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

Terahertz radiation as a pump and probe for studying low frequency vibrations in proteins

Ida Lundholm

UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology
Göteborg, Sweden
2015

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

Terahertz radiation as a pump and probe for studying low frequency vibrations in proteins

Ida Lundholm

Cover: Electron density differences emerging in lysozyme upon terahertz illumination.

ISBN 978-91-628-9543-3 (Pdf)
ISBN 978-91-628-9544-0 (Print)
Available online at http://hdl.handle.net/2077/40171

Department of Chemistry and Molecular Biology
Biochemistry and Biophysics
University of Gothenburg
SE-413 90 Göteborg, Sweden

Printed by Ale Tryckteam AB
Göteborg, Sweden, 2015

Abstract

Many functionally important structural changes in proteins proceed along the direction of their lowest frequency vibrations. These vibrations correspond to picosecond collective dynamics. Establishing the fundamental relationship between these vibrations and protein function remains a challenge within biophysics. Electromagnetic radiation in the terahertz frequency range (0.1-10 THz) can excite collective picosecond vibrations which makes it suitable as a probe for direct observation as well as a pump for the selection of functionally relevant vibrations for detection by other methods. The use of terahertz radiation for biological applications is hampered by several technical difficulties such as water absorption and thermal effects. For these reasons, method development is an important aspect when applying terahertz radiation to biological problems. In this work, terahertz radiation has been used to identify and characterize low frequency vibrations in three different proteins by applying both novel experimental design and analysis methods.

Terahertz absorption spectroscopy was used to identify the change in collective dynamics upon photo activation of a photosynthetic reaction centre. The collective vibrations were of non thermal origin and localized to the chromophore containing subunits implying the involvement of collective dynamics in photosynthesis.

By combining X-ray crystallography with 0.4 THz excitation the presence of collective dynamics was detected in both lysozyme and thermolysin. In lysozyme, the vibrational mode was localized to a central α-helix. The vibrational mode had a lifetime longer than expected which most likely arise from a hypothetical Fröhlich condensation process not previously observed. The interaction of terahertz radiation with thermolysin was identified through a Bayesian statistical analysis of X-ray diffraction data.

List of publications

This thesis is based on the following publications:

Paper I **I. Lundholm**, W. Y. Wahlgren, F. Piccirilli, P. Di Pietro, A. Duelli, O. Berntsson, S. Lupi, A. Perucchi and G. Katona, *Terahertz absorption of illuminated photosynthetic reaction centre solution: a signature of photo activation?*, RSC Advances (2014) 4(49):25502-25509

Paper II **I. Lundholm**, H. Rodilla, W. Y. Wahlgren, A. Duelli, G. Bourenkov, J. Vukusic, R. Friedman, J. Stake, T. Schneider and G. Katona, *Terahertz radiation induces non-thermal structural changes associated with Fröhlich condensation in a protein crystal*, Submitted manuscript (2015)

Paper III **I. Lundholm**, H. Rodilla, M. J. Garcia-Bonete, G. Gotthard, A. Royant, D. di Sanctis, J. Stake and G. Katona *Bayesian inference detects diffraction intensity changes upon terahertz irradiation of thermolysin single crystals*, Manuscript (2015)

Related publications

A. Duelli, B. Kiss, **I. Lundholm**, A. Bodor, M. V. Petoukhov, D. I. Svergun, L. Nyitray and G. Katona *The C-terminal Random Coil Region Tunes the Ca^{2+}-Binding Affinity of S100A4 through Conformational Activation*, PLOS ONE (2014) 9(5):p.e97654

Contribution report

Paper I I prepared the samples, conducted and planned two experiments at the synchrotron, performed the analysis of absorption spectra, was involved in writing the manuscript and prepared figures.

Paper II I prepared lysozyme crystals and collected X-ray diffraction data. I developed Python code for data handling and analysis and did all diffraction data analysis. I took a major role in writing the manuscript and prepared most figures.

Paper III I prepared crystals, collected data at ESRF and analysed the data. I was also involved in developing code for the Bayesian statistical inference. I wrote the manuscript and prepared all figures.

Abbreviations

BChl	Bacteriochlorophyll	**LSP**	Lipidic Sponge Phase
BPhe	Bacteriopheophytin	**MD**	Molecular Dynamics
CCD	Charge Coupled Device	**MO**	Monoolein
CS	Conformational Substates	**NMD**	Normal Mode Dynamics
FTIR	Fourier transform infrared spectroscopy	$\mathbf{P}_{870}$	the special pair
		$\mathbf{Q_A}$	Ubiquinone A
		$\mathbf{Q_B}$	Ubiquinone B
HDI	Highest Density Interval	$\mathbf{RC}_{sph}$	Reaction centre from *R. sphaeroides*
HEWL	Hen Egg White Lysozyme	$\mathbf{RC}_{vir}$	Reaction centre from *Bl. viridis*
IR	Infrared	**THz-TDS**	Terahertz Time Domain Spectroscopy
LCP	Lipidic Cubic Phase		
$\mathbf{LM}_{sph}$	L and M subunits from *R. sphaeroides* reaction centre	**UV**	Ultra Violet

Contents

Chapter 1

Introduction

"Everything that living things do can be understood in terms of the jigglings and wigglings of atoms"

Richard P. Feynman 1963

1.1 Protein dynamics

Proteins play an essential role in all forms of life. Proteins are versatile macromolecules which are responsible for nearly all tasks in a cell including for example catalysis of chemical reactions, transportation of molecules and cell mobility. The three dimensional structure of a protein is a prerequisite for acquiring its proper function but the structure is far from static, its dynamic properties is very important for its function [1]. The famous quote from Richard P. Feynman [2] in the opening of this introductory chapter, beautifully summarizes the importance of a dynamic view on biomolecules for a full under-

standing of their properties. This will also be the central theme of this thesis. The first protein structures were solved for myoglobin by Kendrew in 1958 [3] and haemoglobin by Perutz in 1964 [4] and these structures gave the first indications on the important structure-dynamics relationship. The conformational dynamics of myoglobin and haemoglobin let the proteins change their structures in response to oxygen or carbon dioxide binding. At the same time the different conformational states display dynamics of their own, dynamics which are essential for leaving oxygen and carbon dioxide a path to pass through.

The realization that proteins are flexible molecules led to the formulation of the induced-fit model for enzyme-substrate binding by Koshland in 1958 [5]. The induced-fit model states that an enzyme will adapt its structure to the substrate upon binding. This model embraces the importance of conformational dynamics for enzymatic function which opposed the previously believed rigid nature of an enzyme. At thermal equilibrium, dynamics are also important for enzymatic function. Dynamics are involved in the usually rate limiting product release steps which affects the turnover [6].

Protein dynamics within biochemistry includes both equilibrium and non-equilibrium effects while in the field of physics, dynamics is only ascribed to non-equilibrium effects. This semantic issue has made it difficult for physicists and biochemists to productively combine their knowledge. To avoid further confusion it is important to define the meaning of the term. In this thesis, I will refer to protein dynamics as any change in atomic coordinates over time and the focus will be on thermal equilibrium dynamics.

The dynamic fluctuations within a protein can be small and large, from ångströms to nanometers, and occur on a wide range of timescales,

from femtoseconds to seconds. The dynamic motions of proteins can be divided into three categories depending on which timescale they occur [7, 8] (Figure 1.1 and Table1.1):

1. **Tier 0 dynamics** - Structural changes on a millisecond time scale or slower between kinetically distinct states with an energy barrier of several kT between the states. k is the Boltzmann constant ($1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$) and T is the temperature in Kelvin.

2. **Tier 1 dynamics** - Picosecond to nanosecond fluctuations between closely related conformations separated by an energy barrier lower than kT involving collective motions of atoms.

3. **Tier 2 dynamics** - Only differs from tier 1 dynamics by describing localized events (non-collective behaviour.)

Table 1.1: Classification of dynamic tiers and Conformational Substates (CS)

Tier	Timescale	Energy barrier	Transition type
0	ms or slower	nkT	Kinetically distinct structural changes between CS_0
1	ps-ns	$< kT$	Collective vibrations between CS_1
2	ps-ns	$< kT$	Localized (non-collective) vibrations between CS_2

1.2 The free energy landscape

The free energy landscape is an important concept for bridging biological and physical sciences by providing a common ground for understanding protein function. The free energy landscape is mainly associated with the field of protein folding. When a protein gets translated by the ribosome from mRNA it spontaneously folds into a three dimensional structure, the native state. The correct folding of a protein is key for its proper function and several diseases are coupled to protein misfolding. Often misfolding causes loss of function, but if misfolding results in a modified structure it can even be toxic. This happens for example in Alzheimers and Creuzfeld Jacobs disease where the affected proteins acquire new structures and aggregate together to form fibrils. With the exceptions of misfolding diseases and intrinsically disordered proteins, a certain protein always acquire the exact same structure. The reproducibility of protein folding is a fundamental property invaluable for life to exist. In 1973 Anfinsen found that the proper folding into a functional protein do not need any biological machinery to form and the structure is instead ultimately encoded in the primary sequence [9]. Anfinsen postulated that at a certain set of environmental conditions a protein primary sequence of amino acids will fold into a unique, stable lowest energy conformation. One way to describe the protein folding problem is by a free energy landscape [10, 11] where all possible conformations of the primary sequence are mapped against their associated Gibbs free energy. The energy landscape generally takes the form of a funnel and the lowest energy state in the bottom of the funnel is the native folded state of the protein (Figure 1.1, left). The surface of the folding funnel is rugged and has several local minima which corresponds to folding

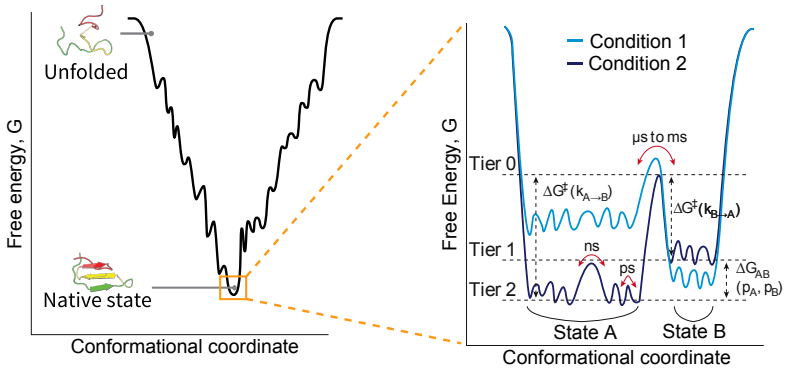

Figure 1.1: One dimensional representation of folding funnel (left) and the free energy landscape for the folded protein (right). Light and dark blue lines represent two different environmental conditions for the same protein at equilibrium. Two CS_0 states are shown, denoted A and B, with an energy barrier $> kT$ separating them. Within the wells of state A and B are several local minima corresponding to CS_1 and CS_2 that can be inter-converted through tier 1 and tier 2 dynamics, respectively. Right part of the figure reprinted with permission from reference [8].

pathway intermediates that guides the folding process towards the global minima of the landscape.

The concept of a free energy landscape can also be used to describe the dynamics of folded proteins and was pioneered by Frauenfelder and coworkers to classify and characterize the dynamics of myoglobin [7, 12]. The structure with a free energy corresponding to the global minima of the protein folding funnel can be further classified into Conformational Substates (CS) each occupying a local minima within the global minima. The free energy landscape is unique for a certain set of environmental conditions and describes the CS and the energy barriers separating them(Figure 1.1, right). The CS can be

inter-converted through the dynamic transition tiers previously described (Table 1.1). For example, myoglobin with carbon monoxide bound to its haeme group has three identified CS with different kinetic properties separated by tier 0 energy barriers, called CS_0 [7, 12, 13]. CS_0 are generally distinct and few in numbers and can therefore be characterized individually. The CS_0 states can be further divided into CS_1 and CS_2 which are separated by smaller energy barriers and can be inter-converted through tier 1 and tier 2 dynamics respectively. Within a CS_0 there are a large number of CS_1 and CS_2 with similar free energies. Therefore, CS_1 and CS_2 can not be characterized individually as the CS_0 and instead they need to be described statistically.

The transition state theory for describing enzyme catalysis can be combined with the free energy landscape concept. An enzymatic reaction depicted in a two dimensional transition state diagram assumes that only one given structure of the protein exists at a certain time point during the reaction. The enzyme can have a rugged energy landscape with several isoenergetic CS available during the course of the reaction. It has therefore been proposed that enzymes can catalyse a reaction through several parallel structural pathways where the structure at a certain point in time is defined in terms of the statistical free energy landscape [14]. The idea of a free energy landscape at thermal equilibrium also modifies the accepted induced-fit model of enzyme substrate binding. Instead of the substrate inducing a conformation in the enzyme, the enzyme can already adapt the optimal conformation for substrate binding (or close to optimal) at thermal equilibrium [15].

1.3 Low frequency vibrations

Low frequency vibrations correspond to both collective and localized dynamics in the picosecond to nanosecond time range (tier 1 and tier 2 dynamics, Table 1.1). Low frequency vibrations are fast compared to the microsecond to millisecond time scales of enzymatic reactions but slow compared to localized bond vibrations in proteins studied by Infrared (IR) spectroscopy. Since the 80's when Normal Mode Dynamics (NMD) and Molecular Dynamics (MD) simulations became available for proteins, functionally relevant vibrations have been found in the picosecond time range [16–23] but there is still a lack of experimental proof thereof. Normal modes are the resonant frequencies for a system at equilibrium. Protein normal modes describes collective dynamic behaviour at a certain frequency and are calculated through an harmonic approximation around a local minima in the free energy landscape. MD simulations are more computationally intensive and are not limited to one local minima and can instead explore the full energy landscape. Early modelling showed that large scale movements of proteins can be described by one or a few low frequency vibrational modes [24–31], demonstrating the direct coupling between function and these modes. NMD simulations have shown that the large amplitude structural changes between open and closed conformations of several different proteins can be extrapolated from a small amplitude low frequency normal mode [32].

The involvement of dynamics in enzyme function is a heavily debated topic, especially when it comes to dynamics being involved in chemical catalysis [6, 33–35]. Enzymatic reactions generally has a turnover in the microsecond to millisecond time range which is a result of the combined time scales of substrate binding, chemical catalysis

and product release. Dynamics on a wide range of time scales works in concert to give rise to a proteins' function and low frequency vibrations may be important for both enzymatic turnover and chemical catalysis. Low frequency vibrations can increase the turnover of an enzymatic reaction through assisting the rate limiting steps of substrate binding and product release. For example, collective low frequency vibrations have been shown to facilitate the slower tier 0 dynamics in the opening of the active site "lid" of adenylate kinase [36]. Fast vibrations can enhance chemical catalysis through so called promoting vibrations, which are vibrations along the reaction coordinate in the free energy landscape. Promoting vibrations increases the chance of an improbable event, such as crossing the transition barrier, and will ultimately increase the turnover [35]. Low frequency vibrations may also be important for enzyme function through mediating allosteric effects. Long range correlated motions provide the means for connecting distant regions without the need of a net conformational change [37].

Even if the exact role of low frequency modes are not fully understood they are of importance for protein function since these modes are evolutionary conserved [36, 38]. For adenylate kinase, two isoforms of the protein have similar normal modes present in the hinge region which assist the lid opening important for product release from the active site pocket. Proteins with similar fold but different functions have several normal modes in common, especially the lowest frequency mode is conserved while the modes important for specificity differs between different proteins [38].

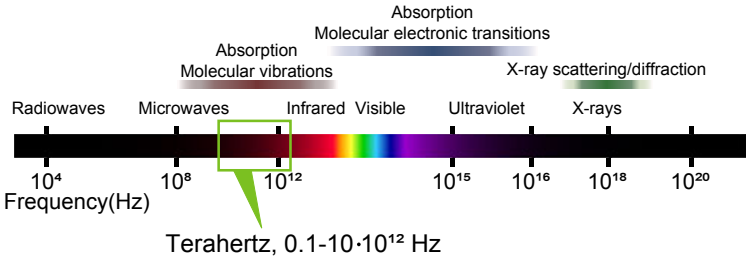

Figure 1.2: The electromagnetic spectrum and the typical interactions between radiation of different energies and a protein.

1.4 Terahertz radiation

Interactions of electromagnetic radiation with proteins is the basis for a wide range of scientific methods and what can be studied depends mainly on the energy of the radiation. High energy radation, X-rays, can be scattered by atoms giving information of the structure or be absorbed through the excitation of core electrons and give information of local structure and oxidation states of metal centres in proteins. Electromagnetic radiation in the Ultra Violet (UV) to visual range on the other hand, has a lower energy and can excite molecular electronic transitions in a protein. UV/Visual absorption spectroscopy is used for characterization, following a reaction, determining concentration and purity. Radiation in the IR and terahertz regions can excite vibrational and rotational state transitions. IR radiation interacts with bond vibrations while terahertz radiation (0.1-10 THz) excites collective low frequency vibrational modes on a timescale of 0.1-10 ps [39] (Figure 1.2).

Ever since the importance of low frequency vibrations for protein function started to be discussed more than three decades ago, tera-

hertz radiation was believed to hold the key for experimental proof thereof. Terahertz absorption spectroscopy could be used to verify results obtained with NMD and MD calculations but several difficulties with the technique has hampered its use for easy identification of low frequency vibrational modes. The main problems with terahertz techniques are the lack of affordable and practical systems for both production and detection of terahertz radiation as well as difficulties with measuring and interpreting terahertz absorption spectra. Despite the difficulties, several research groups have successfully measured terahertz absorption spectra of proteins under both dry and aqueous conditions [40–54]. Terahertz absorption spectra of proteins has been shown to be sensitive to excitation state [48], hydration [43, 46], oxidation state [51], ligand binding [53] and mutations [42]. Terahertz irradiation at 1.52 THz affects the binding rate of myoglobin which indicates an involvement of low frequency vibrations important for its function [55]. Terahertz vibrational modes of biological interest has been determined with a combination of terahertz absorption and calculations. For example, the bacteriorhodopsin conformational change at 3.45 THz [40], the porphyrin "doming" mode at 1.17 THz important for oxygen acceptance in haeme proteins [56] and the primary event of vision at 1.8 THz [57]. With careful experimental design and by using a combination of biophysical techniques, terahertz radiation can be a powerful probe for collective vibrational modes in proteins.

1.5 Photosynthetic reaction centre

Photosynthesis converts solar energy to chemical energy and is without exaggeration the most important chemical process on earth. Pho-

tosynthesis ultimately provides all animal life with energy and is the reason for our oxygen rich atmosphere. Photosynthesis is carried out by plants, algae and bacteria and the molecular machinery share similarities between all photosynthetic organisms, one being the presence of a photosynthetic reaction centre. The reaction centre is a membrane embedded protein complex that absorbs photons and performs the first step towards creating energy usable by the organism.

The most well known reaction centres are those from the two purple bacteria, *Rhodobacter sphaeroides* and *Blastochloris viridis*. The first structure of a membrane protein complex was solved for a Reaction centre from *Bl. viridis* (RC_{vir}) in 1984 [58] which later rewarded Deisenhofer, Michel and Huber a Nobel prize in 1988. The same year the structure was also solved for the the Reaction centre from *R. sphaeroides* (RC_{sph}) [59]. The bacterial reaction centres RC_{vir} and RC_{sph} has become two of the most studied integral membrane proteins [60]. In **Paper I**, the low frequency mode density involved in light activation of RC_{sph} is studied.

RC_{sph} comprises of three subunits, Light(L), Medium(M) and Heavy(H) (Figure 1.3). The L and M subunits are quasi symmetrical and are embedded in the membrane while the H subunit is associated outside the membrane on the cytoplasmic side with one long α-helix anchoring it to the membrane. The L and M subunits are coordinating nine cofactors, four bacteriochlorophylls (BChl), two bacteriopheophytins (BPhe), Ubiquinone A (Q_A), Ubiquinone B (Q_B) and one non-haeme iron [61]. The cofactors forms two symmetrical branches, A and B, where only A is active in electron transport [62]. The electron transport starts with the special pair (P_{870}), formed by the first two BChls, absorbing one photon and thereby reaching an excited state, P^*. P^* then transfers one electron to Q_A via $BChl_A$,

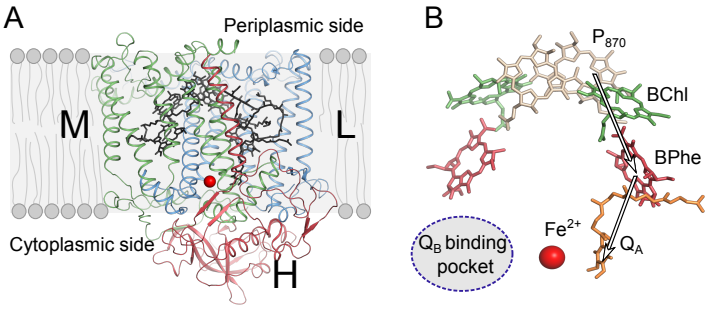

Figure 1.3: A. Overall structure of RC_{sph} and position in the membrane. **B.** The cofactors and the electron transfer (arrows) leading to the charge separation. Q_B is not present in the structure.

$BPhe_A$, creating a charge separated state $P^+Q_A^-$ [63] (Figure 1.3B). After the absorption of two photons, Q_A then transfers the electrons to the mobile quinone Q_B which thereby gets reduced and takes up two protons from the cytoplasm. The reduced Q_B continues the electron transfer chain by oxidizing the bc_1-cytochrome complex.

The H-subunit is evolutionary conserved among bacterial reaction centres but its function is not fully known. It has been suggested that the H-subunit function is to protect the quinones [64] or that it stabilizes the charge separated state [65]. The membrane anchoring α-helix may be involved in the preferential electron transfer along the cofactor A branch [66].

Long lived charge separated state

In the absence of Q_B or if RC becomes light saturated, P_{870} gets reduced by Q_A and the excess energy will dissipated as heat on an unknown time scale. This charge recombination is potentially harm-

ful for the protein both due to the generated heat and due to the possibility of unwanted reduction reactions by Q_A. The charge recombination is prevented by a stabilization of the charge separated state. Charge separation happens already within 200 ps after excitation [67] while the recombination takes 100 ms after excitation by a short light pulse [68]. If however RC is subjected to longer illumination, the charge separated state becomes further kinetically stabilized and recombination is prolonged to several minutes depending on the illumination protocol [69, 70]. The question is *how* the charge recombinated state is stabilized. Under continuous light conditions two kinetic phases can be identified for both charge separation and charge recombination while short light exposure results in only one kinetic phase [69]. The charge recombination process is complex and as a result, measured recombination kinetics is the average of life times given by the distribution of several conformational substates with different kinetic properties [70, 71]. Speculatively, the excess energy after recombination could give rise to a vibrationally excited state instead of thermalizing instantaneously. The vibrationally excited state could instead of, or in addition to, the putative structural changes give rise to different recombination kinetics.

There are several studies of light adapted reaction centers, both X-ray structures [72–74] and spectroscopic studies [69–71, 75–80], concluding that the long life time is associated with conformational changes in the excited state of the protein. Furthermore, light and dark adapted RC_{sph} cleaves into different fragments by trypsin. The different exposed cleaving sites for light and dark adapted RC_{sph} is explained by a structural change on the acceptor side of RC_{sph} upon light activation [81, 82]. Indication of structural changes upon illumination in different parts of the protein (for example references

[65, 72, 82]) as well as the different reaction time scales reported can be the result of a complex free energy landscape from which reaction centre samples a large number of CS. The population of CS as well as the free energy surface are dependent on the illumination protocol [12, 83].

Vibrations in electron transfer reactions

Tier 1 and 2 dynamics are believed to play a key role in electron transfer reactions. Theoretical models indicate the importance of several low frequency modes for both charge separation and charge recombination kinetics in reaction centres [84–89]. The theoretical models has also been strengthened by experimental observations showing the importance of dynamics for the electron transfer process [90]. Furthermore, the electron transfer in reaction centres are faster at lower temperatures [91] which indicates that there are no tier 0 dynamics involved. This also has the implication that the dynamics involved in charge transfer can be of quantum mechanical nature [92].

1.6 Enzyme model systems

In **Paper II** and **III** terahertz excited dynamics were studied on two model systems: the soluble enzymes lysozyme and thermolysin.

Lysozyme

Ever since its discovery by Fleming in 1922 [95] lysozyme has been used as a model protein for a wide range of studies within biology, biochemistry and biophysics. Lysozyme is an antibacterial enzyme present as a component of the immune system of organisms

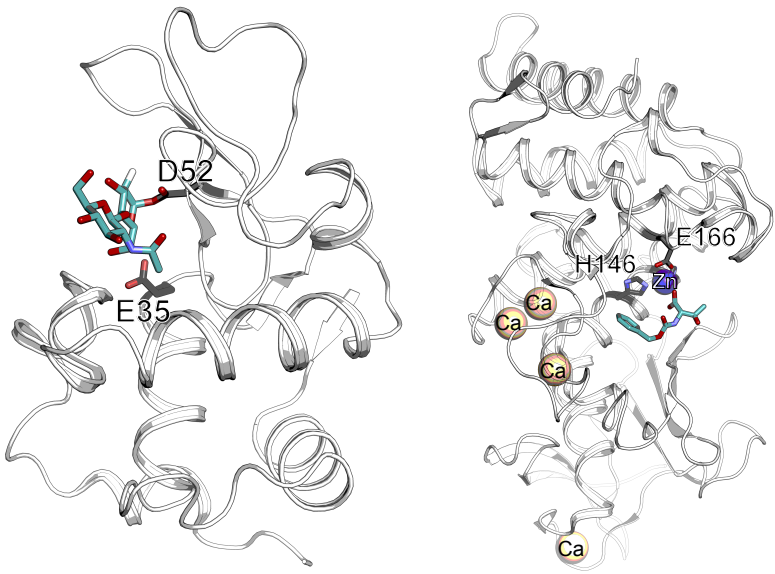

Figure 1.4: Structure of lysozyme (left) with active site residues E35 and D52 colored in dark grey with a covalently bound substrate (cyan) to the active site residue D52. Figure based on structures by Vocadlo [93], PDBID 1H6M and Cheetam [94], PDBID 1HEW. Structure of thermolysin(right) with active site zink (blue) coordinated by residues H146, H142 and E166 with a peptide inhibitor bound to the active site (cyan) and calcium in yellow. Structure from Senda *et. al*, PDBID 1KJO.

all throughout the animal and plant kingdoms. Lysozyme hydrolyses the glycosidic bond between the polysaccharide components, N-acetylmuramic acid and N-acetylglucosamine, in the bacterial cell wall. The hydrolysis leads to cell wall degradation and subsequent cell lysis. Lysozyme can thereby defend against gram-positive bacteria as well as degrade bacteria killed by other defence mechanisms. In 1965, lysozyme became the first enzyme with a determined structure [96]. The structure of lysozyme has an ellipsoidal shape with a large substrate binding cleft that can accommodate the long polysaccharide substrate and coordinates the two catalytic residues Glu35 and Asp52 (Figure 1.4). The structure contain five α-helices and a few β-sheets. The most well studied lysozyme is Hen Egg White Lysozyme (HEWL) that can be purified in large amounts directly from egg white (one egg contains around 5 g lysozyme).

The first terahertz absorption spectrum (3-6 THz) of a protein was measured on dry lysozyme samples already in 1971 [97] and 20 years later measured as a function of hydration for lower frequencies in the terahertz region (0.45-1.3 THz) [98]. These first pioneering absorption spectra showed a broad featureless absorption profile with a large water absorption background. More recent studies show that both dry HEWL [41] and HEWL in solution [44] with the water background removed give rise to smooth spectra with increasing absorption for higher frequencies. The measured absorption spectra also correspond well to calculated normal mode densities [19, 41] with a frequency cut off around 0.2-0.3 THz [44]. It is also interesting to note that the normal mode density for folded and partially unfolded lysozyme is only different for frequencies below 0.45 THz [22].

Early NMD and MD calculations of lysozyme revealed a hinge bending motion that opens and closes the substrate binding cleft [16,

18,99,100]. The free enzyme hinge bending mode has a calculated frequency around 0.09 THz while in the inhibitor bound enzyme the frequency is red shifted to 0.13 THz [100]. A recent study shows that the absorption spectrum above 0.75 THz blue shifts upon inhibitor binding which according to NMD simulations is related to the hinge bending motion [53]. Optical Kerr Effect spectroscopy has also successfully identified two strong vibrational bands at 1.15 THz and 2.80 THz that blue shift upon inhibitor binding to 1.29 THz and 2.89 THz respectively [101]. Using Raman spectroscopy, vibrational bands between 0.6-3 THz has been identified for lysozyme but without determining any biochemical relevance [102–104]. These experiments performed on lysozyme show that the protein has collective vibrational modes in the terahertz frequency region. Some of the recent research manages to relate observed vibrational modes to functionally relevant movements but the exact role of collective dynamics in lysozyme function is still not known. Both NMD and MD simulations show functionally relevant vibrations that with continuous efforts from experimentalists may be identified.

Thermolysin

Thermolysin is a thermostable extracellular metalloendopeptidase from the gram-positive bacteria *Bacillus stearothermophilus* that hydrolyses peptide bonds on the N-terminal side of hydrophobic amino acid residues. Thermolysin has five cofactors, one zinc ion important for catalytic activity and four calcium ions important for stability [105]. The three dimensional structure of thermolysin was determined by Matthews and co-workers in 1972 [106]. Thermolysin has a C-terminal domain mainly composed of α-helices and an N-terminal domain mainly composed of β-sheets that are connected by a central

helix. The active site coordinating the zink ion is situated in the cleft between the two domains (Figure 1.4).

Thermolysin undergoes a hinge bending motion (similar to lysozyme) that opens and closes the active site according to structural studies [107, 108]. The hinge bending motion in thermolysin has also been verified by MD simulations [109]. Thermolysin has not been as extensively studied as lysozyme when it comes to low frequency vibrations but the presence of a hinge bending motion makes thermolysin an interesting candidate for research on the relation between tier 1 dynamics and its enzymatic function.

1.7 Scope of this thesis

In this thesis, terahertz radiation has been used to examine the presence and location of low frequency vibrations in three different proteins, RC_{sph}, lysozyme and thermolysin. The aim was to both detect tier 1 dynamics shown to be important for protein function as well as to develop experimental and analytical techniques for doing so.

In **Paper I**, the low frequency dynamics of the membrane model protein, RC_{sph}, was investigated under extended illumination. The dynamics were probed using synchrotron based terahertz Fourier transform infrared spectroscopy (FTIR). A non-thermal increase in vibrational mode density upon photo activation was observed and the effect was found to arise mainly from the L and M subunits of RC_{sph} coordinating the cofactors.

In **Paper II** and **Paper III**, terahertz was instead used as a pump for exciting low frequency vibrations in lysozyme and thermolysin. The terahertz excitation was visualized at atomic resolution using X-ray crystallography employing a new data collection strategy.

In **Paper II**, the non-thermal excitation of a low frequency mode in lysozyme was found to affect a central α-helix in the protein. The excited state was exceptionally long lived which indicates the presence of a Fröhlich condensate not previously detected experimentally despite the five decades that has passed since its formulation.

In **Paper III** the method was extended to investigate how terahertz radiation changes Bragg peak intensities in the diffraction pattern of thermolysin. A Bayesian statistical method was developed to detect the intensity changes also to estimate structure factor amplitudes for map calculation from the distribution of non-merged reflections. The sensitive statistical method revealed the presence of tier 1 dynamics in thermolysin that could be detected directly from the diffraction pattern and could be visualized mainly as the movement of two residues in the structure.

Chapter 2

Methodology

2.1 Production and purification of RC_{sph}

Due to its high abundance in the photosynthetic membranes, RC_{sph} can be extracted directly from *R. sphaeroides* without any need for recombinant expression. *R. sphaeroides* can grow under anaerobic conditions by phototrophy and under aerobic conditions by chemoheterotrophy. By growing the bacterial culture first in dark conditions, oxygen will decrease which in turn induces the production of chromatophores. Chromatophores are pseudo-organelles formed as an bulb-like extension of the cytoplasmic membrane. After the dark aerobic phase, the culture is grown in light anaerobic conditions whereby the reaction centre is expressed and populates the chromatophores. By optimizing the growth time under dark and light conditions the reaction centre yield is increased. [110]

Reaction centre is a membrane protein and the first step in the protein purification is thus to isolate the photosynthetic membranes. The *R. sphaeroides* cells can be disrupted either by liquid sheer pres-

sure (e.g. French press) or ultra sonication and the membranes containing the photosynthetic unit are isolated by ultra centrifugation. The protein then needs to be solubilized in order to produce a pure enough membrane protein sample in an aqueous media suitable for further analysis or crystallization. Solubilization is the most critical step in membrane protein purification since the native membrane environment is exchanged with an unnatural detergent environment and at the same time the protein needs to be kept in its native and active state. RC_{sph} solubilization is performed in the dark by careful addition of the detergent lauryldimethylamine-N-oxide (LDAO). After solubilization RC_{sph} is separated from other solubilized proteins by cellite column chromatography and subsequent ion exchange chromatography [110]. The purity of the RC_{sph} sample is determined as the fraction between the absorption at 280 nm and 800 nm. Fractions with $A_{280nm}/A_{800nm} > 1.25$ were used for the terahertz absorption spectroscopy in **Paper I**. The purified RC_{sph} can be concentrated to high concentrations (1.1 mM) without significant precipitation and can also be stored at $-80\,^{\circ}C$ without loss of activity.

The H-subunit can be removed from a purified RC_{sph} sample to give a photoactive reaction centre only composed of the L and M subunits (LM_{sph}) [64]. After incubation with the chaotrophic agent $LiClO_4$, the H-subunit precipitates and dissociates from LM_{sph}. LM_{sph} and the H-subunit is then easily separated by centrifugation.

2.2 Lipidic sponge phase

Keeping membrane proteins stable in their native functional state over time is more problematic than for soluble proteins. The unnatural detergent environment introduced when solubilizing membrane

proteins is one of the causes for instability. In order to provide a more native like environment for membrane proteins the Lipidic Cubic Phase (LCP) system was developed for crystallization applications in 1996 [111]. By mixing the lipid Monoolein (MO) with water, several lipidic phases can be formed depending on the water/MO ratio [112]. LCP is formed at a 20-40 % water content. Lipidic Sponge Phase (LSP) can be formed from LCP by the addition of an additional solvent, for example dimethyl sulfoxide (DMSO), polyethylene glycol (PEG) or Jeffamine M600 (as used in **Paper I**) [113]. Depending on the solvent and the addition of further additives, the water content of the sponge phase can be varied [114]. LCP is a rather stiff liquid crystal with small water pores and highly curved lipid bilayers while LSP is instead a liquid that has larger water pores and less curved bilayers. The larger water pores of LSP makes it suitable for membrane proteins with larger water soluble domains and the lower viscosity makes it easier to work with compared to LCP. RC_{sph} has been succesfully crystallized in both LCP [65] and LSP [115]. In **Paper I**, LSP is mixed with RC_{sph} in order to investigate the terahertz absorption properties of the protein in a more native like environment.

2.3 Terahertz technology

There are four main problems with using terahertz radiation in protein research:

1. **Production and detection of terahertz radiation,**
   Practical and stable devices for both production and detection of terahertz radiation is lacking which has hindered its use for studying collective vibrations in proteins. This is due to the so called ”terahertz gap” in the electromagnetic spectrum. The

terahertz gap exists because neither the instrumentation used for the microwave nor the infrared frequency regions is compatible with the terahertz frequency range. The field of terahertz technology is evolving fast at the moment [116] but the problem persists to produce a terahertz beam with a high enough power to be comparable to what can be produced in other frequency regions.

2. **Featureless absorption spectrum,**
   Absorption spectra of both dehydrated and hydrated protein samples are featureless in the terahertz region. This is likely caused by a combination of high mode density at terahertz frequencies, presence of multiple conformations in the sample and variability in the protein environment. [43]

3. **Water absorption,**
   Water has the property of forming hydrogen bonding networks with neighbouring molecules. The extensive dynamic hydrogen bonding network behaves in a collective manner and absorbs radiation in the terahertz region. Since the native environment for proteins is aqueous, the strong water absorption makes it difficult to separate protein absorption from background water absorption.

4. **Absorption is temperature dependent,**
   The terahertz absorption by water increases with temperature [117] which complicates the study of terahertz absorption spectra. A temperature effect must always be evaluated when working with terahertz radiation and the experimental setup should be designed to minimize heating.

Terahertz absorption spectroscopy

Terahertz absorption spectra can be measured either by FTIR or Terahertz Time Domain Spectroscopy (THz-TDS). The absorption measurements in **Paper I** were measured with FTIR at the SISSI beamline of the Elettra synchrotron [118]. By the use of synchrotron terahertz radiation, it is possible to perform FTIR experiments that on conventional lab sources are hampered by the lack of broad band sources with high enough energy [119]. Terahertz synchrotron radiation is also highly collimated as opposed to other sources which makes it possible to measure on small samples. FTIR differs from dispersive spectrophotometers by collecting data over a wide spectral range simultaneously instead of measuring one wavelength per time. FTIR has several advantages over dispersive techniques, for example better signal to noise and faster spectral acquisition time.

In FTIR absorption spectroscopy, the detector records the transmission of the radiation through the sample. The transmission through the sample has to be related to the transmission through the surroundings in order to deduce the absorption by the molecule under study. For a protein solution, this is done by measuring the transmission spectrum of the sample, I_{sample}, and the buffer, I_{buffer}, separately and calculate the absorption by the protein, A, according to Lambert-Beers law:

$$A = log \frac{I_{\text{buffer}}}{I_{\text{sample}}} \tag{2.1}$$

The protein terahertz absorption will be negative since the protein sample has a lower water concentration than the buffer solution and water absorbs more strongly than the protein at terahertz frequencies. To acquire an absolute protein absorption spectrum the water

concentration has to be known and accounted for when calculating the absorption.

Terahertz radiation as a pump

In **Paper II** and **III** terahertz radiation was used to excite low frequency vibrations in proteins by irradiating protein crystals with a terahertz beam. The terahertz source used in both papers can deliver 195.2 GHz and 390.4 GHz radiation at total power of 80 mW and 12 mW, respectively. The source can be pulsed by applying 0 V(terahertz on) or 5 V (terahertz off) through a TTL port. An external pulse can thereby synchronize the terahertz pulse with for example an X-ray detector readout as was done in **Paper II** and **III**.

Diffraction effects heavily influence the terahertz beam profile since the wavelength of terahertz radiation ($\lambda = 7.5$ mm for 0.4 THz) has a comparable size to the optical components of terahertz devices. When the terahertz beam is exiting from the horn antenna, the beam will get diffracted at the antenna edge. This causes the terahertz beam to spread out and the power attenuates fast with distance. Therefore, care needs to be taken to align the center of the beam on the sample as well as to minimize the distance between the antenna and the sample in the experimental setup.

2.4 X-ray crystallography

Protein crystallization

A protein first has to be crystallized before its structure can be determined by X-ray crystallography. Production of high quality crystals is crucial since the quality of the crystal ultimately determines the

quality of the collected diffraction data. In a crystal the proteins are packed in a repetitive manner in three dimensions and are held together by non-covalent interactions. The crystal can be described in terms of its space group, the asymmetric unit and the unit cell. The unit cell is the smallest unit that through translation can describe the full crystal. The unit cell in turn is composed of the asymmetric unit repeated according to the symmetry operators of the space group.

A protein crystal is formed by gently forcing the protein out from solution to a crystalline state. A supersaturated solution is formed by adding precipitants and/or changing the water content of the protein solution so that the solubility limit of the protein is exceeded. The supersaturated solution is metastable and not in thermodynamic equilibrium. In the supersaturated state, a nucleation event (either spontaneous or triggered) will induce the excess protein molecules to move towards equilibrium from the solution phase to a protein rich phase that may be either precipitate or a protein crystal depending on the conditions. To successfully crystallize a protein is an iterative process where the right conditions have to be found. Protein purity and concentration, choice of precipitate solution and concentration, temperature and pH are the most important parameters that affect crystal formation.

The most common crystallization method is vapour diffusion where a drop of protein solution mixed with precipitant solution is sealed in a chamber together with precipitant containing reservoir solution. The reservoir will through vapour diffusion reduce the water concentration in the protein drop thereby moving it into a supersaturated state from which a crystal can form.

Diffraction theory

When X-rays are impinging on an atom its electrons are set into motion creating an oscillating dipole. In the case of elastic scattering, the oscillating electrons in turn are acting as secondary sources that re-emit X-ray photons with an energy equal to the incoming X-ray photon in all directions. When electrons are ordered in a crystal, the periodicity of the crystal makes the scattered waves from all electrons to interact with positive or negative interference to give rise to a diffraction pattern. In order to observe diffraction, the radiation wavelength (λ) has to be similar to the spacing between the scattering objects. X-rays (λ around 1 Å) are therefore used for determining the atomic structure of a molecule, with atomic spacing around 1 Å.

Lattice points in parallel planes of a crystal act like a mirror in the sense that the angle between the incoming beam and the plane is equal to the angle between the diffracted beam and the plane. Diffraction from a crystal can thus be approximated as reflections from different sets of crystal planes, this is why sometimes diffraction spots are called reflections. This approximation is the core of the well known Bragg equation:

$$n\lambda = 2d_{hkl} \cdot \sin\theta \tag{2.2}$$

which states that for positive interference to occur the product of the distance between the planes, d_{hkl}, and the sinus of the angle between the incoming radiation and the plane has to be equal to any integer, n, times the wavelength. As can be seen from the equation, diffraction only occurs at a small set of angles. If the unit cell is large, as in the case of protein crystals, several Bragg planes are in diffracting position at every angle, θ. A Bragg peak, (h, k, l), corresponds to diffraction

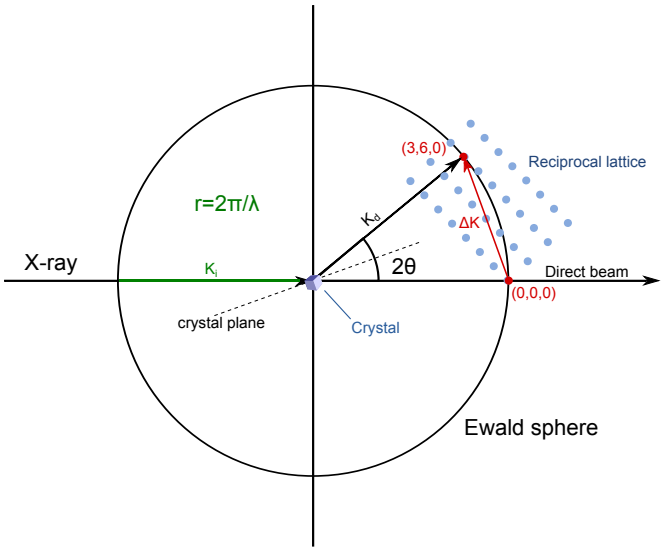

Figure 2.1: The Ewald construction, when a reciprocal lattice point crosses the Ewald sphere diffraction occurs. Rotation of the crystal rotates the reciprocal lattice, hence more reciprocal lattice points are brought into a diffracting position. The red arrow is the scattering vector, δK, which is the difference between the incoming wave vector, K_i with length $= 2\pi/\lambda$, and the diffracted wave vector, K_d. For elastic scatteing $K_i = K_d = 2\pi/\lambda$ and the scattering vector thus lies on a sphere with radius $2\pi/\lambda$.

from one set of crystal planes defined by Miller indices hkl and the intensity of the spot is proportional to the number of electrons in that specific crystal plane. The diffraction from a crystal can be described as a reciprocal lattice which has the same Laue symmetry as the real space lattice.

The relation between reciprocal space and direct space and the diffraction criterion is explained in the so called Ewald construction (Figure 2.1). The reciprocal lattice is described by the 3 lattice vectors

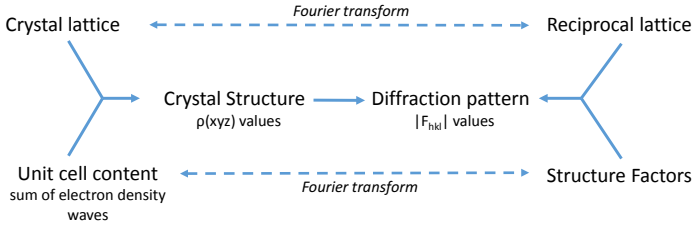

Figure 2.2: Relationship between crystal structure and the measured diffraction pattern.

a^*, b^* and c^* which are related to the unit cell vectors a,b and c according to equation 2.3.

$$a^* = 2\pi\frac{b \times c}{V},\ b^* = 2\pi\frac{c \times a}{V},\ c^* = 2\pi\frac{a \times b}{V} \tag{2.3}$$

If the reciprocal lattice is known the crystal lattice can be calculated. Direct space and reciprocal space are both periodic and they can be described by a Fourier series, furthermore they are reciprocally related and can be inter converted through a Fourier transform. The electron density, $\rho(xyz)$, can be calculated through Fourier synthesis:

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l F(hkl, xyz) \tag{2.4}$$

$$F_{hkl} = |F_{hkl}|e^{i\alpha_{hkl}} = \sum_j f_j e^{-2\pi i(hx_j+ky_j+lz_j)} \tag{2.5}$$

where F_{hkl} is the structure factor, the Fourier transform of the unit cell content (the sum of the scattering contribution of all atoms j) sampled at reciprocal lattice point hkl. f_j is the atomic structure factor and x, y, z the atomic coordinates. The relation between the electron density and the diffraction pattern is shown in Figure 2.2.

The phase problem

The structure factors are complex numbers, and unfortunately their phase is not easily accessible experimentally since in a diffraction experiment the intensity is measured. The intensity is proportional to the square of the amplitude of the structure factor, $|F(hkl)|^2$, thus information about the phase is lost. Since the phase cannot be retrieved directly from an experiment it gives rise to the so called phase problem. The phase problem can be solved through experimental techniques where the structure factor phase and amplitude is changed by heavy atom incorporation either through a dominating total scattering from the higher electron content of a heavy atom (isomorphous replacement) or through an anomalous signal from measuring diffraction close to the absorption edge of the heavy atom (anomalous dispersion). Isomorphous replacement needs diffraction data collected from several crystals, both native and with heavy atoms incorporated. The difficulty with this technique is to produce isomorphous crystals. Changes in unit cell dimensions are not tolerated and the molecule has to be only locally affected by the heavy atom. In anomalous dispersion, isomorphism is not an issue since data can be collected from one crystal only, instead radiation damage becomes a major problem since diffraction data has to be collected at several wavelengths on the same crystal.

If possible, molecular replacement is the most popular method used for solving the phase problem since available experimental techniques are both time consuming and difficult to perform. In molecular replacement the phase is acquired through a model calculated from a homologous structure with at least 30% sequence identity. To build an atomic model with the correct phase the homologous structure has to be positioned correctly in the unit cell, thus the problem

can be divided into finding 3 translational operators and 3 rotational operators. There are several computer programs that can perform molecular replacement either based on the Patterson method (Molrep [120], AMoRe [121]) or based on maximum likelihood methods (Phaser [122]) which has proven to be better at discerning correct solutions from noise. Phaser, the molecular replacement program used throughout this thesis, is an automated program based on maximum likelihood probability theory.

Data collection

X-ray diffraction data from protein crystals are mainly collected at synchrotrons where electromagnetic radiation is produced by the acceleration of electrons at relativistic speed through a magnetic field. There are several properties of synchrotron radiation that makes it superior over conventional X-ray sources. The properties important for X-ray crystallography are mainly the high flux and brilliance. The large number of photons make it possible to focus the beam in the micrometre range and use monochromators while still having a brilliance several orders of magnitudes higher than conventional lab sources.

Protein crystals suffer from X-ray radiation damage caused mainly by ionization by high energy electrons produced by photoelectric absorption or inelastic scattering and mechanical stress caused by lattice disruption. Radiation damage causes the crystal to loose diffraction power with time which affects the quality of the data. The high flux of synchrotron radiation shortens the total acquisition time and is thus limiting the time dependent radiation damage effects. By cryo cooling the crystal to 100 K diffusion of free radicals through the crystal can be stopped and also the low temperature stabilizes the lattice. [123]

To be able to accurately construct an electron density map from

measured diffraction spots a large part of the reciprocal lattice has to be sampled. Since a rotation in real space results in a rotation in reciprocal space (Equation 2.3) the crystal is rotated in the X-ray beam while diffraction images are collected. The crystal is mounted in a nylon loop and the loop is then positioned on a goniostat. For data collected at room temperature the crystal is shielded from drying out by a capillary with a small drop of liquid in the top sealed of with vacuum grease. The typical goniostat used for protein crystallography is a mini-kappa goniometer which has three rotation axes, Ω, K and Φ. During data collection the crystal is aligned by two of the axes (Ω and K or Φ and K) so that rotation is only around the third axis (Φ or Ω). The total rotation needed for a complete dataset depends on the symmetry and orientation of the crystal.

In a normal experiment the crystal is oscillated within a range of 0.1-2° per frame. A real crystal is mosaic which means that the crystal planes are slightly misoriented in relation to each other, this affects the angular spread of the Bragg peaks. When the oscillation is lower than the mosaicity of a Bragg peak it will be collected over several frames, so called fine slicing, which improves the peak profile reconstruction in three dimensions.

X-ray detectors used for protein crystallography is either Charge Coupled Device (CCD) detectors or hybrid pixel detectors (e.g. Pilatus detector). Pilatus detectors are gaining popularity over the more traditional CCD detectors due to the very fast readout time (in ms), 1 pixel point spread function and the lack of noise. For a fine slicing data collection strategy a Pilatus detector is necessary since such a data collection with a CCD would take too much time.

Data processing

All diffraction images recorded is converted into a list containing Miller index and integrated intensity of all reflections through data processing. XDS [124] is an automated software for processing that can be called through pipeline scripts for processing large amount of datasets given one standard input file. In the first step of processing detector related corrections are defined and pixels containing signal is discerned from noise. Based on the strongest reflections in the data the direction and parameters of the crystal unit cell is found and crystal symmetry is suggested. Geometric parameters are refined and agreement between all possible Bravais lattices are reported. Diffraction images are then masked depending on user specified high resolution cut off and shadows on the detector caused by intruding hardware, for example the spindle axis and beam stop.

All diffraction spots are then integrated, spot profiles are determined on the grid constructed from the strong reflections from earlier processing steps. The output from the integration step is a list of all detected indexed spots, hkl, and their integrated intensity together with their standard deviation. Correction factors are applied to the intensities and the quality and completeness of the data is reported. The final step of processing is scaling and merging where data from several crystals can be merged together and put on the same scale.

After processing the results are evaluated mainly according to the data **completeness** (percentage of possible reflections recorded) , **R-merge** (spread in intensity of symmetry related reflections(Equation 2.6)), $\langle I/I(\sigma)\rangle$ (signal to noise, the mean of the ratio between all intensities and their associated standard error), **CC(1/2)** [125](the percentage of Pearson correlation between random half datasets) and the redundancy (total number of recorded reflections divided by the

number of unique reflections).

$$\mathrm{R}_{\mathrm{merge}} = \frac{\sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_i |I_{hkl,i}|} \tag{2.6}$$

where $I_{hkl,i}$ is the ith intensity measurement of a reflection and $\langle I_{hkl} \rangle$ the average intensity from multiple reflections.

Refinement and model building

After processing and initial phasing through molecular replacement the structural model is far from perfect and the phases and atomic coordinates are improved throughout several iterative steps of model building and refinement. Refinement is carried out through simulated annealing or other minimization algorithms that minimizes a maximum likelihood target in reciprocal space. The crystallographic R-factor (Equation 2.7) is reported after every refinement cycle to assess the progress of the refinement.

$$R = \frac{\sum ||F_{\mathrm{obs}}| - |F_{\mathrm{calc}}||}{\sum |F_{\mathrm{obs}}|} \tag{2.7}$$

The R factor represents the difference between observed ($|F_{\mathrm{obs}}|$) and calculated ($|F_{\mathrm{calc}}|$) structure factor amplitudes and should thus decrease for models that better fit the data. For cross validation usually 5% of the data is left out from refinement to calculate R_{free}. Several restraints are put on the molecule, for rigid body refinement the whole molecule or defined parts of it is treated like a rigid body that is translated and rotated in the unit cell. Restrained refinement instead lets the atoms in the molecule to move freely within a pre set range of restraints, for example on bond lengths and bond angles. The automated refinement in reciprocal space is accompanied

by model building and refinement in real space. In model building the model is changed to better fit the electron density, for example by mutating residues, adding waters or adding alternative conformations of a residue. Different conformations of an amino acid is described by its alternative coordinates together with the occupancy which is the fraction of a conformational state present in the structure. Model building can be both automated or be done by hand using graphic software such as Coot [126]. In real space refinement the model coordinates are modified to better fit the electron density. After model building and real space refinement, another round of refinement in reciprocal space is performed which will hopefully yield improved phases and a better electron density map. With a better map, the model can be further improved. The cycle of refinement and model building proceeds until no obvious improvements of the model can be made and model errors are as few as possible.

Every atom in a structural model is at least described by its coordinates in the unit cell and additionally a B-factor, all four parameters are refined. The B-factor (also called temperature factor and Debye-Waller factor) is part of the atomic scattering factor f, and describes the disorder or relative vibrational motion of the atom.

$$f = f_0 e^{-B(\sin^2\theta/\lambda^2)} \tag{2.8}$$

$$B = 8\pi^2\mu^2 \tag{2.9}$$

where f is the corrected atomic structure factor, f_0 the structure factor for a given atom at 0 K, θ is the scattering angle, λ the X-ray wavelength and μ^2 is the mean square displacement of the atom. The B-factor is an isotropic model of motion described as a sphere with its radius as the only parameter, a spherical model assumes atomic

motion to be equal in all directions. A more valid description would be an anisotropic B-factor described as an ellipsoid with six parameters but that is less used since too many parameters may lead to over fitting.

There is a fall off in intensity as a function of $\sin\theta/\lambda$ for the atomic structure factor because of the size of the electron cloud of an atom is of comparable size to the incoming X-ray wavelength, this gives rise to a path difference between waves scattered by different parts of the electron cloud which in turn produces some negative interference, the larger the atom the more fall off in intensity. The higher order Bragg peaks are because of this fall off more difficult to detect and needs positive interference from more crystal planes. For a crystal to yield high resolution data it thus needs to be well ordered (low mosaicity) and have a large volume to increase the path length for the X-ray beam through the crystal. The choice of wavelength also affects the resolution, the smaller the wavelength the larger the Ewald sphere radius and a bigger fraction of the reciprocal lattice becomes available for measurement (Figure 2.1). The detector distance also affects how high resolution data can be collected and a short distance is needed when detecting high resolution data. Both short detector distances and short X-ray wavelengths results in small peak separation which may result in overlapping Bragg peaks. Therefore, both the wavelength and detector distance should be chosen depending on the diffraction power of the crystal to ensure highest possible resolution and peak separation.

The model is validated in the end of refinement according several methods. One validation approach is to study the Ramachandran plot which is a 2D plot with the Ψ and Φ angles of the protein backbone on each axis. Certain amino acids are found to cluster in specific re-

gions of the Ramachandran plot determined by their steric hindrance. Amino acids outside these allowed regions are probably due to errors in the model. Other methods validate the model according to for example rotamers, standard bond lengths and angles and B-factor distributions.

Electron density maps

There are several different kinds of electron density maps. The electron density maps are named after the structure factor amplitudes used in the Fourier transform, the phase is coming from the calculated structure factor for all maps mentioned here. The structural model is built into a 2Fo-Fc map (structure factor amplitude = $2|F_{\text{obs}}| - |F_{\text{calc}}|$) map since it is less model biased than a Fo map. Furthermore a Fo-Fc map (structure factor amplitude=$|F_{\text{obs}}| - |F_{\text{calc}}|$) is also calculated which yields positive difference peaks in the map where the model fails to account for some electron density and a negative difference peak where the model gives rise to features that are not real. The Fo-Fc map is used to guide the real space model building. Fo-Fo maps are calculated to display a difference between two experimental datasets, in this thesis used to display the difference in electron density between terahertz irradiated and non-irradiated protein crystals. What is important to remember is that all electron density maps are biased towards the model since the phase is always retrieved from the model and never directly from the experimental data.

The electron density is a continuous function but for the sake of reducing computing power the function is solved on a grid in the unit cell. The grid spacing need to be at least $d_{\text{min}}/2$ to preserve the information in the Fourier transform. For visualization the map is contoured at set electron density values much like the contouring of

a map where different lines correspond to a certain height. Instead of absolute electron density, maps are in practice contoured at levels expressed as number of sigma away from the mean electron density, sigma is the square root of the density variance. The lower the sigma level of a map the more details and noise are shown.

2.5 Bayesian statistics

In science, nothing is ever known with absolute certainty. Instead, statistical methods are employed to evaluate the confidence in our data and models. Bayesian inference and the related maximum likelihood methods, as opposed to the more well known frequentist inference, are applied to many parts of X-ray crystallography, from data processing to refinement and model building [127–132]. The strength of a Bayesian approach is the inclusion of prior beliefs in the analysis. To determine the confidence of for example our structural model to be true given our data, prior beliefs can be incorporated in the form of stereo chemical plausibility for a protein molecule. By doing so, we increase the chance of producing a model that is describing the data well and at the same time is plausible according to prior knowledge.

Bayesian inference follows Bayes theorem stated already in 1763 [133]:

$$P(Model|Data, I) = \frac{P(Data|Model, I)P(Model|I)}{P(Data|I)} \tag{2.10}$$

For parameter estimation, $P(Data|I)$ is a proportionality constant and the theorem can be simplified to:

$$P(Model|Data, I) \propto P(Data, I|Model)P(Model|I) \tag{2.11}$$

which states that the probability of the model to be true given our

data and conditional information, I, equals the probability of the data to be measured given the model times the probability of the model to be true given conditional information without considering the data. The result gained from Bayesian statistical methods is the posterior distribution, $P(Model|Data, I)$, given the prior distribution $P(Model|I)$ and the data likelihood function, $P(Data, I|Model)$. The posterior distribution describes the probability of values for all parameters in the model and will thus have one dimension per parameter. The certainty of the values of one parameter can be evaluated with a credible interval. For example, a 95 % Highest Density Interval (HDI) credible interval states that we are 95 % certain that the parameter has a value between x_1 and x_2. The credible interval can also give the certainty of differences between two groups. If the 95 % credible interval for the estimated difference has strictly positive or negative values for x_1 and x_2 we are 97.5 % certain that there is a non-zero difference. In addition, a credible interval also gives information on the difference magnitude and precision which is information that is not gained from the more commonly used confidence intervals.

A Bayesian approach for statistical group comparison is superior to null hypothesis t-tests. The p-value in null hypothesis testing is dependent on a hypothetical distribution of possible outcomes from an experiment which in turn is dependent on the sampling intentions. By increasing the number of samplings, the null hypothesis can always be rejected since it is never literally correct. A Bayesian approach on the other hand gives results only dependent on the data actually sampled and is not dependent on the sampling intentions. In a Bayesian approach, any sample size can be used and larger sample sizes will generally decrease the posterior distribution width as we become more certain of the parameter value [134, 135].

Bayesian methods also allow for the use of complex models. The Bayesian model can be described by any number of parameters but increasing number of parameters will increase the difficulty of finding the combination of parameters giving the maximum of the posterior distribution. Bayesian methods are both mathematically and computationally intensive which has limited its use in science, especially among researchers without a strong background in statistics. More software is becoming available mainly based on numerical algorithms (e.g. Markov Chain Monte Carlo simulations) that may pave the way for Bayesian methods to be more frequently used.

Chapter 3

Terahertz absorption spectroscopy of RC_{sph}

Given the probably complex energy landscape and the importance of dynamics for the electron transfer reaction in RC_{sph} we investigated the difference in vibrational density in the terahertz region of RC_{sph} upon illumination.

RC_{sph} is an excellent test system for studies of membrane protein dynamics for several reasons: for a membrane protein it is relatively easy to produce and purify at high yields, it is well characterized and due to its photosynthetic activity it can be reversibly cycled between activated and ground state by the use of a laser. The reaction center studied here is a mutant called R26 which lacks carotenoids. The lack of carotenoids simplifies the absorption spectrum in the UV/Vis region which is used during purification and to follow light activation. Figure 3.1 shows an absorption spectrum with the three typical absorption peaks for RC_{sph} arising from the cofactors BPhe, accessory BChl and P_{870}. The absorption peak of P_{870} bleaches upon illumina-

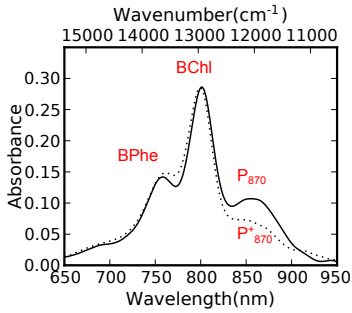

Figure 3.1: Optical absorption spectra of 0.62 mM RC_{sph} solution in ground state (solid line) and illuminated by a 532 nm laser with a power of 7.7 kW m^{-2}(dotted line). The cofactors BPhe, accessory BChl and P_{870} give rise to the three absorption peaks typical for RC_{sph}.

tion and relaxes back to the ground state in the dark [61]. From the peak height at 855 nm for illuminated and non illuminated sample we determined a fraction of 33 % to be in the photo activated state when using a 532 nm laser with power of 7.7 kW m^{-2}.

3.1 Self referencing strategy

The exact water content in the background sample and protein sample has to be known in order to calculate the protein absorption. The water content is not easily determined for a RC_{sph} sample due to the unknown concentration of detergent and other rest components from the purification. Instead of analysing the absolute spectra, difference absorption spectra were calculated using a self referencing technique previously used for terahertz absorption studies on photoactive yellow protein and bacteriorhodopsin [42, 48]. By using a self referencing strategy, the background (RC_{sph} in dark state) and signal (RC_{sph}

in light excited state) spectra are recorded from the same sample, thereby circumventing the problem with unknown water concentrations. Spectra from the ground state and laser excited state of the protein were recorded alternately and the difference absorption was calculated for three consecutive spectra (dark-light-dark) according to Equation 3.1:

$$\Delta A^i = -\log\left(\frac{2I^i_{light}}{I^{i-1}_{dark} + I^{i+1}_{dark}}\right) \tag{3.1}$$

ΔA is the difference absorption, i denotes the spectrum number, I_{light} and I_{dark} is the measured transmittance intensity for illuminated and non illuminated sample, respectively. Calculating difference spectra from measurements made close in time minimizes artefacts in the spectrum due to drifts in the experimental station, such as fluctuating radiation intensity and sample degradation.

3.2 Experimental setup

The sample was sandwiched between two Z-cut quartz windows separated by a 25 µm teflon spacer and alternately illuminated with a 532 nm laser with a power of 7.7 $\mathrm{kW\,m^{-2}}$. The quartz windows had high transmittance in the terahertz region as well as having high thermal conductivity to ensure fast heat transfer from the sample. The quartz windows were in direct contact with the metal sample holder that was temperature controlled by a circulating thermal bath to further reduce heating. A drawing of the experimental setup is shown in Figure 3.2.

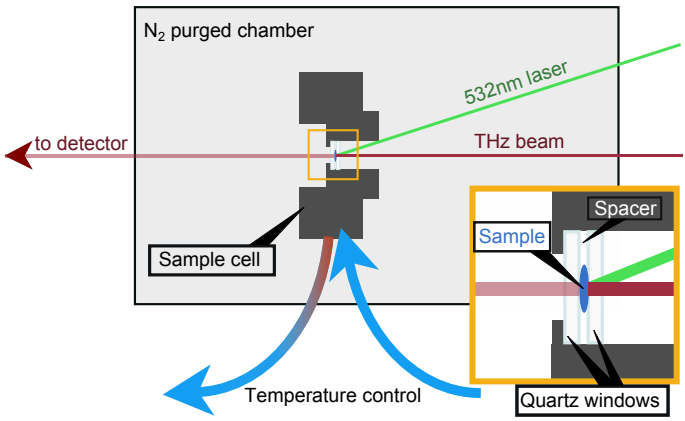

Figure 3.2: Illustration of the experimental setup. The sample cell was temperature controlled through a circulating water bath kept at constant temperature (blue arrows). The laser and synchrotron terahertz beam illuminated the sample that was kept in the sample cell between two quartz windows separated with a spacer (orange box). The experiment chamber was purged with N_2 to minimize atmospheric water absorption.

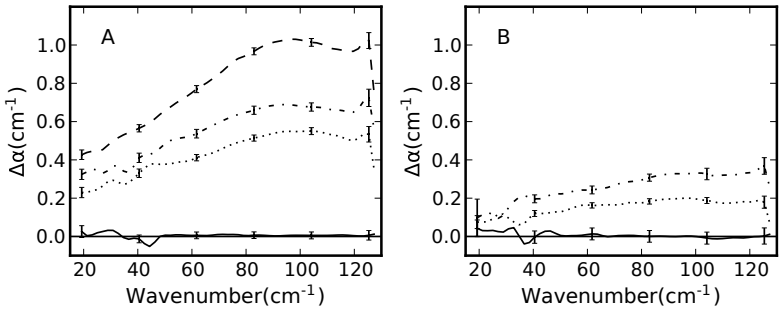

Figure 3.3: Difference in terahertz absorption ($\alpha_{\text{illuminated}} - \alpha_{\text{ground}}$) of RC$_{sph}$ and LM$_{sph}$ in detergent, **A**, and sponge phase, **B**. The error bars are representing the standard error $\sigma/\sqrt{N}$. **A.** 1.1 mM RC$_{sph}$ (dashed line), 0.76 mM RC$_{sph}$ (dash-dot line), 0.71 mM LM$_{sph}$ (dotted line) and detergent based buffer (solid line). **B.** Pure LSP spectrum (solid line), RC$_{sph}$ (0.76 mM) mixed 1:1 with LSP samples incubated for 14 days at room temperature (dotted line) and non incubated (dash-dot line).

3.3 Difference terahertz absorption of RC$_{sph}$

The difference absorption spectrum for RC$_{sph}$ between the illuminated and non illuminated sample shows an overall positive difference over the whole spectral region (20-125 cm^{-1}) with a maximum around 90 cm^{-1} (Figure 3.3A). Similarly, light activation of photoactive yellow protein has also been shown to produce an increase in terahertz absorption [48]. The difference magnitude increases with both higher protein concentration (Figure 3.3A) and laser power (not shown) which confirms the signal arising from the laser somehow interacting with the protein. The difference absorption spectrum is smooth and shows no distinct features alike previous terahertz absorption studies on proteins.

Absorption spectra were also measured of LM_{sph} to investigate the influence of the H-subunit on the difference signal. As can be seen in Figure 3.3A, the difference absorption profile from LM_{sph} has a very similar shape to RC_{sph}. LM_{sph} shows a stronger difference signal when the absorption spectra of both LM_{sph} and RC_{sph} are scaled according to their volumetric concentration. It can be deducted from these results that the difference absorption is originating from the LM subunit and that the H-subunit plays a minor role. It is not surprising that LM shows more vibrations upon illumination since they coordinate the cofactors were the electron transport is taking place. Low frequency vibrations can very well be involved in the electron transfer by structural stabilization and through increasing the heat transfer rate out from the structure. The H subunit does not add to the increase in vibrational density upon photo activation despite previously observed light induced structural changes [65]. On the other hand, structural changes related to tier 0 dynamics may still take place in the H-subunit even if no difference in tier 1 dynamics is detected. Furthermore, the H-subunit may have functionally important low frequency modes that do not change significantly upon light activation.

3.4 Effect of protein environment

A hydrated protein is not expected to behave in the same way when it is dehydrated due to the functional importance of solvent interactions. Integral membrane proteins, such as RC_{sph}, are interacting with the membrane, the cytoplasmic and periplasmic solvent in its native environment. This makes the interactions more complex than for soluble globular proteins. In order to investigate the effect of protein

environment on the vibrational spectrum we measured spectra from samples in detergent solution and in LSP. As can be seen in Figure 3.3A and B, both RC_{sph} and LM_{sph} in LSP and detergent buffer produce spectra with similar shapes. The smaller difference absorption for LSP samples can be explained by its lower protein concentration compared to the protein samples in detergent buffer. When normalizing the difference absorption spectra of protein in their different environments according to protein concentration they overlap well. Protein in two different solvents thus give rise to very similar difference absorption signals which indicates that the vibrations involved in electron transfer are located within the protein and do not extend significantly to the surroundings.

3.5 Ruling out a thermal effect

Absorption in the terahertz frequency region is highly dependent on the temperature of the sample. Since the RC_{sph} samples were illuminated with high laser power, heat will be generated. The question is how much this heating affects the absorption spectra. While the experimental setup design minimized heating, even a small temperature increase in the sample can affect the difference signal substantially. A water sample heating up 1 °C will increase the absorption at 33 cm^{-1} with 4 cm^{-1} according to Rønne *et al.* [117].

We measured terahertz absorption of RC_{sph} without laser excitation at two sample temperatures, 25 °C and 26 °C, to investigate how a change in sample temperature influence the terahertz absorption. The sample temperature was set by changing the circulating water bath temperature. It takes around 30 min to switch between two temperatures this way compared to the instantaneous switch pos-

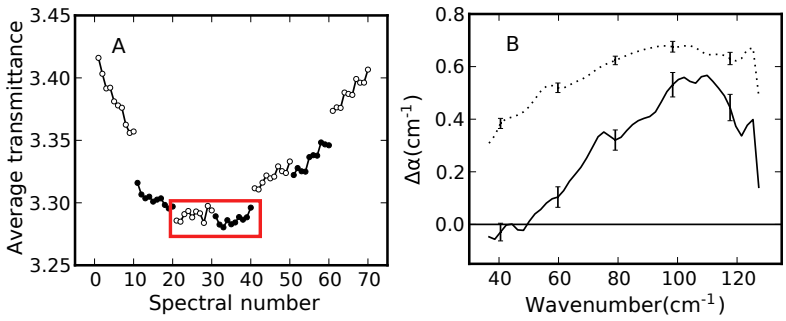

Figure 3.4: A. Average transmittance measured at 25 °C (open circles) and 26 °C (filled circles) plotted in the order they were collected (=spectrum number). The spectra used for the difference absorption calculation is highlighted in the red box. **B.** Difference absorption spectrum of 0.76 mM RC_{sph}, $\alpha_{25^{\circ}C} - \alpha_{26^{\circ}C}$, (solid line). The difference absorption spectrum for the 0.76 mM RC_{sph} sample from Figure 3.3A is shown for comparison (dotted line). The error bars are representing the standard error ($\sigma/\sqrt{N}$).

sible with a laser. As such, the drift in the terahertz beam intensity made it impossible to employ the self referencing strategy. A faster temperature switch could have been achieved through the use of an IR laser but then the sample temperature is instead difficult to determine. For these reasons we used an alternative strategy: five spectra were measured at each temperature and spectra collected during a period with stable transmission signal were used in the calculation of the difference absorption spectrum. Figure 3.4A shows how the average transmittance was changing over time during the collection of the temperature control data and the time range with a stable transmission signal.

As can be seen in Figure 3.4B, the 1 °C increase in sample temperature is not enough to produce the same difference absorption

acquired with laser excitation. The difference absorption maximum is around $100\,\mathrm{cm}^{-1}$ for both the temperature and laser excited protein spectra. On the other hand, the signal declines faster and is around zero for frequencies below $50\,\mathrm{cm}^{-1}$ when only changing the temperature compared to when the protein is excited with the green laser. This shape difference between the two spectra demonstrates that pure thermal fluctuations differs from functionally relevant low frequency vibrations upon photo activation of the protein.

According to our simulations, the temperature increase caused by the laser was at most $0.09\,^{\circ}\mathrm{C}$ when estimating the sample as water. Theoretically, combining the heat dependency from reference [117] and our simulated temperature increase would produce a terahertz absorption spectrum with a slightly smaller magnitude at $33\,\mathrm{cm}^{-1}$ compared to what was experimentally observed. On the other hand, the simulated temperature increase is regarding a worst case scenario and the real temperature increase is most likely lower. If heating was the sole reason for the difference signal, the difference in absorption should mainly depend on the water content of the sample. For RC_{sph} in LSP (25 % water) and detergent samples (62 % water) we do not find such a relationship. Taken together, the photo excitation of RC_{sph} is the main source of the difference absorption signal and heating of the sample represents only a small fraction of the spectral density difference.

3.6 Conclusions

We managed to achieve a sensitivity high enough to record a significant 0.005 % difference in absorption signal by using synchrotron terahertz radiation in combination with a self referencing strategy

for calculation of difference absorption spectra. The difference signal originated from the LM subunits containing the chromophores and was not caused by heating. The increase in terahertz absorption upon photo activation of RC_{sph} can be a result of either an increase in amplitude of the low frequency vibrations or by an increase in the number of normal modes in the spectral region caused by a red shift in the normal mode frequencies.

Chapter 4

Terahertz radiation induces structural changes in lysozyme

Instead of probing vibrational modes with a broad band terahertz source, a single wavelength terahertz pulse can be used to excite a few normal modes in the structure and the effect can be probed by other means. In **Paper II** we used terahertz radiation as a pump on a model protein and probed the induced structural change with X-ray crystallography. X-ray crystallography provide structural information with atomic resolution on how terahertz radiation interacts with a protein. X-ray crystallography also provides a contrast to the unordered water environment and only shows the ordered structural elements of the protein and bound waters. Lysozyme was chosen as a test protein for this novel technique since it display collective dynamics in the terahertz frequency region both according to models and experimental observations. Lysozyme is furthermore easily crystallized

and diffracts well to high resolution even at room temperature. The lowest frequency mode is believed to have the strongest link to protein function. Therefore, we chose to investigate the structural response to 0.2 THz and 0.4 THz irradiation which are close to the observed frequency cut off in lysozyme absorption spectra [44]. Furthermore, a 0.36 THz vibrational band, close to the 0.4 THz excitation frequency applied here, has been recently observed in lysozyme crystals [136].

4.1 Experimental setup and data collection

According to our simulations of the terahertz source and a protein crystal, 50 % of the 0.4 THz radiation is absorbed by the crystal already after a 50 µm penetration into the crystal. Diffraction data was collected near the terahertz illuminated surface of the crystal to ensure that the X-ray probed volume of the crystal was also receiving highest possible terahertz power. Since the crystal is rotating during X-ray diffraction data collection, a large crystal face was aligned orthogonal to the rotation axis and the terahertz source was illuminating that same face (Figure 4.1). This way, the terahertz field will be evenly distributed in the probed volume during data collection. Lysozyme crystallizes in the tetragonal space group $P4_32_12$ and has crystal faces as seen in Figure 4.2. Crystal alignment is done by first recording a few diffraction images over 5° and process the images to find the orientation of the unit cell axes. When the unit cell orientation is known, the goniometer Φ and K angles are chosen so that the Ω axis is orthogonal to one of the crystal faces. The crystal was mounted in a 90° bent loop which increased the chances finding a face on the crystal to align. The crystal was kept temperature controlled by an air flow with constant temperature (297 K). A plastic capillary

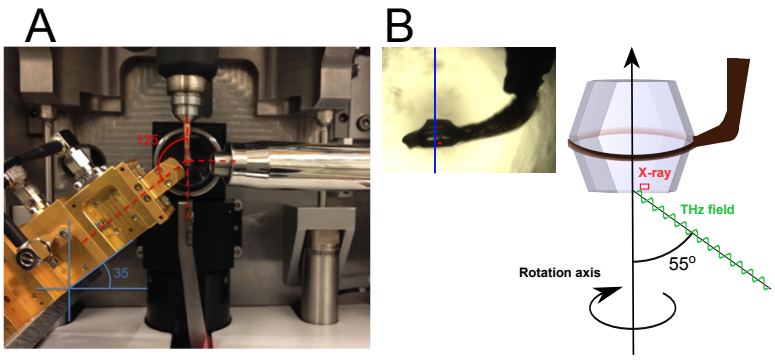

Figure 4.1: **A.** Image of the experimental setup showing the position of the terahertz source, a crystal mounted on the goniometer and the temperature control stream. **B.** Schematic illustration of a typical lysozyme crystal mounted in the loop, terahertz field direction, rotation axis and relative X-ray beam size. The inset shows an image of an aligned crystal from the experiment.

was mounted over the crystal to protect it from from drying out.

In order to see the terahertz radiation induced effect in the protein crystal, data collected from the terahertz illuminated protein needs to be compared to data collected from a non-illuminated protein. Simply comparing two datasets collected from two different crystals will show differences mainly depending on crystal differences and not on terahertz irradiation. Measuring two consecutive datasets from the same crystal will show differences mainly due to radiation damage. To maximize the chance of recording a difference due to terahertz irradiation, we employed a self referencing strategy (as in **Paper I**). The terahertz source was timed with the readout pulse from the Pilatus detector so that every second diffraction frame was collected from the crystal illuminated by 0.2 or 0.4 THz and every second frame from a non-illuminated crystal. Every frame was collected over 20 ms. The

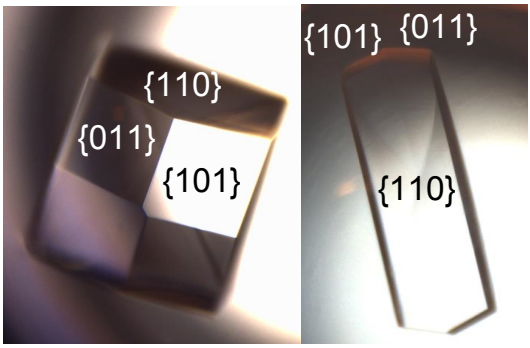

Figure 4.2: Lysozyme crystals in two different projections showing the different faces that could be oriented with the terahertz field.

data set was divided into two parts: one containing the terahertz illuminated data and one with the non-illuminated data. The data was collected with an rotation of only 0.01° for every frame (extreme fine slicing) which made it possible to get high quality data even when removing every second frame. As can be seen in Table 4.1 $CC_{1/2}$ and $<I/\sigma(I)>$ is unaffected when removing every second image compared to the processing of sequential data. R_{merge} increases slightly due to the higher redundancy of the data (collected over 360° instead of 180° to match the number of images).

Diffraction data were collected from crystals illuminated by both 0.2 THz and 0.4 THz. A reference data set was also collected without any illumination and the data was separated into odd and even frames (depending on the frame number). THz_{on} and THz_{off} frames as well as odd and even frames from the reference data set were processed separately. Diffraction data was collected from a large number of crystals and data with high quality was merged together. The 0.4 THz data is collected from 7 different crystals on 142 500 images for THz_{on} and THz_{off} each. The 0.2 THz data is collected from 6

crystals on 49 500 images for both THz_{on} and THz_{off}. The reference data is collected from 12 crystals on 155 400 images for both odd and even frames. Data processing statistics for the merged data is shown in Table 4.2.

Table 4.1: Processing statistics for the first half, second half and when removing every second frame for a 360° data set collected with the extreme fine slicing method.

	0-180°	180-360°	Every second (0-360°)
Resolution(Å)	15.0-1.7	15.0-1.7	15.0-1.7
R_{merge}(%)[a]	4.7(29.9)	5.7(47.6)	7.1(52.7)
$CC_{1/2}$ [a]	100.0(98.2)	99.9(96.5)	99.9(97.1)
$<I/\sigma(I)>$[a]	32.6(5.52)	26.6(3.26)	30.79(3.86)
Completeness(%)[a]	97.5(92.9)	97.1(92.0)	96.7(88.9)
Redundancy	12.4	12.4	24.3

[a] Values for the highest resolution shell (1.70-1.74) are shown in parentheses.

Table 4.2: Crystallographic table for merged data from crystals illuminated with 0.4 and 0.2 THz illuminated and reference data

	0.4 THz		**0.2 THz**		**Reference**	
	THz_{on}	THz_{off}	THz_{on}	THz_{off}	"Odd"	"Even"
Data collection						
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions						
a,b(Å)	79.35	79.35	79.32	79.32	79.24	79.24
c(Å)	37.42	37.42	37.51	37.51	37.5	37.5
α, β, γ(°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution(Å)	15.0-1.7	15.0-1.7	15.0-1.7	15.0-1.7	15.0-1.7	15.0-1.7
R_{merge}(%)[a]	9.4 (73.7)	9.4(76.3)	8.9(59.9)	8.9(58.6)	8.0(57.2)	8.0(56.9)
$CC_{1/2}$ [a]	100.0(99.1)	100.0(99.1)	100.0(97.7)	100.0(98.1)		
$<I/\sigma(I)>$[a]	75.56(10.49)	75.58(10.10)	46.81(5.51)	46.92(5.63)	84.49(12.95)	84.03(13.28)
Completeness(%)[a]	91.1(66.7)	90.9(65.2)	92.1(73.8)	92.2(74.0)	93.4(74.8)	93.6(76.3)
Redundancy	205.0	197.0	68.8	68.7	182.0	184.2
Refinement						
No. reflections	11 830	11 829	11 954	11 929	12 140	12 140
R_{work}/R_{free}(%)	0.1658/0.1908	0.1659/0.1902	0.1689/0.1939	0.1683/0.1962	0.1615/0.1842	0.1612/0.1836
No. atoms	1042	1042	1042	1042	1042	1042
R.m.s. deviations						
Bondlengths(Å)	0.0229	0.0229	0.0221	0.0223	0.0231	0.0231
Bond angles(°)	1.9907	1.9804	1.9695	1.9823	2.0222	2.0224

[a] Values for the highest resolution shell (1.70-1.74) are shown in parentheses.

4.2 Difference electron density maps

Difference electron density Fourier maps, $F_{0.4THz}$-F_{off}, $F_{0.2THz}$-F_{off} and F_{even}-F_{odd} were calculated using the structure factor amplitudes from the merged data. Phases were refined for F_{off} and F_{odd} data for the terahertz and reference map calculations, respectively. The three maps are shown in Figure 4.3. Electron density differences emerges in the $F_{0.4THz}$-F_{off} map but only noise can be seen in the $F_{0.2THz}$-F_{off} and F_{even}-F_{odd} maps. The strongest features in the $F_{0.4THz}$-F_{off} map are four positive difference peaks that are associated with helix 3. The four helix peaks is among the six strongest peaks in the whole map. Since these peaks are not randomly situated close to the helix and the reference map contains only noise we can conclude that the positive difference peaks in the $F_{0.4THz}$-F_{off} map is caused by the terahertz irradiation. Interestingly, the $F_{0.2THz}$-F_{off} shows no apparent electron density differences which indicates that the terahertz effect is frequency specific. Worth noting is that the 0.2 THz data is collected on fewer frames than the 0.4 THz data which may cause a 0.2 THz induced effect not to be visible due to lack of data. A low frequency vibration in helix3 may be related to lysozyme function since it lines the substrate binding cleft and is directly linked to one of the hinge residues, Gly102 [53].

4.3 Structural analysis

In order to understand the structural origin of the positive electron density features, B-factors and atomic coordinate changes were compared from structural models refined against the 0.4 THz data (THz_{on} and THz_{off} separately) and reference data (odd and even frames separately). B-factors describes both static disorder arising from crystal inhomogeneity and dynamic disorder due to atomic movements around their equilibrium position. The dynamic disorder is directly coupled to the proteins' dynamics and a terahertz induced effect can therefore be visible in the difference in B-factors between the ground state and terahertz excited state. B-factors estimated with NMD

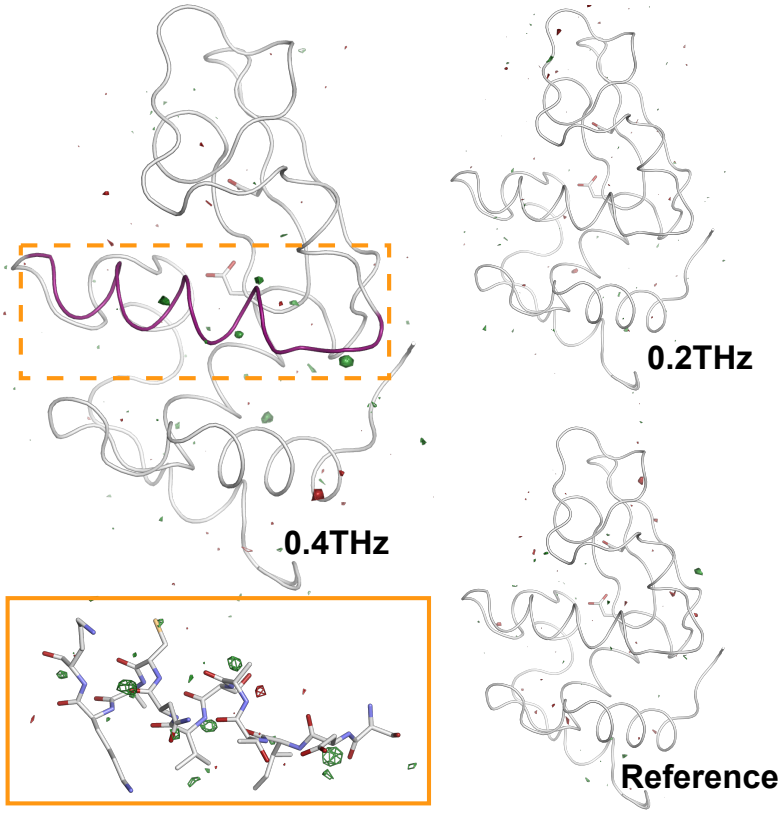

Figure 4.3: $F_{0.4THz}$-F_{off} (top left), $F_{0.2THz}$-F_{off} (top right) and reference F_{even}-F_{odd}(bottom right) difference electron density maps. Positive and negative differences are shown in green and red, respectively. The inset in the lower left is a full atom model of helix 3 with the positive electron difference peaks from the $F_{0.4THz}$-F_{off} map. The σ level is 3.5 for all maps and 2.8 for the inset.

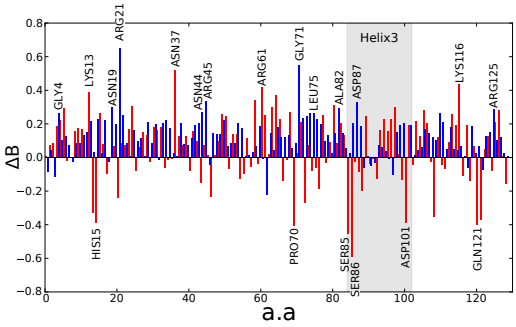

Figure 4.4: B-factor differences, THz_{on}-THz_{off}, from structural models refined against the 0.4 THz data (red) and B-factor difference between structural models refined against odd and even frames of the reference data (blue).

corresponds well with B-factors from several crystallographic structures [137, 138] showing the impact of low frequency vibrations on the B-factors. The difference B-factors for both the 0.4 THz data and reference data are mainly positive for all residues in the protein (Figure 4.4). The positive difference between B-factors from odd and even frames of the reference data is due to the difference in radiation damage (X-ray dose difference of 5.5 Gy). The same radiation damage difference is present between 0.4 THz illuminated data and non illuminated data but the plot in Figure 4.4 shows more negative differences in B-factors. Removing the radiation damage effect from the THz_{on}-THz_{off} difference would most certainly produce an even more negative B-factor difference. The more negative differences in B-factors is consistent with a structural ordering upon terahertz illumination. Interestingly, the structural ordering reveals an interaction between terahertz and the protein not caused by pure heating since thermal vibrations would cause an expansion and more disorder in the protein.

The same structural models were used for analysing the change in atomic coordinates. When projecting the difference in atomic coordinates of all atoms in helix 3 onto the helix axis, we observe a

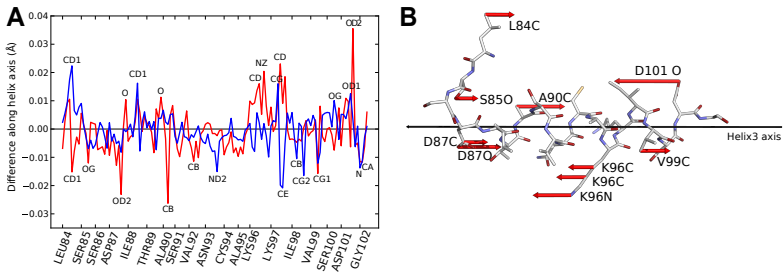

Figure 4.5: A. Atomic coordinate changes projected onto the helix axis from structural models refined against the 0.4 THz data (THz_{on}-THz_{off}) (red) and reference data (odd-even frames) (blue). **B.** Enhanced atomic coordinate changes along the helix axis upon 0.4 THz illumination.

compression of the helix upon terahertz illumination. A compression together with the higher ordering of the structure indicated by the negative B-factors can be explained by that the electrons occupy a smaller volume upon terahertz illumination which in turn could give rise to the observed positive electron density differences.

4.4 The protein crystal environment

Low frequency vibrations in a crystallized protein can be both intra and inter molecular. The inter molecular vibrations are phonons, an intrinsic property of the crystal lattice. Lysozyme crystal lattice phonons are expected to have frequencies below 0.3 THz. The 0.4 THz induced vibration observed here is therefore most likely an intra molecular vibration with a frequency similar to a previously observed intra molecular vibration of 0.36 THz in a tetragonal lysozyme crystal [136]. In our experiment, it is unlikely that we observe any orientation dependent effect. This is because the eight unique orientations of helix3 in the crystal will have an even distribution of angles to the incoming terahertz radiation during the data collection (Figure 4.6).

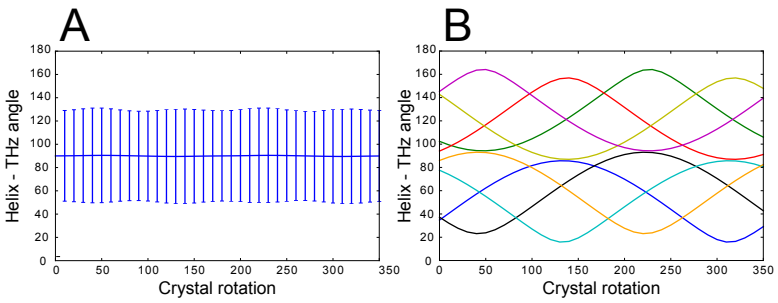

Figure 4.6: **A.** The average and standard deviation of the angles between helix3 and terahertz polarization vector over a 360° rotation in the {110} plane. **B.** The angle between every unique helix 3 in the unit cell (different colours) and the terahertz polarization vector over a 360° rotation in the {110} plane.

The terahertz absorption of a lysozyme crystal is partly due to the water content [139]. Water within a protein crystal can be separated into hydration water important for protein function [140] and bulk water. Hydration water form layers that creates a water-protein interaction network with strong collective properties [103, 141]. Taking together that a tetragonal lysozyme crystal contains 33 % water [142] and that a typical globular protein coordinates two hydration shells with a radius of 3.65 Å [143], the protein separation in the crystal leaves space for not much more than the two hydration layers. The majority of the water in a lysozyme crystal is therefore involved in a water-protein interaction network. Consequently, even if the coordinated water is absorbing the radiation, the excitation energy can be transferred to the protein. According to NMD calculations on hydrated lysozyme, a water content equal to a tetragonal lysozyme crystal is not changing the calculated normal mode density at a frequency of 0.4 THz [46].

4.5 Terahertz heating effects

When irradiating a protein crystal with terahertz radiation at room temperature we expect the radiation to interact with the protein. The crystal water could absorb the radiation and the energy be transferred to the protein as heat which then would cause the electron density difference. By experimentally measuring the heating rate in the protein crystal upon terahertz irradiation we estimated 0.4 THz radiation to heat the crystal 1 mK during the collection of one diffraction image. This small heating is not enough to shift a conformational population more than 0.1 % and it is therefore highly unlikely that heating is the cause of the difference signal. Also, heating would cause an expansion of the protein which goes against our observation of the helix compression and decrease in B-factors.

Furthermore, the higher power for 0.2 THz compared to 0.4 THz would lead to higher temperature upon illumination. If heating was the cause of the observed effect we would expect to see a stronger electron density for 0.2 THz than for 0.4 THz.

4.6 Fröhlich condensation

By illuminating lysozyme with terahertz radiation we excited low frequency vibrational modes in the structure that became visible through positive electron density differences. After a vibrational mode gets excited by a photon, how long does it stay in this state before the excitation energy is dissipated to the surroundings?

Taking into account the power at 0.4 THz and the number of lysozyme molecules in the X-ray probed volume, one lysozyme molecule receives around one photon per millisecond. To keep a terahertz excited population of 10 %(which is the detection limit for a difference electron density map) during the 20 ms collection time of one diffraction image, the lifetime of the excited vibrational mode has to be in the micro to millisecond time range.

Protein low frequency vibrations are underdamped [26, 28, 40, 100, 101, 144]. An underdamped vibrational mode will thermalize slower

than the cycle time of the exciting radiation which can explain a long life time. According to MD simulations of lysozyme, the underdamping at 0.4 THz can at most prolong the life time to the nanosecond time scale [145]. Therefore, underdamping cannot be the sole reason for the observed micro to millisecond lifetime.

We found that the only theory able to explain this exceptionally long life time of low frequency vibrations is Fröhlichs theoretical model postulated more than four decades ago [146]. According to Fröhlichs model, excitation energy from a terahertz photon can be stored (condense) into a low frequency mode for an extended period of time [147]. Fröhlichs model describes a protein as a collection of coupled oscillators in a thermal bath. Upon illumination, the oscillators will get excited and then transfer the excitation energy in between each other and to the thermal bath. If the transfer rate between the oscillators is faster than the energy transfer to the surrounding thermal bath, a Fröhlich condensate can form (Figure 4.7). Fröhlichs theory states that polar characteristics of the biological material is needed for the condensate to form [148]. Since an α-helix is a macro dipole it thereby has the prerequisites for a Fröhlich condensate formation.

4.7 Normal mode analysis

We performed NMD to investigate if a redistribution of a normal mode population in line with Fröhlich model can give rise to positive electron density differences. NMD is a harmonic approximation of motion, and even if a fraction of low frequency vibrations are expected to be anharmonic [29, 149, 150] the harmonic normal modes has the same directionality as anharmonic modes but tends to underestimate the amplitude [26, 151–153]. For an energy minimized lysozyme molecule, 25 normal modes were calculated with frequencies between 0.2-0.5 THz. The 10 normal modes with the highest root mean square fluctuations in helix3 were chosen for the analysis. The ground state molecule was described by an ensemble of the 10 modes evenly populated. The excited state was described by only one mode since the protein motion of a Fröhlich condensate would

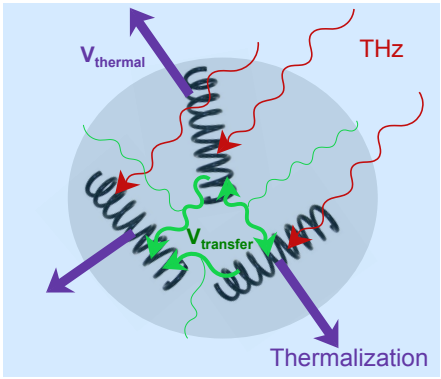

Figure 4.7: Three coupled oscillators (black springs) in a system (circle) in contact with a thermal bath (blue). The oscillators are getting energy from terahertz photons (red) and the energy can redistribute within the system(green) or thermalize to the bath (purple). If the transfer rate between the oscillators, $V_{transfer}$, is faster than the thermalization, $V_{thermal}$, a Fröhlich condensate can form.

be dominated by one vibration. Ten electron difference maps were calculated between the ground state ensemble and all ten different modes describing the Fröhlich condensate. Positive electron density differences emerges in several maps but especially one had striking similarities with the $F_{0.4THz}$-F_{off} map (Figure 4.8A). The type of motion giving rise to the observed electron density differences are shown in Figure 4.8B and gives indications of the Fröhlich condensate vibration observed in our experiment. This NMD approach also shows that redistribution of vibrational motions can give rise to measurable electron density features.

4.8 Conclusions

In **Paper II** we showed for the first time how terahertz radiation can give rise to positive electron density differences in a protein. The experimental design and data analysis made it possible to detect a

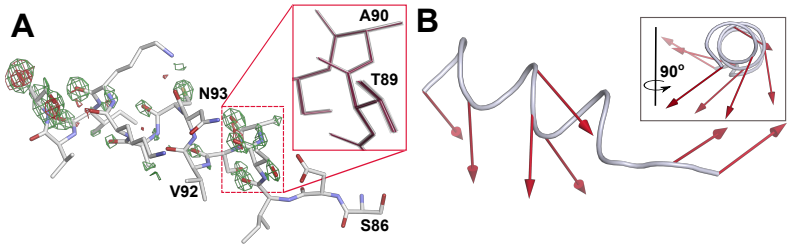

Figure 4.8: A. Electron density map between ground state described as an ensemble of ten modes and an excited state described as one fully occupied mode. The inset shows an overlay of the extreme positions of all ten modes around two amino acids in the helix. **B**. Arrows shows the direction of movement of the mode representing the excited state in A.

small but significant signal caused by 0.4 THz irradiation. The observed positive electron density differences were associated with helix 3 in lysozyme that lines the substrate binding cleft and is neighbouring one of the hinge residues, therefore the excited vibrational mode may be related to lysozyme function. An exceptionally long lifetime of the vibrational mode was observed which can only be explained by the formation of a Fröhlich condensate in the protein. Fröhlich formulated his theory over 40 years ago and this is to my knowledge the first experimental observation of a Fröhlich condensate in a protein.

Chapter 5

Terahertz effect on thermolysin diffraction data

In **Paper III** we examined the terahertz induced effects in thermolysin with a similar approach as described in **Paper II**. In **Paper II**, several days of beam time was used to collect the large amount of diffraction data for lysozyme that was needed for a difference signal to become visible. Lysozyme was expected to have functionally relevant vibrations with frequencies in the terahertz region but for the technique to be applicable to a wider range of more unknown proteins, is there a way to detect a terahertz induced effect with less data? In **Paper III** we answer this question by employing a Bayesian statistical method to detect Bragg peak intensity changes caused by terahertz illumination in thermolysin diffraction data. Any structural change in a protein visible in a electron density map should already be detectable in the diffraction image since the Bragg peak intensities is the only experimentally measured data. Investigating the Bragg peak intensities directly may also be less error prone than the calculation of electron density difference maps since inaccurate phases is a large

source of errors [154] and for map calculation several scaling and data reduction steps are performed that may introduce systematic errors. On the other hand, electron density maps provide the structural information on *where* the induced vibrational effect is located. The two approaches can be used together: the Bragg peak intensity analysis to investigate *if* a terahertz induced effect is present and afterwards, the calculation of electron density maps to provide structural information. The certainty of a electron density difference being the cause of terahertz radiation is increased by the accompanying reflection intensity analysis.

5.1 Still image diffraction

To increase the redundancy per reflection without increasing the data collection time a small part of reciprocal space can be examined by measuring several diffraction frames from the same crystal orientation, so called still images. The termolysin crystal was mounted in a loop and kept at 99 % humidity using a humidifier stream. 0.4 THz radiation were illuminating one uniform face of the crystal and the X-ray micro focus beam probed close to the surface similarly to the data collection in **Paper II**. Every second frame was collected from a crystal illuminated by terahertz and every second without any illumination through synchronization with the detector read-out signal. $\mathrm{THz_{on}}$ and $\mathrm{THz_{off}}$ frames were separated into two groups. In total, 2000 images were recorded with a 10 ms exposure per image which made it possible to study the intensity profile over time. From a randomly oriented crystal the 95 strongest peaks were identified (max resolution 2.7 Å) and used in the analysis (Figure 5.1).

5.2 Terahertz radiation causes intensity changes

Using two different Bayesian statistical approaches, the time dependent differences in photon count between $\mathrm{THz_{on}}$ and $\mathrm{THz_{off}}$ frames were analysed. In the first approach, the Bragg peak intensities from

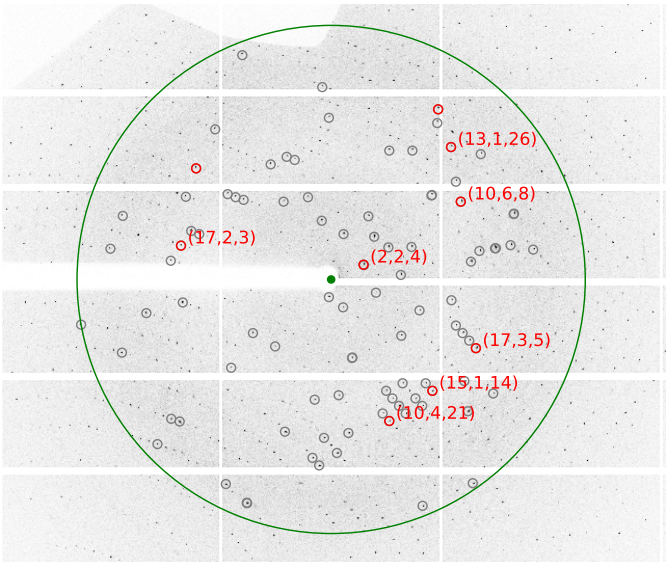

Figure 5.1: Sum of 100 diffraction still images of thermolysin. All 95 Bragg peaks investigated in the statistical analysis are encircled in grey, the nine peaks with a credible interval (95 % HDI) not including zero from the statistical analysis between THz_{on} and THz_{off} frames are highlighted with red. The green circle (resolution 2.7 Å) shows the maximum resolution of the 95 peaks used in the analysis. The shadow in the top left corner comes from the terahertz source.

THz_{on} and THz_{off} frames were modelled as two Poisson stochastic variables and the difference between these variables were estimated in the statistical analysis. In the second approach the difference in Bragg peak intensity were modelled directly as a normal distribution. In both approaches a parameter describing a linear decay on the time dependent photon count was incorporated to account for any time dependent effect. The difference in photon count, $I_{0.4THz}$-I_{dark}, was close to zero for all peaks but using the Bayesian statistical method small differences could be detected. Both models identified six out of 95 peaks to have a significant difference in Bragg peak intensity upon terahertz illumination (95 % HDI credible intervals). When modelling the separate observations as Poisson variables three more peaks were found to have a non-zero difference intensity (shown in Figure 5.1). For all nine peaks, the photon count difference was estimated to be at maximum (around 5 % photon count difference) in the first frames and decayed with time probably due to radiation damage. Eight of the peaks are found within a 3.4-4.6 Å resolution range indicating that the terahertz effect is more pronounced for low intensity, high resolution reflections. The estimated intensity differences are both negative and positive which can be explained by an electron density redistribution in the protein upon terahertz irradiation probably due to the excitation of low frequency normal modes.

5.3 Analysis of rotation diffraction data

As a next step, we analysed how terahertz affects diffraction intensities from a much larger part of reciprocal space compared to the still images. This was done by collecting rotation data using the same strategy as for the still images but collecting over 360° rotation and using the extreme fine slicing strategy (0.01° oscillation) employed in **Paper II**. After initial processing of the THz_{on} and THz_{off} frames the reflections were scaled but the normally employed step of merging multiple observations was omitted. By not merging the reflections, the information contained in the distribution of measured intensities for each reflection can be utilized when estimating the intensity and

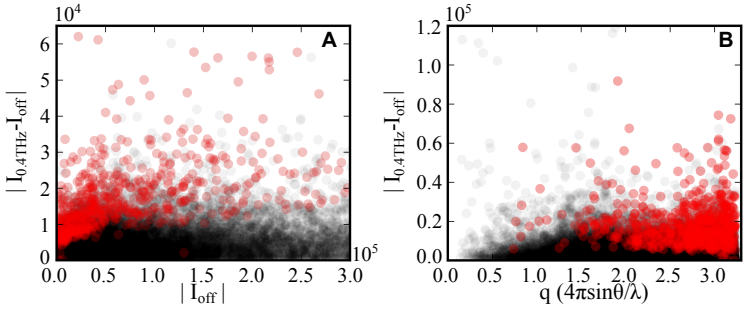

Figure 5.2: Distribution of the estimated intensities of all reflections (black) and reflections with an estimated significant intensity change upon 0.4 THz illumination (red). Panel A shows the estimated absolute intensity of all reflections recorded with the terahertz source switched off versus the estimated difference between intensities recorded during terahertz illumination, I_{THz}, and intensities recorded with no terahertz illumination, I_{off}. Panel B shows the difference in estimated intensities, $|I_{THz} - I_{off}|$, as a function of q (momentum transfer magnitude, $\propto$ resolution). Intensity values are given on an arbitrary scale.

error with an Bayesian approach. Each reflection was modelled as a normal distribution in the Bayesian estimation. The mean and error of the posterior distribution corresponds to I and σ(I) respectively, for each reflection. A total of 24 078 unique reflections with a redundancy of 140 collected on 12 crystals were used in the analysis. Significant changes (I/σ>1.7) in intensity upon terahertz illumination were found for 710 reflections out of the 24 078 reflections analysed (2.95 %). The terahertz influenced reflections are generally more abundant among low intensity and high resolution reflections similarly to what was observed for the still image (Figure 5.2).

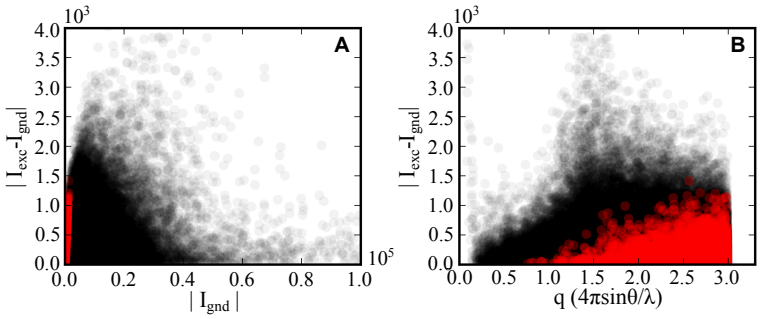

Figure 5.3: Distribution of all intensities calculated from normal mode structural ensembles representing a ground state, I_{gnd}, and excited state, I_{exc} (black dots). Red dots represents reflections changing more than 10 % between I_{gnd} and I_{exc}. The absolute of the difference between simulated I_{gnd} and I_{exc} plotted versus absolute intensity of the ground state (**A**) and as a function of q (**B**).

5.4 Normal mode analysis

A normal mode analysis was performed on thermolysin to investigate how a redistribution of occupancy between vibrational modes changes the distribution of Bragg peak intensities compared to the intensities estimated with the Bayesian method from oscillation data (Figure 5.2). Ten modes with frequencies around 0.4 THz were selected for the analysis. Similarly to the normal mode analysis in **Paper II**, the ground state was described by an ensemble of all ten modes with equal occupancy and the excited state as one fully occupied mode. Intensities were calculated for the ground state and excited state up to a resolution of 2.1 Å. Distribution of the calculated intensities are shown in Figure 5.3 in a similar representation as Figure 5.2. The distribution of the reflections with at least 10 % change in intensity (3.1 % of all reflections) are more abundant for low intensity reflections (Figure 5.3A) at higher resolution (Figure 5.3B). This analysis shows that the excitation of one low frequency normal mode can produce the intensity distribution observed experimentally.

5.5 Structure factor amplitude estimation

Structure factor amplitudes, $F_{THz\ on}$, $F_{THz\ off}$ and ΔF, were estimated together with the reflection intensities from unmerged reflections using the Bayesian method. For comparison, the rotation data was also merged according to standard procedure. The standard approach of estimating structure factors from merged intensities and the strategy described here where structure factors are estimated from the distribution of unmerged reflections, are both using a Bayesian parameter estimation. The difference is that our approach uses less assumptions on the estimation, the prior only forces positivity on the reflections compared to the prior defined by Wilson [155] used in French and Wilsons algorithm [132] implemented in most merging software (e.g. cTRUNCATE and XDSCONV). Estimating structure factors from unmerged reflections is more computationally heavy but since less assumptions are made in the prior, the posterior distribution is less biased by prior beliefs. Comparing the two approaches, estimation from non-merged reflections yields a 96 % completeness while XDSCONV rejects a large proportion of the data which results in a 86 % completeness. The rejected reflections are mainly negative which is probably due to a faulty background subtraction and not caused by outliers since the whole distribution of several rejected reflections were found to be negative.

5.6 Electron density difference map

Structural models were refined against the Bayesian estimated $F_{THz\ off}$ structure factors and a second model were refined against merged THz_{off} data. The structural models as well as the 2Fo-Fc maps are similar for both models but R and R_{free} are higher for the model refined against non-merged data due to the large error on the weak reflections otherwise excluded. The $F_{0.4THz}$-F_{dark} electron density map is shown in Figure 5.4 and is calculated from $F_{0.4THz}$ and F_{dark} from the Bayesian estimate from the non-merged reflections. Electron density differences emerges in several parts of the structure. A

terahertz induced movement of the residue Y76 in one of the helices in the N-terminal lobe and of V253 in a loop in the C-terminal region of thermolysin is observed. The residue movements are also observed in the difference map calculated from merged intensities but is less pronounced and the strongest feature is seen at Y76.

5.7 Conclusions

We have shown that terahertz radiation interacts with thermolysin molecules in a crystal, first by observing changes in Bragg peak intensities and secondly as electron density differences. A sensitive Bayesian method was employed that can detect subtle differences while introducing less bias towards prior beliefs compared to a standard approach. This is a first demonstration of the Bayesian approach for estimating structure factor amplitudes from unmerged reflections. The approach can be extended to handle background subtraction of diffraction images to improve the estimation of weak reflections. By combining the still image analysis and difference electron density map calculations the certainty for ascribing observations as a cause of terahertz irradiation is increased. The still image analysis can be used to screen for proteins that interact strongly with terahertz radiation so that the more time consuming rotation data collection can be performed on the most suitable target.

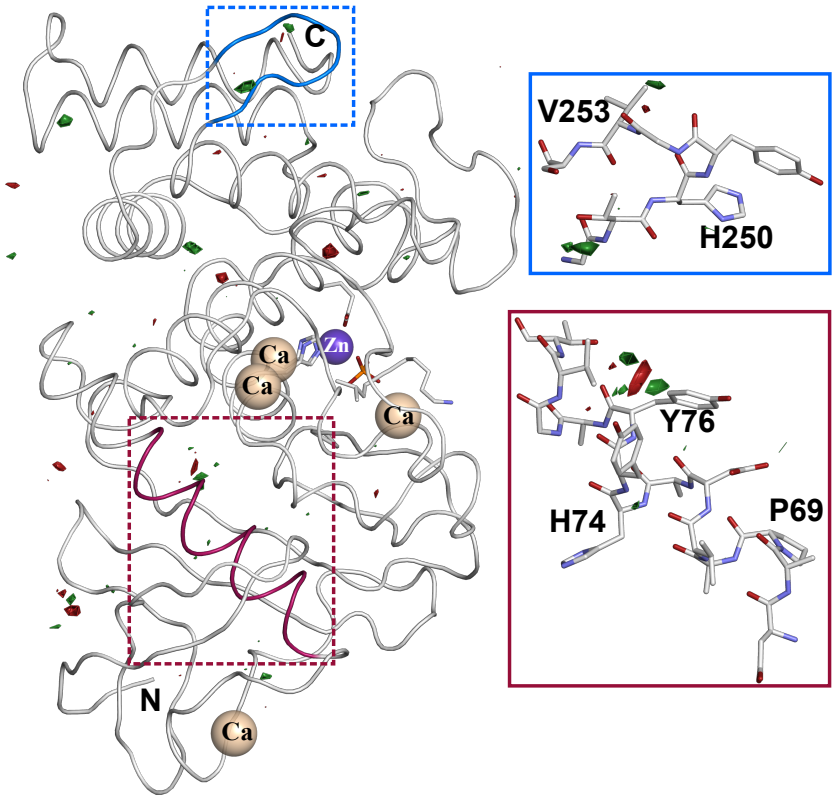

Figure 5.4: The difference Fourier electron density map of thermolysin at 2.1 Å resolution. Positive and negative electron density differences are shown in green and red respectively. Two different regions of the structure is highlighted with different colors, the two insets show a full atom model of each highlighted region to visualize the electron density differences in more detail. The sigma level of the full structure map is 3.5 and for the zoomed in full atom insets the sigma level is 3.4. The structure contains four calcium and one zink that are shown in beige and purple respectively. The zink ion is coordinated in the active site by H146, H142 and E166 drawn in stick representation. A phosphate group is bound to the zink instead of water and in the active site a dipeptide (K-L) is modeled.

Chapter 6

Concluding remarks

The work presented in this thesis has exploited the property of terahertz radiation to excite collective low frequency vibrations in proteins. The only direct way of detecting low frequency vibrational modes is to use broadband terahertz radiation as a probe. Unfortunately the terahertz absorption bands of proteins are too broad and featureless for an easy identification of specific normal modes. On the other hand, investigating the difference absorption between ground and excited state of RC_{sph} revealed a change in the density of vibrational modes available to the system together with the photo induced charge separation. In light saturating conditions, the excited state of RC_{sph} has a complex energy landscape with a large number of isoenergetic CS. Instead of detecting one or a few vibrational modes in common between these CS, a broad and featureless difference spectrum was recorded similar to an absolute protein terahertz absorption spectrum. Whether or not certain modes in RC_{sph} are crucial for the stability of the charge separated state is a question that remains to be answered. The study however, gives insight into the vast number of vibrational modes present in the excited state ensemble. Importantly, the change in vibrational mode density was localized to the L and M subunits accommodating the cofactors. We also found that the difference signal was not dependent on the protein being in detergent or

lipid based environments.

A challenging yet crucial part of the data analysis was to rule out if a thermal effect was the main cause of the signal or not. It would be interesting to further investigate the thermal response on RC_{sph} light activation in the terahertz frequency region and to separate thermal response from the light response. There is a large number of measurements that would be interesting to perform but synchrotron time is limited. Broad band terahertz absorption is extremely limited at providing the detailed structural information needed for understanding the role of low frequency vibrations and their involvement in protein function. Instead, the technique can be used to study proteins from a material perspective by investigating how the environment is affecting the low frequency vibrations. The solvent affects protein dynamics and more studies towards understanding how the protein interacts with the surroundings can be pursued with terahertz absorption spectroscopy.

The use of terahertz radiation as a pump in X-ray crystallography provided a more structurally detailed view of terahertz excitation compared to absorption spectroscopy studies. The choice of pump frequency is limited because of the lack of high power terahertz sources with tunability over a broad frequency range that are also both affordable and compact. NMD and MD modelling are crucial tools for identifying possible normal modes. Modelling and experiments go hand in hand, and mapping the solvated protein properties can also assist in acquiring better models in the future. A more accurate estimation of important modes can assist in choosing the correct frequency of terahertz excitation in an experiment.

The experimental design presented in **Paper II** and **III** is not limited to terahertz studies but it can also be applied to study other inducible structural changes in crystallized proteins. **Paper II** presented the first successful detection of electron density changes introduced by terahertz irradiation. A putative Fröhlich condensate was for the first time detected in a protein. No other theory is able to explain the exceptionally long lifetime of the excited vibrational mode in our experiment. Self organization within living organisms

on a macroscopic as well as microscopic level seemingly goes against the second law of thermodynamics which intrigued prominent physicists, for example Schrödinger, in the 20th century [156]. Fröhlich theory is a general model that may describe in detail the origin of self organization within the limits of thermodynamics [157]. Our experiment is not ultimately proving Fröhlich hypothesis but gives a strong indication of its presence. The relevance of Fröhlichs condensation to biological processes may be revealed in the future as the technology for production of terahertz radiation will evolve and provide researchers with more powerful pulsed sources. There is a huge number of other exciting experiments that can be done with pulsed and tunable high power terahertz sources. For example, elucidating frequency dependency and performing time resolved experiments.

The presence of terahertz excited vibrations in a crystallized protein can be demonstrated both through the calculation of electron density difference maps as well as through Bragg peak intensity changes. A direct study of the diffraction images gives a less biased result compared to map calculation but the structural information is of course lacking. The Bayesian statistical approach could detect small but significant changes in Bragg peak intensities upon terahertz irradiation. The presented method can be used for cross validation of electron density maps and a faster way of screening for terahertz induced effects due to a change of experimental parameters at the synchrotron. Dynamic information can also be obtained from the diffuse scattering from a crystal. Terahertz induced effects may be found in the weak diffuse scattering after further method development for still image diffraction. The Bayesian approach used to estimate structure factor amplitudes from unmerged reflections use less assumptions in the data treatment compared to a more standard approach. The method can be improved by adapting other steps in the data processing, for example the background subtraction. The method can become a less biased tool for analysing diffraction data with high redundancy since both structure factor amplitude and error can be estimated from the recorded intensity distribution.

The field of protein terahertz research is ready to move from convincing that low frequency vibrations exist in proteins to determining the actual coupling to function, if any. To get there, terahertz techniques should continue to be improved and developed for protein applications. I hope future research on protein dynamics will fill the current "teraherz gap" in published research as the technological advances now are bridging it regarding terahertz production and detection.

Acknowledgements

Firstly, I would like to thank my supervisor, **Gergely**, for taking me in as a student in your group. I am very grateful for everything I got the chance to learn from you and it has been great to discuss science on topics ranging from physics to biology. Your courage in tackling the big questions is truly inspirational!

Thank you **Richard** for being my examiner as well as coming up with great ideas for the papers.

Thanks to the members of Gergelys group **Weixiao**, **Annette** and **Majo**. We all had fun times both in and outside the lab. **Weixiao**, thanks for all the help in the lab during the first years, you became like a co-supervisor to me and I learned a lot working with you. **Annette**, I think we really had a lot of fun together, I especially remember the loooong walks with you trying to convince me that "it's not so far" and also that evening with the Slovenian schnaps. **Majo**, the beam-times with you were always fun! Welcome to the group **Hassan** and good luck with the continuation of the terahertz project.

Thanks to **Jordi** for doing your master project with me, it was really fun to have you in the lab and I was really inspired by your passion for science. Hope you will enjoy your time as a PhD student as much as I have!

Thanks also to **Helena**, **Jan**, **Josip** for a great collaboration. A special thanks to **Helena**; I think we were a great team when re-writing the paper together and you taught me a lot about terahertz. I think we also managed to have a lot of fun during the long beam times all over Europe.

Mike and **Jennie**, we did this PhD adventure together from day one and I am so happy for always being welcome at your place for movie nights and dinners. **Mike**, I'm truly happy you decided to befriend me that first week, we always managed to find interesting topics to talk about, be it science or philosophy and it's nice to have someone to enjoy a good wine with. **Jennie**, all the times training at Friskis were always fun with you, your energy is really contagious :) Hope both of you will have a great time in Australia!

Rebecka, why didn't you start your PhD earlier? You always manage to make me laugh with your random topics and it's always great to hang out with you.

Rosie Your laugh really lightens up the corridor. I'm so happy you got the structure, congratulations! Also, thanks for chairing my defence.

Petra B, Oskar, Emil, Rob, Linnéa, David, Elin J, Matthijs, Ash, Rajiv, Petra E, Alex J - Beer clubs, lunches, coffee breaks, and BBQ's have always been so much fun at the Lundberg lab thanks to you! I will really miss it.

Elin D, thanks for sharing the office with me, I think we had some really fun moments. **Rhawnie**, soon it's your turn! you can do it :) **Maria H**, thanks for the nice after-works and discussions about feminism, hope to see you soon in Germany! **Stephan** Smaklig måltid! **Alex B**, I really enjoyed our after choir beers on the tuesdays.
Cecilia, it has always been nice to talk to you about everything from programming to music.

The rest of the Lundberg lab; **Gisela, Sebastian W, Kristina, Jenny, Tobias, Jiao, Raphal, Parveen, Amit**, it's been great to have you around. The new people, **Cecilia, Rob, Greger**, welcome to the lab and good luck with your projects!

Thanks to all former Lundberg lab people for all the fun times during my first years in the lab, **Karin, Peter, Mikael, Anna, Sebastian P, Linda, Erik, Gerhard**. **Karin** ESRF trips were always fun with you, what is a night shift without sugar saturated coffee? **Peter** thanks for the pep talks during the more difficult times.

Thanks to **Lars** and **Bruno** for all the help with fixing misbehaving computers and lab equipment and thanks to **Örjan** and **Martin** for organizing the courses.

Tack till alla mina vänner för att ni alltid får mig på bra humör.

Tack till min familj för att ni finns där och stöttar mig även fast jag aldrig skaffar ett "riktigt" jobb. Tack för dom fantastiska sommrarna i Nyhult där jag kunde studera sniglarnas anatomi och vandringsrutter vilket helt klart var första steget mot att vilja doktorera. **Mamma**, det finns ingen bättre lyssnare än du och med kloka ord hjälper du mig alltid upp när allt känns som tyngst. **Pappa**, du har alltid stöttat mig och fått mig att vilja upptäcka och lära mig nya saker, finns ingen person med sådant driv som du och jag slutar aldrig förundras över hur du orkar hålla sådant tempo.

The time as a PhD student has been really great, I've learned so much and met so many awesome people. But the best thing with it all was that this PhD made me attend that synchrotron school in Duino, Italy, 2011 where I met the love of my life, **Federico**. Thank you for always cheering me up, for always listening, for sharing everything with me and for making me remember what's really important in life. I love you and I can't wait to marry you next year!

References

[1] R. J. P. Williams. The conformation properties of proteins in solution. *Biological Reviews*, 54(4):389–437, 1979.

[2] M. S. Richard P. Feynman, Robert B. Leighton. *Six easy pieces*, chapter Atoms in motion, page 59. Addison-Wesley, 1964.

[3] J. C. Kendrew, G. Bodo, H. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181:662–666, March 1958.

[4] H. Muirhead and M. F. Perutz. Structure Of Haemoglobin: A Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5.5Å Resolution. *Nature*, 199:633–638, Aug. 1963.

[5] D. Koshland Jr. Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):98, 1958.

[6] A. Kohen. Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Accounts of chemical research*, 48(2):466–473, 2014.

[7] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. Protein states and protein quakes. *Proceedings of the National Academy of Sciences*, 82(15):5000–5004, 1985.

[8] K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450:964–972, december 2007.

[9] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.

[10] J. D. Bryngelson and P. G. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21):7524–7528, 1987.

[11] J. D. Bryngelson and P. G. Wolynes. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *The Journal of Physical Chemistry*, 93(19):6902–6915, 1989.

[12] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

[13] R. H. Austin, K.-W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–5373, 1975.

[14] S. J. Benkovic, G. G. Hammes, and S. Hammes-Schiffer. Free-energy landscape of enzyme catalysis. *Biochemistry*, 47(11):3317–3321, 2008.

[15] D. D. Boehr, R. Nussinov, and P. E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, 5(11):789–796, 2009.

[16] B. Brooks and M. Karplus. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences*, 80(21):6571–6575, 1983.

[17] N. Gō, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences*, 80(12):3696–3700, 1983.

[18] B. Brooks and M. Karplus. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences*, 82(15):4995–4999, 1985.

[19] M. Levitt, C. Sander, and P. S. Stern. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, 181(3):423 – 447, 1985.

[20] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins Structure Function and Genetics*, 33(3):417–429, 1998.

[21] P. Dauber-Osguthorpe, D. J. Osguthorpe, P. S. Stern, and J. Moult. Low frequency motion in proteins: comparison of normal mode and molecular dynamics of *streptomyces griseus* protease A. *Journal of computational physics*, 151(1):169–189, 1999.

[22] H. W. van Vlijmen and M. Karplus. Analysis of calculated normal modes of a set of native and partially unfolded proteins. *The Journal of Physical Chemistry B*, 103(15):3009–3021, 1999.

[23] F. Tama, F. X. Gadea, O. Marques, and Y.-H. Sanejouand. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function, and Bioinformatics*, 41(1):1–7, 2000.

[24] R. W. Harrison. Variational calculation of the normal modes of a large macromolecule: methods and some initial results. *Biopolymers*, 23(12):2943–2949, 1984.

[25] J.-F. Gibrat and N. Gō. Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins: Structure, Function, and Bioinformatics*, 8(3):258–279, 1990.

[26] A. Kitao, F. Hirata, and N. Gō. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chemical Physics*, 158(2):447–472, 1991.

[27] A. E. García. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68:2696–2699, Apr 1992.

[28] S. Hayward, A. Kitao, F. Hirata, and N. Gō. Effect of solvent on collective motions in globular protein. *Journal of Molecular Biology*, 234(4):1207–1217, 1993.

[29] A. Amadei, A. Linssen, and H. Berendsen. Essential dynamics of proteins. *Proteins Structure Function and Bioinformatics*, (17):412–25, 1994.

[30] D. Perahia and L. Mouawad. Computation of low-frequency normal modes in macromolecules: Improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Computers & Chemistry*, 19(3):241–246, 1995.

[31] O. Marques and Y.-H. Sanejouand. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins: Structure, Function, and Bioinformatics*, 23(4):557–560, 1995.

[32] F. Tama and Y.-H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein engineering*, 14(1):1–6, 2001.

[33] M. P. Frushicheva, M. J. Mills, P. Schopf, M. K. Singh, R. B. Prasad, and A. Warshel. Computer aided enzyme design and catalytic concepts. *Current Opinion in Chemical Biology*, 21:56 – 62, 2014.

[34] J. Villa and A. Warshel. Energetics and dynamics of enzymatic reactions. *The Journal of Physical Chemistry B*, 105(33):7887–7907, 2001.

[35] S. Hay and N. S. Scrutton. Good vibrations in enzyme-catalysed reactions. *Nature chemistry*, 4:161–168, January 2012.

[36] K. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–916, 2007.

[37] K. A. Niessen, M. Xu, and A. Markelz. Terahertz optical measurements of correlated motions with possible allosteric function. *Biophysical Reviews*, 7(2):201–216, 2015.

[38] O. Keskin, R. L. Jernigan, and I. Bahar. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophysical Journal*, 78(4):2093–2106, 2000.

[39] J. A. McCammon and M. Karplus. The dynamic picture of protein structure. *Accounts of Chemical Research*, 16(6):187–193, 1983.

[40] A. Xie, A. F. van der Meer, and R. H. Austin. Excited-state lifetimes of far-infrared collective modes in proteins. *Physical review letters*, 88(1):018102, 2001.

[41] A. Markelz, S. Whitmire, J. Hillebrecht, and R. Birge. THz time domain spectroscopy of biomolecular conformational modes. *Physics in Medicine and Biology*, 47(21):3797, 2002.

[42] S. Whitmire, D. Wolpert, A. Markelz, J. Hillebrecht, J. Galan, and R. Birge. Protein flexibility and conformational state: A comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin. *Biophysical journal*, 85(2):1269–1277, 2003.

[43] C. Zhang, E. Tarhan, A. Ramdas, A. Weiner, and S. M. Durbin. Broadened far-infrared absorption spectra for hydrated and dehydrated myoglobin. *The Journal of Physical Chemistry B*, 108(28):10077–10082, 2004.

[44] J. Xu, K. W. Plaxco, and S. J. Allen. Collective dynamics of lysozyme in water: terahertz absorption spectroscopy and comparison with theory. *The Journal of Physical Chemistry B*, 110(47):24255–24259, 2006.

[45] J. Xu, K. W. Plaxco, and S. J. Allen. Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy. *Protein Science*, 15(5):1175–1181, 2006.

[46] J. R. Knab, J.-Y. Chen, Y. He, and A. G. Markelz. Terahertz measurements of protein relaxational dynamics. *Proceedings of the IEEE*, 95(8):1605–1610, 2007.

[47] R. Balu, H. Zhang, E. Zukowski, J.-Y. Chen, A. Markelz, and S. Gregurick. Terahertz spectroscopy of bacteriorhodopsin and rhodopsin: similarities and differences. *Biophysical journal*, 94(8):3217–3226, 2008.

[48] E. Castro-Camus and M. Johnston. Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy. *Chemical Physics Letters*, 455(4-6):289–292, 2008.

[49] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, M. Gruebele, D. M. Leitner, and M. Havenith. Protein sequence-and pH-dependent hydration probed by terahertz spectroscopy. *Journal of the American Chemical Society*, 130(8):2374–2375, 2008.

[50] G. M. Png, R. J. Falconer, B. M. Fischer, H. A. Zakaria, S. P. Mickan, A. P. Middelberg, and D. Abbott. Terahertz spectroscopic differentiation of microstructures in protein gels. *Optics express*, 17(15):13102–13115, 2009.

[51] Y. He, J.-Y. Chen, J. Knab, W. Zheng, and A. Markelz. Evidence of protein collective motions on the picosecond timescale. *Biophysical Journal*, 100(4):1058 – 1065, 2011.

[52] T. Q. Luong, P. K. Verma, R. K. Mitra, and M. Havenith. Do hydration dynamics follow the structural perturbation during thermal denaturation of a protein: A terahertz absorption study. *Biophysical Journal*, 101(4):925–933, 2011.

[53] K. Woods. THz time scale structural rearrangements and binding modes in lysozyme-ligand interactions. *Journal of Biological Physics*, 40(2):121–137, 2014.

[54] F. Piccirilli, G. Schirò, V. Vetri, S. Lupi, A. Perucchi, and V. Militello. Decoding vibrational states of concanavalin A amyloid fibrils. *Biophysical chemistry*, 199:17–24, 2015.

[55] R. H. Austin, M. W. Roberson, and P. Mansky. Far-infrared perturbation of reaction rates in myoglobin at low temperatures. *Physical review letters*, 62(16):1912, 1989.

[56] D. D. Klug, M. Z. Zgierski, S. T. John, Z. Liu, J. R. Kincaid, K. Czarnecki, and R. J. Hemley. Doming modes and dynamics of model heme compounds. *Proceedings of the National Academy of Sciences*, 99(20):12526–12530, 2002.

[57] Q. Wang, R. W. Schoenlein, L. A. Peteanu, R. A. Mathies, C. V. Shank, et al. Vibrationally coherent photochemistry in the femtosecond primary event of vision. *Science*, 266(5184):422–422, October 1994.

[58] J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel. X-ray structure analysis of a membrane protein complex: electron density map at 3Å resolution and a model of the chromophores of the photosynthetic reaction center from *rhodopseudomonas viridis*. *Journal of molecular biology*, 180(2):385–398, 1984.

[59] J. Allen and G. Feher. Crystallization of reaction center from *rhodopseudomonas sphaeroides*: preliminary characterization. *Proceedings of the National Academy of Sciences*, 81(15):4795–4799, 1984.

[60] K. E. McAuley, P. K. Fyfe, J. P. Ridge, N. W. Isaacs, R. J. Cogdell, and M. R. Jones. Structural details of an interaction between cardiolipin and an integral membrane protein. *Proceedings of the National Academy of Sciences*, 96(26):14706–14711, 1999.

[61] G. Feher. Some chemical and physical properties of a bacterial reaction center particle and its primary photochemical reactants*. *Photochemistry and photobiology*, 14(3):373–387, 1971.

[62] M. Michel-Beyerle, M. Plato, J. Deisenhofer, H. Michel, M. Bixon, and J. Jortner. Unidirectionality of charge separation in reaction centers of photosynthetic bacteria. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 932:52–70, 1988.

[63] A. J. Hoff and J. Deisenhofer. Photophysics of photosynthesis. structure and spectroscopy of reaction centers of purple bacteria. *Physics reports*, 287(1):1–247, 1997.

[64] R. Debus, G. Feher, and M. Okamura. LM complex of reaction centers from *rhodopseudomonas sphaeroides* R-26: characterization and

reconstitution with the h subunit. *Biochemistry*, 24(10):2488–2500, 1985.

[65] G. Katona, U. Andreasson, E. M. Landau, L.-E. Andreasson, and R. Neutze. Lipidic cubic phase crystal structure of the photosynthetic reaction centre from *rhodobacter sphaeroides* at 2.35Å resolution. *Journal of molecular biology*, 331(3):681–692, 2003.

[66] J. Allen, G. Feher, T. Yeates, H. Komiya, and D. Rees. Structure of the reaction center from *rhodobacter sphaeroides* R-26: the cofactors. *Proceedings of the National Academy of Sciences*, 84(16):5730–5734, 1987.

[67] M. G. Rockley, M. W. Windsor, R. J. Cogdell, and W. W. Parson. Picosecond detection of an intermediate in the photochemical reaction of bacterial photosynthesis. *Proceedings of the National Academy of Sciences*, 72(6):2251–2255, 1975.

[68] R. K. Clayton and H. F. Yau. Photochemical electron transport in photosynthetic reaction centers from *rhodopseudomonas sphaeroides*: I. kinetics of the oxidation and reduction of P-870 as affected by external factors. *Biophysical journal*, 12(7):867, 1972.

[69] U. Andréasson and L.-E. Andréasson. Characterization of a semistable, charge-separated state in reaction centers from *rhodobacter sphaeroides*. *Photosynthesis research*, 75(3):223–233, 2003.

[70] F. van Mourik, M. Reus, and A. R. Holzwarth. Long-lived charge-separated states in bacterial reaction centers isolated from *rhodobacter sphaeroides*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1504(2):311–318, 2001.

[71] V. N. Kharkyanen, Y. M. Barabash, N. M. Berezetskaya, E. P. Lukashev, P. P. Knox, and L. N. Christophorov. Peculiarities of light-induced slow protein dynamics in the photosynthetic reaction center. *Chemical Physics Letters*, 512(1):113–117, 2011.

[72] M. Stowell, T. McPhillips, D. Rees, S. Soltis, E. Abresch, and G. Feher. Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. *Science*, 276(5313):812–816, 1997.

[73] G. Fritzsch, J. Koepke, R. Diem, A. Kuglstatter, and L. Baciou. Charge separation induces conformational changes in the photosynthetic reaction centre of purple bacteria. *Acta Crystallographica Section D: Biological Crystallography*, 58(10):1660–1663, 2002.

[74] G. Katona, A. Snijder, P. Gourdon, U. Andréasson, Ö. Hansson, L.-E. Andréasson, and R. Neutze. Conformational regulation of charge recombination reactions in a photosynthetic bacterial reaction center. *Nature structural & molecular biology*, 12(7):630–631, 2005.

[75] P. Knox, E. Lukashev, K. Timofeev, and N. Seifullina. Effects of oxygen on the dark recombination between photoreduced secondary quinone and oxidized bacteriochlorophyll in *rhodobacter sphaeroides* reaction centers. *Biochemistry (Moscow)*, 67(8):901–907, 2002.

[76] F. Müh, J. C. Williams, J. P. Allen, and W. Lubitz. A conformational change of the photoactive bacteriopheophytin in reaction centers from *rhodobacter sphaeroides*. *Biochemistry*, 37(38):13066–13074, 1998.

[77] L. Kálmán and P. Maróti. Conformation-activated protonation in reaction centers of the photosynthetic bacterium *Rhodobacter sphaeroides*. *Biochemistry*, 36(49):15269–15276, 1997.

[78] A. O. Goushcha, V. N. Kharkyanen, and A. R. Holzwarth. Nonlinear light-induced properties of photosynthetic reaction centers under low intensity irradiation. *The Journal of Physical Chemistry B*, 101(2):259–265, 1997.

[79] M. Olenchuk and N. Berezetska. Study of the recombination process of light-induced charge separation in reaction centers of purple bacteria under long-term exposition. *Molecular Crystals and Liquid Crystals*, 497(1):121–453, 2008.

[80] D. Kleinfeld, M. Okamura, and G. Feher. Electron-transfer kinetics in photosynthetic reaction centers cooled to cryogenic temperatures in the charge-separated state: evidence for light-induced structural changes. *Biochemistry*, 23(24):5780–5786, 1984.

[81] P. Brzezinski and L.-E. Andreasson. Trypsin treatment of reaction centers from *rhodobacter sphaeroides* in the dark and under illumination: protein structural changes follow charge separation. *Biochemistry*, 34(22):7498–7506, 1995.

[82] I. Smirnova, A. Blomberg, L.-E. Andréasson, and P. Brzezinski. Localization of light-induced structural changes in bacterial photosynthetic reaction centers. *Photosynthesis research*, 56(1):45–55, 1998.

[83] L. N. Christophorov and V. N. Kharkyanen. Synergetic mechanisms of structural regulation of the electron transfer and other reactions

of biological macromolecules. *Chemical Physics*, 319(1-3):330 – 341, 2005.

[84] A. Warshel, Z. Chu, and W. Parson. Dispersed polaron simulations of electron transfer in photosynthetic reaction centers. *Science*, 246(4926):112–116, 1989.

[85] J. N. Gehlen, M. Marchi, and D. Chandler. Dynamics affecting the primary charge transfer in photosynthesis. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 263(5146):499–501, 1994.

[86] I. A. Balabin and J. N. Onuchic. Dynamically controlled protein tunneling paths in photosynthetic reaction centers. *Science*, 290(5489):114–117, 2000.

[87] H. R. Treutlein, M. A. Lemmon, D. M. Engelman, and A. Brunger. The glycophorin A transmembrane domain dimer: Sequence-specific propensity for a right-handed supercoil of helixes. *Biochemistry*, 31(51):12726–12732, 1992.

[88] K. Schulten and M. Tesch. Coupling of protein motion to electron transfer: Molecular dynamics and stochastic quantum mechanics study of photosynthetic reaction centers. *Chemical Physics*, 158(2):421–446, 1991.

[89] B. H. McMahon, J. D. Müller, C. A. Wraight, and G. U. Nienhaus. Electron transfer and protein dynamics in the photosynthetic reaction center. *Biophysical journal*, 74(5):2567–2587, 1998.

[90] H. Wang, S. Lin, J. P. Allen, J. C. Williams, S. Blankert, C. Laser, and N. W. Woodbury. Protein dynamics control the kinetics of initial electron transfer in photosynthesis. *Science*, 316(5825):747–750, 2007.

[91] C. Kirmaier, D. Holten, and W. W. Parson. Temperature and detection-wavelength dependence of the picosecond electron-transfer kinetics measured in *rhodopseudomonas sphaeroides* reaction centers. resolution of new spectral and kinetic components in the primary charge-separation process. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 810(1):33–48, 1985.

[92] A. Warshel and W. W Parson. Dynamics of biochemical and biophysical reactions: insight from computer simulations. *Quarterly reviews of biophysics*, 34(04):563–679, 2001.

[93] D. J. Vocadlo, G. J. Davies, R. Laine, and S. G. Withers. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*, 412(6849):835–838, 2001.

[94] J. Cheetham, P. Artymiuk, and D. Phillips. Refinement of an enzyme complex with inhibitor bound at partial occupancy: Hen egg-white lysozyme and tri-n-acetylchitotriose at 1.75Å resolution. *Journal of molecular biology*, 224(3):613–628, 1992.

[95] A. Fleming. On a remarkable bacteriolytic element found in tissues and secretions. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 93(653):pp. 306–317, May 1922.

[96] C. Blake, D. Koenig, G. Mair, A. North, D. Phillips, and V. Sarma. Structure of hen egg-white lysozyme. a three-dimensional fourier synthesis at 2 angstrom resolution. *Nature*, (206):757–61, 1965.

[97] U. Buontempo, G. Careri, P. Fasella, and A. Ferraro. Far-infrared spectra of some globular proteins. *Biopolymers*, 10(12):2377–2386, 1971.

[98] K. Moeller, G. Williams, S. Steinhauser, C. Hirschmugl, and J. C. Smith. Hydration-dependent far-infrared absorption in lysozyme detected using synchrotron radiation. *Biophysical journal*, 61(1):276, 1992.

[99] J. A. McCammon, B. R. Gelin, M. Karplus, and P. G. Wolynes. The hinge-bending mode in lysozyme. *Nature*, 262:325–326, 1976.

[100] R. E. Bruccoleri, M. Karplus, and J. A. McCammon. The hinge-bending mode of a lysozyme-inhibitor complex. *Biopolymers*, 25(9):1767–1802, 1986.

[101] D. A. Turton, H. M. Senn, T. Harwood, A. J. Lapthorn, E. M. Ellis, and K. Wynne. Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature Communications*, 5:3999, June 2014.

[102] L. Genzel, F. Keilmann, T. Martin, G. Wintreling, Y. Yacoby, H. Fröhlich, and M. W. Makinen. Low-frequency Raman spectra of lysozyme. *Biopolymers*, 15(1):219–225, 1976.

[103] H. Urabe, Y. Sugawara, M. Ataka, and A. Rupprecht. Low-frequency Raman spectra of lysozyme crystals and oriented DNA films: Dynamics of crystal water. *Biophysical Journal*, 74(3):1533–1540, 1998.

[104] A. Hédoux, R. Ionov, J.-F. Willart, A. Lerbret, F. Affouard, Y. Guinet, M. Descamps, D. Prevost, L. Paccou, and F. Danede. Evidence of a two-stage thermal denaturation process in lysozyme: a Raman scattering and differential scanning calorimetry investigation. *The Journal of chemical physics*, 124(1):014703, 2006.

[105] D. Tsuru, H. Kira, T. Yamamoto, and J. Fukumoto. Studies on bacterial protease. *Agricultural and Biological Chemistry*, 30(9):856–862, 1966.

[106] B. Matthews, J. Jansonius, P. Colman, B. Schoenborn, and D. Dupourque. Three-dimensional structure of thermolysin. *Nature*, 238(80):37–41, 1972.

[107] D. R. Holland, D. E. Tronrud, H. W. Pley, K. M. Flaherty, W. Stark, J. N. Jansonius, D. B. McKay, and B. W. Matthews. Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry*, 31(46):11310–11316, 1992.

[108] A. C. Hausrath and B. W. Matthews. Thermolysin in the absence of substrate has an open conformation. *Acta Crystallographica Section D: Biological Crystallography*, 58(6):1002–1007, 2002.

[109] D. M. F. van Aalten, A. B. M. Linssen, A. Amadei, V. G. H. Eijsink, G. Vrielid, and H. J. C. Berendsenl. The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins: Structure, Function, and Genetics*, 22:45–54, 1995.

[110] G. Feher and M. Y. Okamura. *The Photosynthetic Bacteria*, chapter Chemical composition and properties of reaction centers in The Photosynthetic Bacteria, pages 349–386. Plenum Press, N.Y., 1978.

[111] E. M. Landau and J. P. Rosenbusch. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proceedings of the National Academy of Sciences*, 93(25):14532–14535, 1996.

[112] H. Qiu and M. Caffrey. The phase diagram of the monoolein/water system: metastability and equilibrium aspects. *Biomaterials*, 21(3):223–234, 2000.

[113] S. Engström, K. Alfons, M. Rasmusson, and H. Ljusberg-Wahren. Solvent-induced sponge (L3) phases in the solvent-monoolein-water system. In *The Colloid Science of Lipids*, pages 93–98. Springer, 1998.

[114] A. Ridell, K. Ekelund, H. Evertsson, and S. Engström. On the water content of the solvent/monoolein/water sponge (L3) phase. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 228(1-3):17–24, 2003.

[115] P. Wadsten, A. B. Wöhri, A. Snijder, G. Katona, A. T. Gardiner, R. J. Cogdell, R. Neutze, and S. Engström. Lipidic sponge phase crystallization of membrane proteins. *Journal of Molecular Biology*, 364(1):44 – 53, 2006.

[116] M. Tonouchi. Cutting-edge terahertz technology. *Nature photonics*, 1(2):97–105, 2007.

[117] C. Rønne, L. Thrane, P.-O. Åstrand, A. Wallqvist, K. V. Mikkelsen, and S. R. Keiding. Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation. *The Journal of chemical physics*, 107(14):5319–5331, 1997.

[118] S. Lupi, A. Nucara, A. Perucchi, P. Calvani, M. Ortolani, L. Quaroni, and M. Kiskinova. Performance of SISSI, the infrared beamline of the ELETTRA storage ring. *JOSA B*, 24(4):959–964, 2007.

[119] P. Roy, J.-B. Brubach, P. Calvani, G. deMarzi, A. Filabozzi, A. Gerschel, P. Giura, S. Lupi, O. Marcouillé, A. Mermet, A. Nucara, J. Orphal, A. Paolone, and M. Vervloet. Infrared synchrotron radiation: from the production to the spectroscopic and microscopic applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 467-468, Part 1:426–436, 2001.

[120] A. Vagin and A. Teplyakov. MOLREP: an automated program for molecular replacement. *Journal of applied crystallography*, 30(6):1022–1025, 1997.

[121] J. Navaza. AMoRe: an automated package for molecular replacement. *Acta Crystallographica Section A: Foundations of Crystallography*, 50(2):157–163, 1994.

[122] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. *Phaser* crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674, Aug 2007.

[123] C. Nave. Radiation damage in protein crystallography. *Radiation Physics and Chemistry*, 45(3):483 – 490, 1995.

[124] W. Kabsch. XDS. *Acta Crystallographica Section D*, 66(2):125–132, Feb 2010.

[125] P. A. Karplus and K. Diederichs. Linking crystallographic model and data quality. *Science*, 336(6084):1030–1033, 2012.

[126] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of coot. *Acta Crystallographica Section D - Biological Crystallography*, 66:486–501, 2010.

[127] B. Rupp. *Biomolecular crystallography: principles, practice, and application to structural biology*, chapter Statistics and probability in crystallography, pages 313–370. Garland Science, 2009.

[128] T. Ursby and D. Bourgeois. Improved estimation of structure-factor difference amplitudesfrom poorly accurate data. *Acta Crystallographica Section A: Foundations of Crystallography*, 53(5):564–575, 1997.

[129] G. Bricogne. A Bayesian statistical theory of the phase problem. I. A multichannel maximum-entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta Crystallographica Section A*, 44(4):517–545, Jul 1988.

[130] R. J. Read. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica Section A: Foundations of Crystallography*, 42(3):140–149, 1986.

[131] T. C. Terwilliger. MAD phasing: Bayesian estimates of FA. *Acta Crystallographica Section D: Biological Crystallography*, 50(1):11–16, 1994.

[132] S. French and K. Wilson. On the treatment of negative intensity observations. *Acta Crystallographica Section A*, 34(4):517–525, Jul 1978.

[133] M. Bayes and M. Price. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. Bayes, frs communicated by mr. Price, in a letter to John Canton, amfrs. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763.

[134] E.-J. Wagenmakers. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5):779–804, 2007.

[135] J. K. Kruschke. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573, 2013.

[136] G. Acbas, K. A. Niessen, E. H. Snell, and A. Markelz. Optical measurements of long-range protein vibrations. *Nature communications*, 5, 2014.

[137] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173 – 181, 1997.

[138] S. Kundu, J. S. Melton, D. C. Sorensen, and G. N. P. Jr. Dynamics of proteins in crystals: Comparison of experiment with simple models. *Biophysical Journal*, 83(2):723 – 732, 2002.

[139] K. M. Tych, A. D. Burnett, C. D. Wood, J. E. Cunningham, A. R. Pearson, A. G. Davies, and E. H. Linfield. Applying broadband terahertz time-domain spectroscopy to the analysis of crystalline proteins: a dehydration study. *Journal of Applied Crystallography*, 44(1):129–133, 2010.

[140] J. A. Rupley, E. Gratton, and G. Careri. Water and globular proteins. *Trends in Biochemical Sciences*, 8(1):18 – 22, 1983.

[141] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith. An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences*, 104(52):20749–20752, 2007.

[142] L. Steinrauf. Preliminary X-ray data for some new crystalline forms of β-lactoglobulin and hen-egg-white lysozyme. *Acta Crystallographica*, 12(1):77–79, 1959.

[143] X. Chen, I. Weber, and R. W. Harrison. Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. *The Journal of Physical Chemistry B*, 112(38):12073–12080, 2008.

[144] J. Ma. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, 13(3):373 – 380, 2005.

[145] L. Meinhold, J. C. Smith, A. Kitao, and A. H. Zewail. Picosecond fluctuating protein energy landscape mapped by pressure–temperature molecular dynamics simulation. *Proceedings of the National Academy of Sciences*, 104(44):17261–17265, 2007.

[146] H. Fröhlich. Long-range coherence and energy storage in biological systems. *International Journal of Quantum Chemistry*, 2(5):641–649, 1968.

[147] J. R. Reimers, L. K. McKemmish, R. H. McKenzie, A. E. Mark, and N. S. Hush. Weak, strong, and coherent regimes of Fröhlich condensation and their applications to terahertz medicine and quantum consciousness. *Proceedings of the National Academy of Sciences*, 106(11):4219–4224, 2009.

[148] J. Pokorný. Conditions for coherent vibrations in the cytoskeleton. *Bioelectrochemistry and Bioenergetics*, 48(2):267 – 271, 1999.

[149] T. Horiuchi and N. Gō. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. *Proteins: Structure, Function, and Bioinformatics*, 10(2):106–116, 1991.

[150] S. Hayward, A. Kitao, and N. Gō. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins: Structure, Function, and Bioinformatics*, 23(2):177–186, 1995.

[151] S. Hayward, A. Kitao, and N. Gō. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein Science*, 3(6):936–943, 1994.

[152] A. Kitao and N. Gō. Investigating protein dynamics in collective coordinate space. *Current opinion in structural biology*, 9(2):164–169, 1999.

[153] A. Nicolaï, P. Delarue, and P. Senet. Low-frequency, functional, modes of proteins: All-atom and coarse-grained normal mode analysis. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, pages 483–524. Springer, 2014.

[154] R. Henderson and J. K. Moffat. The difference Fourier technique in protein crystallography: errors and their treatment. *Acta Crystallographica Section B*, 27(7):1414–1420, Jul 1971.

[155] A. Wilson. The probability distribution of X-ray intensities. *Acta Crystallographica*, 2(5):318–321, 1949.

[156] E. Schrödinger. *What is life? - the Physical Aspect of the Living Cell*. Cambridge University Press, 1944.

[157] P. Weightman. Prospects for the study of biological systems with high power sources of terahertz radiation. *Physical biology*, 9(5):053001, 2012.