved cell-free RNA and protein synthesis system," *PLoS ONE*, vol. 9, no. 9, 2014.
- [84] D. Gagoski, M. E. Polinkovsky, S. Mureev, A. Kunert, W. Johnston, Y. Gambin, and K. Alexandrov, "Performance benchmarking of four cell-free protein expression systems," *Biotechnology and Bioengineering*, vol. 113, no. 2, pp. 292–300, 2016.
- [85] J. R. Rumble, D. R. Lide, and T. J. Bruno, *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data*. 2017.
- [86] A. Barth, "Infrared spectroscopy of proteins," *Biochimica et Biophysica Acta - Bioenergetics*, vol. 1767, no. 9, pp. 1073–1101, 2007.
- [87] P. I. Haris, G. T. Robillard, A. A. Van Dijk, and D. Chapman, "Potential of carbon-13 and nitrogen-15 labeling for studying protein-protein interactions using Fourier-transform infrared spectroscopy," *Biochemistry*, vol. 31, pp. 6279–6284, 7 1992.
- [88] K. P. Das, L. P. Choo-Smith, J. M. Petrash, and W. K. Surewicz, "Insight into the secondary structure of non-native proteins bound to a molecular chaperone alpha-crystallin. An isotope-edited infrared spectroscopic study.," *The Journal of Biological Chemistry*, vol. 274, pp. 33209–12, 11 1999.
- [89] M. Zhang, H. Fabian, H. H. Mantsch, and H. J. Vogel, "Isotope-edited Fourier transform infrared spectroscopy studies of calmodulin's interaction with its target peptides.," *Biochemistry*, vol. 33, pp. 10883–10888, 9 1994.
- [90] J. Marley, M. Lu, and C. Bracken, "A method for efficient isotopic labeling of recombinant proteins," *Journal of Biomolecular NMR*, vol. 20, pp. 71–75, 2001.
- [91] M. Markus, K. Dayie, P. Matsudaira, and G. Wagner, "Effect of Deuteration on the Amide Proton Relaxation Rates in Proteins. Heteronuclear NMR Experiments on Villin 14T," *Journal of Magnetic Resonance, Series B*, vol. 105, pp. 192–195, 10 1994.
- [92] C. H. Arrowsmith, G.-S. Yi, and A. Ayed, "NMR of large (> 25 kDa) proteins and protein complexes," *Dynamics, Structure, and Function of Biological Macromolecules*, vol. 315, p. 133, 2001.
- [93] V. Tugarinov, V. Kanelis, and L. E. Kay, "Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy," *Nature Protocols*, vol. 1, no. 2, pp. 749–754, 2006.
- [94] N. K. Goto and L. E. Kay, "New developments in isotope labeling strategies for protein solution NMR spectroscopy," *Current Opinion in Structural Biology*, vol. 10, no. 5, pp. 585–592, 2000.
- [95] L. P. McIntosh and F. W. Dahlquist, "Biosynthetic Incorporation of ^{15}N and ^{13}C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins," *Quarterly Reviews of Biophysics*, vol. 23, no. 01, pp. 1–38, 1990.

- [96] A. E. L. Busche, A. S. Aranko, M. Talebzadeh-Farooji, F. Bernhard, V. Dötsch, and H. Iwai, "Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein," *Angewandte Chemie International Edition*, vol. 48, no. 33, pp. 6128–6131, 2009.
- [97] P. S. C. Wu, K. Ozawa, S. Jergic, X.-C. Su, N. E. Dixon, and G. Otting, "Amino-acid type identification in 15 N-HSQC spectra by combinatorial selective 15 N-labelling," *Journal of biomolecular NMR*, vol. 34, no. 1, pp. 13–21, 2006.
- [98] S. Sobhanifar, S. Reckel, F. Junge, D. Schwarz, L. Kai, M. Karbyshev, F. Löhr, F. Bernhard, and V. Dötsch, "Cell-free expression and stable isotope labelling strategies for membrane proteins," *Journal of biomolecular NMR*, vol. 46, no. 1, p. 33, 2010.
- [99] T. Yabuki, T. Kigawa, N. Dohmae, K. Takio, T. Terada, Y. Ito, E. D. Laue, J. A. Cooper, M. Kainosho, and S. Yokoyama, "Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis," *Journal of Biomolecular NMR*, vol. 11, no. 3, pp. 295–306, 1998.
- [100] M. Hiller, V. A. Higman, S. Jehle, B. J. Van Rossum, W. Kühlbrandt, and H. Oschkinat, "[2,3-13C]-labeling of aromatic residues - Getting a head start in the magic-angle-spinning NMR assignment of membrane proteins," *Journal of the American Chemical Society*, vol. 130, no. 2, pp. 408–409, 2008.
- [101] K. Ozawa, P. S. C. Wu, N. E. Dixon, and G. Otting, "15N-Labelled proteins by cell-free protein synthesis.," *FEBS Journal*, vol. 273, no. 18, pp. 4154–4159, 2006.
- [102] D. S. Waugh, "Genetic tools for selective labeling of proteins with α -15 N-amino acids," *Journal of Biomolecular NMR*, vol. 8, no. 2, pp. 184–192, 1996.
- [103] M. T. Lin, L. J. Sperling, H. L. F. Schmidt, M. Tang, R. I. Samoilova, T. Kumasaka, T. Iwasaki, S. A. Dikanov, C. M. Rienstra, and R. B. Gennis, "A rapid and robust method for selective isotope labeling of proteins," *Methods*, 2011.
- [104] P. J. Hajduk, D. J. Augeri, J. Mack, R. Mendoza, J. Yang, S. F. Betz, and S. W. Fesik, "NMR-Based Screening of Proteins Containing 13C-Labeled Methyl Groups," *Journal of the American Chemical Society*, vol. 122, pp. 7898–7904, 8 2000.
- [105] M. Hong and K. Jakes, "Selective and extensive 13C labeling of a membrane protein for solid-state NMR investigations," *Journal of Biomolecular NMR*, vol. 14, no. 1, pp. 71–74, 1999.
- [106] T. Kigawa, Y. Muto, and S. Yokoyama, "Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis," *Journal of Biomolecular NMR*, vol. 6, pp. 129–134, 9 1995.
- [107] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, and P. Güntert, "Optimal isotope labelling for NMR protein structure determinations," *Nature*, vol. 440, no. 7080, pp. 52–57, 2006.
- [108] J. Yokoyama, T. Matsuda, S. Koshihara, N. Tochio, and T. Kigawa, "A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution," *Analytical Biochemistry*, vol. 411, no. 2, pp. 223–229, 2011.

BIBLIOGRAPHY

- [109] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, and G. Otting, "Optimization of an *Escherichia coli* system for cell-free synthesis of selectively ^{15}N -labelled proteins for rapid analysis by NMR spectroscopy," *European Journal of Biochemistry*, vol. 271, no. 20, pp. 4084–4093, 2004.
- [110] J. Yokoyama, T. Matsuda, S. Koshiba, and T. Kigawa, "An economical method for producing stable-isotope labeled proteins by the *E. coli* cell-free system," *Journal of Biomolecular NMR*, vol. 48, no. 4, pp. 193–201, 2010.
- [111] J. P. Schwans, F. Sunden, A. Gonzalez, Y. Tsai, and D. Herschlag, "Uncovering the Determinants of a Highly Perturbed Tyrosine pKa in the Active Site of Ketosteroid Isomerase," *Biochemistry*, vol. 52, pp. 7840–7855, 11 2013.
- [112] H. T. Kratochvil, J. K. Carr, K. Matulef, A. W. Annen, H. Li, M. Maj, J. Ostmeier, A. L. Serrano, H. Raghuraman, S. D. Moran, J. L. Skinner, E. Perozo, B. Roux, F. I. Valiyaveetil, and M. T. Zanni, "Instantaneous ion configurations in the K^+ ion channel selectivity filter revealed by 2D IR spectroscopy," *Science*, vol. 353, pp. 1040 LP – 1044, 9 2016.
- [113] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, and H. Lau, "Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten," *Justus Liebigs Annalen der Chemie*, vol. 583, pp. 129–149, 8 1953.
- [114] P. E. Dawson, T. W. Muir, I. Clark-Lewis, and S. B. Kent, "Synthesis of proteins by native chemical ligation," *Science*, vol. 266, pp. 776 LP – 779, 11 1994.
- [115] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, "A general method for site-specific incorporation of unnatural amino acids into proteins," *Science*, 1989.
- [116] J. A. Ellman, D. Mendel, and P. G. Schultz, "Site-specific incorporation of novel backbone structures into proteins," *Science*, vol. 255, no. 5041, pp. 197–200, 1992.
- [117] K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, and G. Otting, "Multiple-site labeling of proteins with unnatural amino acids," *Angewandte Chemie International Edition*, vol. 51, no. 9, pp. 2243–2246, 2012.
- [118] S. M. Hecht, B. L. Alford, Y. Kuroda, and S. Kitano, "" Chemical aminoacylation" of tRNA's.," *Journal of Biological Chemistry*, vol. 253, no. 13, pp. 4517–4520, 1978.
- [119] C. J. Noren, S. J. Anthony-Cahill, K. A. Noren, M. C. Griffith, and P. G. Schultz, "In vitro suppression of an amber mutation by a chemically aminoacylated transfer RNA prepared by runoff transcription," *Nucleic Acids Research*, vol. 18, pp. 83–88, 1 1990.
- [120] S. A. Robertson, J. A. Ellman, and P. G. Schultz, "A general and efficient route for chemical aminoacylation of transfer RNAs," *Journal of the American Chemical Society*, vol. 113, pp. 2722–2729, 3 1991.
- [121] M. Lodder, S. Golovine, A. L. Laikhter, V. A. Karginov, and S. M. Hecht, "Misacylated Transfer RNAs Having a Chemically Removable Protecting Group," *The Journal of Organic Chemistry*, vol. 63, pp. 794–803, 2 1998.
- [122] S. M. Hecht, "Synthesis of Modified Proteins Using Misacylated tRNAs BT - Protein Engineering," pp. 255–270, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

- [123] C. C. Liu and P. G. Schultz, "Adding New Chemistries to the Genetic Code," *Annual Review of Biochemistry*, vol. 79, pp. 413–444, 6 2010.
- [124] J. W. Chin, "Expanding and Reprogramming the Genetic Code of Cells and Animals," *Annual Review of Biochemistry*, vol. 83, pp. 379–408, 6 2014.
- [125] N. Wu, A. Deiters, T. A. Cropp, D. King, and P. G. Schultz, "A Genetically Encoded Photocaged Amino Acid," *Journal of the American Chemical Society*, vol. 126, pp. 14306–14307, 11 2004.
- [126] S. E. Cellitti, D. H. Jones, L. Lagpacan, X. S. Hao, Q. Zhang, H. Y. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz, and B. H. Geierstanger, "In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy," *Journal of the American Chemical Society*, vol. 130, no. 29, pp. 9268–9281, 2008.
- [127] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, and P. G. Schultz, "Control of protein phosphorylation with a genetically encoded photocaged amino acid," *Nature Chemical Biology*, vol. 3, p. 769, 10 2007.
- [128] E. M. Purcell, H. C. Torrey, and R. V. Pound, "Resonance Absorption by Nuclear Magnetic Moments in a Solid," *Physical Review*, vol. 69, pp. 37–38, 1 1946.
- [129] F. Bloch, W. W. Hansen, and M. Packard, "Nuclear Induction," *Physical Review*, vol. 69, p. 127, 2 1946.
- [130] M. Saunders, A. Wishnia, and J. G. Kirkwood, "The nuclear magnetic resonance spectrum of Ribonuclease 1," *Journal of the American Chemical Society*, vol. 79, pp. 3289–3290, 6 1957.
- [131] M. P. Williamson, T. F. Havel, and K. Wüthrich, "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry," *Journal of Molecular Biology*, vol. 182, pp. 295–315, 3 1985.
- [132] D. S. Wishart and B. D. Sykes, "The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data," *Journal of Biomolecular NMR*, vol. 4, no. 2, pp. 171–180, 1994.
- [133] M. P. Williamson, "Using chemical shift perturbation to characterise ligand binding," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 73, pp. 1–16, 2013.
- [134] M. R. Hansen, L. Mueller, and A. Pardi, "Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions," *Nature Structural Biology*, vol. 5, p. 1065, 12 1998.
- [135] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard, "Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution," *Proceedings of the National Academy of Sciences*, vol. 92, pp. 9279 LP – 9283, 9 1995.

BIBLIOGRAPHY

- [136] N. Tjandra and A. Bax, "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium," *Science*, vol. 278, no. 5340, pp. 1111–1114, 1997.
- [137] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, and A. Bax, "Use of dipolar 1H – 15N and 1H – 13C couplings in the structure determination of magnetically oriented macromolecules in solution," *Nature Structural Biology*, vol. 4, p. 732, 9 1997.
- [138] M. Zweckstetter, "NMR: prediction of molecular alignment from structure using the PALES software," *Nature Protocols*, vol. 3, no. 4, p. 679, 2008.
- [139] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley & Sons, 2013.
- [140] J. Keeler, *Understanding NMR spectroscopy*. John Wiley & Sons, 2011.
- [141] L. E. Kay, M. Ikura, R. Tschudin, and A. Bax, "Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins," *Journal of Magnetic Resonance*, vol. 89, no. 3, pp. 496–514, 1990.
- [142] M. Ikura, L. E. Kay, and A. Bax, "A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin," *Biochemistry*, vol. 29, pp. 4659–4667, 5 1990.
- [143] E. Schmidt and P. Güntert, "A New Algorithm for Reliable and General NMR Resonance Assignment," *Journal of the American Chemical Society*, vol. 134, pp. 12817–12829, 8 2012.
- [144] D. E. Zimmerman, C. A. Kulikowski, Y. Huang, W. Feng, M. Tashiro, S. Shimotakahara, C.-y. Chien, R. Powers, and G. T. Montelione, "Automated analysis of protein NMR assignments using methods from artificial intelligence," *Journal of Molecular Biology*, vol. 269, no. 4, pp. 592–610, 1997.
- [145] L. Isaksson, M. Mayzel, M. Saline, A. Pedersen, J. Rosenl w, B. Brutscher, B. G. Karlsson, and V. Y. Orekhov, "Highly Efficient NMR Assignment of Intrinsically Disordered Proteins: Application to B- and T Cell Receptor Domains," *PLoS ONE*, vol. 8, no. 5, pp. 1–8, 2013.
- [146] V. A. Jaravine and V. Y. Orekhov, "Targeted acquisition for real-time NMR spectroscopy," *Journal of the American Chemical Society*, 2006.
- [147] V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, and V. Y. Orekhov, "Hyperdimensional NMR Spectroscopy with Nonlinear Sampling," *Journal of the American Chemical Society*, vol. 130, pp. 3927–3936, 3 2008.
- [148] W. Herschel, "XIV. Experiments on the refrangibility of the invisible rays of the sun," *Philosophical Transactions of the Royal Society of London*, vol. 90, pp. 284–292, 1800.
- [149] I. Newton, *Opticks, or, a treatise of the reflections, refractions, inflections & colours of light*. Courier Corporation, 1704.
- [150] G. G. Suchkova and L. I. Maklakov, "Amide bands in the IR spectra of urethanes," *Vibrational Spectroscopy*, vol. 51, no. 2, pp. 333–339, 2009.

- [151] M. Wolpert and P. Hellwig, "Infrared spectra and molar absorption coefficients of the 20 alpha amino acids in aqueous solutions in the spectral range from 1800 to 500 cm⁻¹," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 64, no. 4, pp. 987–1001, 2006.
- [152] W. Uhmann, A. Becker, C. Taran, and F. Siebert, "Time-Resolved FT-IR Absorption Spectroscopy Using a Step-Scan Interferometer," *Applied Spectroscopy*, vol. 45, no. 3, pp. 390–397, 1991.
- [153] I. Radu, M. Schlegler, C. Bolwien, and J. Heberle, "Time-resolved methods in biophysics. 10. Time-resolved FT-IR difference spectroscopy and the application to membrane proteins," *Photochemical & Photobiological Sciences*, vol. 8, no. 11, pp. 1517–1528, 2009.
- [154] T. Kottke, V. A. Lórenz-Fonfría, and J. Heberle, "The Grateful Infrared: Sequential Protein Structural Changes Resolved by Infrared Difference Spectroscopy," *The Journal of Physical Chemistry B*, vol. 121, pp. 335–350, 1 2017.
- [155] D. Magana, D. Parul, R. B. Dyer, and A. P. Shreve, "Implementation of Time-Resolved Step-Scan Fourier Transform Infrared (FT-IR) Spectroscopy Using a kHz Repetition Rate Pump Laser," *Applied spectroscopy*, vol. 65, pp. 535–542, 5 2011.
- [156] B. Stuart, "Infrared spectroscopy," *Kirk-Othmer Encyclopedia of Chemical Technology*, 2005.