Terahertz spectrometry applied to proteins

JUNYI QIU

A thesis submitted to

THE UNIVERSITY OF LONDON

for the degree of

DOCTOR OF PHILOSOPHY

School of Electronic Engineering and Computer Science

Queen Mary, University of London

May 2016

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

Junyi Qiu

09 May 2016

Acknowledgments

I would like to begin these acknowledgements by sincerely thanking my primary supervisor Dr. Robert Donnan for his constant encouragement and support and keep me on track. He is unfailingly patient to give me guidance throughout my Ph.D.

And my deep gratitude goes to my second supervisor, Professor R W Pickersgill, for always being on hand to help me out. Also for his open-mindedness for supporting me to complete this biochemistry based research in his lab.

I would like to extend my appreciation to my research panel members, Dr. Akram Alomainy, for all his help and guidance in my intermediate reports.

This thesis would not be possible without expertise assistance and support from Dr. Bin Yang, from both aspects of techniques and experiments. His unrelenting enthusiasm, patience, and goodwill have inspired me a lot.

I would great appreciate Dr. Oleksandr Sushko, who has spent lots of time working with me on the THz experiments, and kindly taught and shared the knowledge of THz work.

I am very obliged to Dr. Shuang Gu, for her assistance since the start of my Ph.D., and her continuous support - always try to put me up throughout the four years.

I greatly appreciate the assistance of our THz lab manager Dr. Massimo Candotti for always doing his best to meet any inquiry about equipment or time-slots. Also Miss Yumiko Tashiro, for her patiently support and assistance in the biochemistry lab.

Thanks to Mr. Yang Zeng for always being kind and helpful assisting on my experiment.

A special thank goes to Ms. Liang Wang, for helping me proofreading my thesis, and being a great room mate on the emotional journey of the last period of my Ph.D.

A huge thank you goes to my dearest friends and family, especially my parents. Although we are far apart, your persistent encouragement and support are with me at all times.

Finally, I would like to acknowledge the studentship from the College Doctoral Training fund provided by the Queen Mary University of London, and thanks that the school of Electronic Engineering are giving me this opportunity of study.

Abstract

Electromagnetic radiation from the radio waves used in nuclear magnetic resonance spectroscopy through to X-rays used in crystallography have provided a wealth of knowledge about the structure, function, and dynamics of protein molecules. Terahertz waves, the topic of this thesis, are lower in frequency than radiation from the infrared, not to the frequencies of individual bond vibrations, but to the frequency range where slower longer range protein librations (low frequency vibrations) are expected to occur. The role of low frequency protein dynamics remains controversial, with some arguing that these motions are crucial for enzyme and protein function. Terahertz spectroscopy may provide key evidence to contribute to this interdisciplinary debate.

In this thesis, terahertz (THz) spectroscopy has been applied in studying a number of proteins experimentally. In the first results chapter, the effect of protein concentration and ionic strength in the 0.1-2.5 THz region was investigated using Terahertz time domain spectroscopy. The results confirm the presence of terahertz excess for a number of proteins, which results from the increased absorption of THz waves when protein is introduced into the system. THz spectroscopy was then used to detect the difference between a folded protein, myoglobin, and folding intermediates, including the molten globule form, apomyoglobin. The results collected using THz spectroscopy were unable to differentiate between the folded and molten globule states. A further study was susceptible to the formation of higher order protein complexes and explored structures formed using PduA*. These experiments were primarily biochemical in nature with showing that PduA* assembles into nanotubes of 20nm diameter in vitro. The final results chapter explores the sub-THz circular dichroism signal from a vector network analyser driven by quasi-optical circuits. Wherever possible, the THz experiments were benchmarked using established analytical techniques.

Aims and Objectives

The aim of this work was to apply THz spectroscopy to the study of protein structure and function. The objectives were:

(1) To build an experimental apparatus capable of measuring the THz signal from protein solutions;

(2) To assess the THz absorption spectrum as a function of protein concentration and ionic strength;

(3) To correlate the spectrum feature of the THz signal with protein conformation and explore the molten globule state in comparison to the fully folded state of myoglobin

(4) To explore higher-order structures as exemplified by the PduA* assembly into nanotubes and measure their THz spectra;

(5) To explore the THz circular dichroism signal.

These objectives are discussed in detail in their respective results Chapters 3 through 6. Chapters 1 and 2 are introductory, and Chapter 7 brings the work to a brief conclusion.

Detailed outline of the thesis:

Chapter 1 gives an introduction to THz wave interacting with protein structure and dynamics, including conformation change of protein folding and solvent dynamics. A review is conducted of the work of others in this area.

Chapter 2 introduces both the theoretical background of biological methods and the detailed procedures applied. For THz measurements, current techniques employ terahertz time-domain spectroscopy (THz-TDS) and vector network analyser (VNA).

Chapter 3 firstly investigates the concentration response of selected proteins measured on THz-TDS, and then further elucidates protein structure and stability in different ionic environments, measured in the sub-THz spectral domain as well as by circular dichroism (CD).

Chapter 4 presents the hydration dynamics' response of Apomyoglobin (ApoMb) to protein folding states, by combining THz-TDS spectroscopy with CD measurements.

In Chapter 5, filamentous nanotube structures were produced from PduA* *in vitro* and probed using a VNA-driven quasi-optical circuit.

Chapter 6 develops a 220-325 GHz band VNA-CD spectrometer based on the Quasi-Optical bench to explore the correlated transmittance response of spectroscopic signatures with biomolecules in aqueous solution.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ApoMb	Apomyoglobin
BSA	Bovine serum albumin protein
BMCs	Bacterial microcompartments
CD	Circular dichroism
CYC	Cytochrome C
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
Eut	Ethanolamine utilisation
Etu	Ethanol utilisation
FFT	Fast Fourier transform
FIR	Far infrared
FPLC	Fast protein liquid chromatography
Fs	Femtosecond
FT	Fourier transform
Gdm^+	Guanidinium
GdmCl	Guanidinium chloride
HSA	Human serum albumin
HCl	Hydrochloric acid
HCl/KCl	Hydrochloric Acid-Potassium Chloride
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-thiogalactoside
IR	Infrared
KH_2PO4	Monopotassium phosphate
K_2HPO4	Dipotassium phosphate
KITA	Kinetic terahertz absorption

Mb	Myoglobin
MD	Molecular dynamics
MeV	Mega electron Volt
MW	Molecular Weight
NaCl	Sodium chloride
NMR	Nuclear magnetic resonance
PDB	Protein data bank
Pdu	1,2-propanediol utilisation
Phe	Phenylalanine
PTA	Phosphotungstic acid
Ps	Picosecond
PTFE	Polytetrafluoroethylene
QO	Quasi-optical
QMUL	Queen Mary, University of London
Rg	Radius of gyration
SAXS	Small-Angle X-ray Scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S75/200	Superdex 75/Superdex 200
TEM	Transmission electronic microscopy
T_{CO}	Co-polar transmittance
T_{CR}	Cross-polar transmittance
TF	Transfer function
THz	Terahertz
THz-TDS	Terahertz time domain spectroscopy
Tris-HCl	Tris-Hydrochloride
Trp	Tryptophan
Tyr	Tyrosine

TPX	Poly-4-methyl pentene-1
UV-vis	Ultraviolet visible
VCD	Vibrational circular dichroism
VNA	Vector network analyser
ZnTe	Zinc telluride

Chapter 1 **Foundation of protein and its interaction with THz waves**

1.1 THz radiation

1.1.1 The electromagnetic spectrum

Based on a nuclear magnetic resonance experiment, the frequency of a hydrogen nucleus lies in the radio frequency range 60-800 MHz (Figure 1.1) [1]. Further along the frequency range, infrared absorption spectroscopy provides information about bond stretching and bond vibration around 1500 cm^{-1} (45 THz). The visible range serves as a rich source of information about light absorbing biological entities such as haem, chlorophyll and other entities with conjugated double bonds (400-700 THz). Just beyond this is the ultra-violet range where the absorption of the aromatic side chains in proteins is used to determine protein concentration. At shorter wavelengths and higher energies, X-rays have been utilised to identify protein structures from crystals because the wavelength is similar to that of the carbon-carbon bond. The value of spectroscopic studies in the THz range, around 0.1-10 THz, has yet to be established.

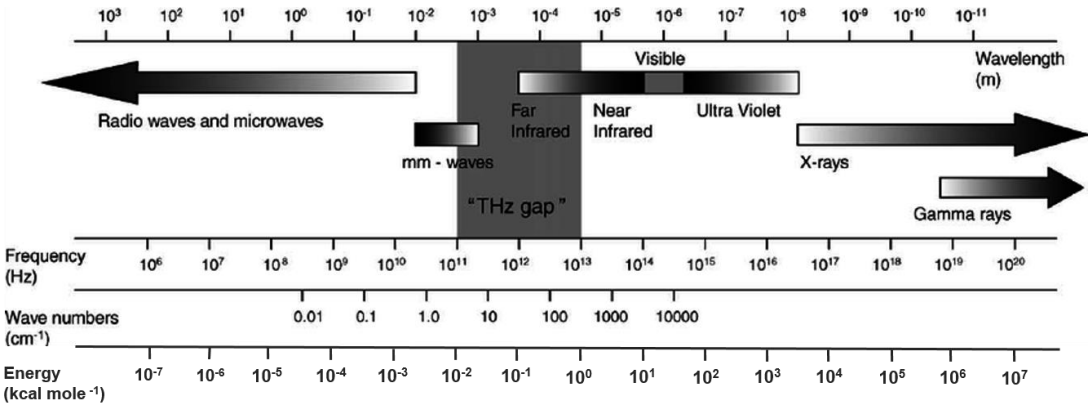


Figure 1.1 The electromagnetic spectrum showing THz frequencies are sandwiched between microwaves and visible light [1].

1.1.2 THz waves

Terahertz waves, which have a commonly used but not strict definition from 0.1-10 THz (shown in Figure 1.1) are of interest in both fundamental research and industrial applications [2–4]. THz techniques have the capacity to serve as an effective instrument in the portrayal of vibrational modes [3]. These methods which are used for the detection of THz radiation can also be utilised in the investigation of physical phenomena, including excitations that involve low energy and carrier dynamics which exist in electronic materials. Other phenomena include combined torsional and vibrational methods in condensed-phase media, as well as vibrational and rotational changes that occur in molecules [5], [6].

The measurement of the electric field amplitude can be conducted directly through THz spectroscopy. It is therefore possible, with the assistance of a Fourier transform, to directly and concurrently acquire the amplitude and phase of the signal that underwent analysis after measuring the sample [7], [8].

Various previous research studies, books and reviews have well explained, in great detail, the THz theory and devices as well as instrumentation [9], [10]. Among its numerous uses, the THz technology is utilised in fields such as nanomaterials [8], defence [11], [12], and industry [13], as well as semiconductors [14], communications [15], [16] and biomedicine [17]. Other fields include pharmaceutics [18], [19], imaging [20], and agro-products [21], as well as food [22].

1.2 Background of protein structure and dynamics

1.2.1 Foundation of protein structure

The four defined classifications of protein structure are primary, secondary, tertiary, and quaternary. The primary structure is composed of a linear sequence of the 20 amino acid residues, which is unique in specific properties and can determine the folding and function of proteins. Based on the propensity of contact with polar solvent, such as water, the residues are classified by hydrophobic/hydrophilic, polar/non-polar and charged/non-charged. As a general rule, the interior of protein structures is most likely composed of hydrophobic amino acids, in order to avoid the contact with the hydrophilic environment *in vivo*. Polar and hydrophilic residues, which are attracted to the polarity of the solvent, are dominated at the protein surface. The protein's backbone, which includes the alpha carbons and the peptide bond, is hydrophilic and is thus not favoured in the hydrophobic core. However, by forming secondary structure elements from the backbone, the hydrophilic NH and CO groups are prevented from exposure towards the hydrophobic surroundings.

The two major secondary structure elements called α-helix and β-sheet, are stabilised by hydrogen bonding between residues that are non-adjacent in sequence. Amino acids such as methionine, alanine, leucine, glutamate and lysine prefer to adopt helical conformations; while amino acid residues containing large aromatic structures (tryptophan, tyrosine and phenylalanine) and Cb-branched amino acids (isoleucine, valine and threonine) have a tendency to adopt β-strand conformations. In α-helices, the hydrogens link to the C=O category of the amino acid are contributed by the NH group (Figure 1.2A). The polypeptide chains (either parallel or anti-parallel) generate β-sheets. The neighbouring residues of the side chains point alternately upwards and

downwards along the β-strand. The linking of hydrogen occurs between neighbouring β-strands to generate a β-sheet (Figure 1.2B). The elements of the secondary structure are frequently packed closely together to form a structural motif. The BMC shell proteins, which are comprised of β-α-β structural motifs (Figure 1.2C), will be addressed in chapter 5.

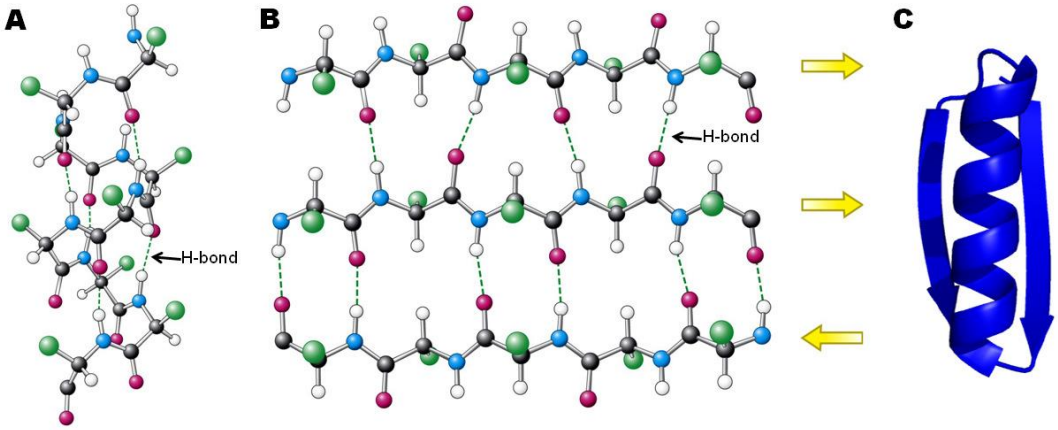


Figure 1.2 Secondary structures. (A) a-helix and (B) β-sheet in a stick and ball presentation. Black circle- carbon atom; Blue circle- nitrogen atom; White circle- hydrogen atom; Red circle- oxygen atom; Green circle- side chains. Hydrogen bonds are depicted as broken green lines; In B panel, neighbouring β-strands can be either in parallel (same direction) or anti-parallel (different direction). (C) Cartoon representation of a right-handed β-α-β motif. Figures are adopted from [23].

The tertiary structure of the protein is built on the base of the secondary structure elements with orientation in space, an example of Pdu subunit tertiary structure which consists of a conserved protein fold was shown in Figure 1.3A. The secondary and tertiary structure uniquely characterises each protein and each is important for its function. Despite the protein structures are complex and may vary from individuals, proteins always attempt to reach the maximum component packing density and thus ensure the resulting 3-dimentional structures are well-packed. This is subject to the

necessity that those amino acid residues having a favourable free energy in how they interact with water, and tending to stay close to the surface of the protein.

After the formation of the resulting protein, protein requires motions at different time scales for their functions, and thus results the molecular dynamics. The quaternary structures shown in Figure 1.3B and C are examples of PduA and PduB subunits assemble into hexamer and three-fold symmetry. Some examples of functions requiring molecular dynamics are protein folding/unfolding, catalysis, membrane transportation, transcription, and protein-protein complex formation. Among these, studying the vital cellular process of protein folding/unfolding is important for the purpose of performing a particular biological function and leading to significant advances in our understanding of the fundamental insights into the protein behaviour.

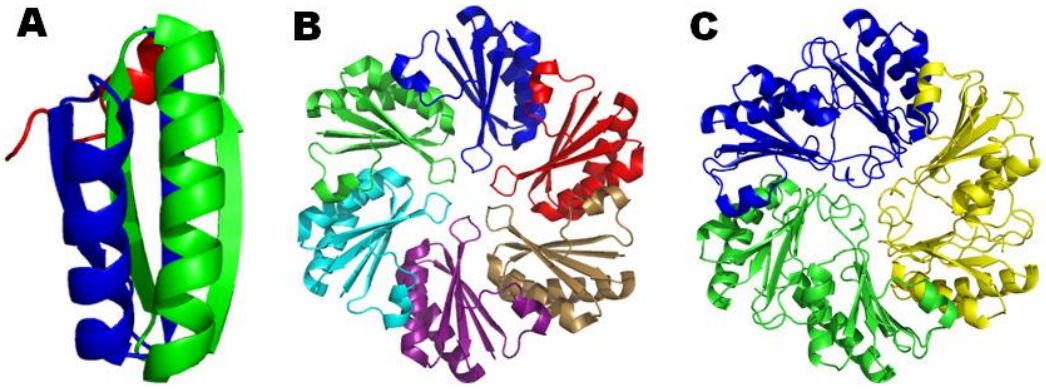


Figure 1.3 Examples of the tertiary and quaternary structure. Tertiary structure of Pdu subunit which consists of a conserved protein fold; (B) PduA subunits further assemble into a six-fold hexamer (PDB: 3NWG); (C) PduB subunits, which are comprised of two Pdu subunit, assemble into a trimer, with a three-fold symmetry (PDB: 4FAY). Figures are adopted from [24].

1.2.2 Protein folding and denaturation

Proteins have a tendency to fold to a certain conformation. Proteins in their native state (folded) are generally of a greater stability and are more favourable energetically (possess lower free energy) than their unfolded condition. This procedure by which proteins achieve their stable three-dimensional conformation is called protein folding. The necessity of protein conformation to physiological function has been studied extensively [25], [26]. Protein will lose its function in the denaturing process under certain conditions, such as adding denaturant agents (GdmCl and urea, for example), treating with high temperature or changes of the pH [27], [28]. The process known as denaturation is a disruption of the native state which leads to the unfolding of the proteins. An unfolded protein is non-functional owing to the lack of arrangement that the functional secondary structures turn into random coils. The unfolded protein tends to aggregate when it encounters lower levels of energy and greater stability. The aggregation of proteins can lead to a number of structural appearances with intermediates (oligomers) which range from unordered amorphous aggregates to highly ordered fibrils, e.g. amyloid fibrils, as shown in Figure 1.4 [29]. Although the amyloid fibrils vary in time, sequence and conditions, they are enriched in the cross-β structure [30]. The majority of the structural fluctuations of amyloid proteins are linked with protein misfolding diseases, for example, the Alzheimer's disease [36].

Far-reaching theoretical and experimental investigations have focused on the means by which a protein folds to different conformations, monitored by circular dichroism, tryptophan fluorescence, IR spectroscopy, X-ray diffraction, NMR spectroscopy, and mass spectrometry [31], [36]. However, obtaining basic knowledge of the molecular mechanism responsible for the folding of these flexible polypeptide chains is still one

of the biggest challenges of structural biology: the relationship between the polypeptide sequence and the folding behaviour of a protein is not entirely understood yet. The relationship between structure and function is essential to determining functional and structural domains, such as their relevance to the protein stability [32], [33]. The crucial nature of the amino acid sequence of a protein is justified by the theory of which all the information, needed for the folding of a protein, is based on this sequence [32], [33]. Thus, the knowledge gained through other studies will help to understand the complicated question of how the folding process occurs.

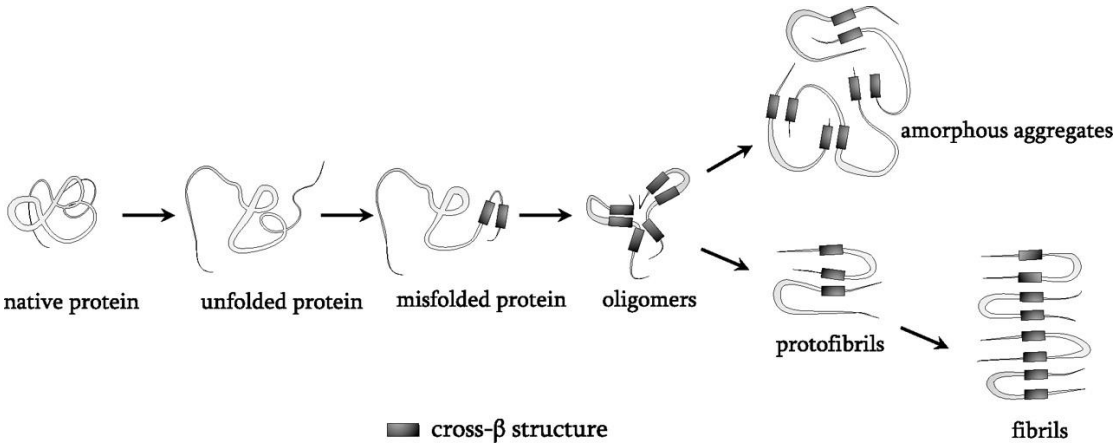


Figure 1.4: Representation diagram of protein misfolding and aggregation [30]. *In extreme circumstances of temperature change or pH, proteins experience conformational changes that cause unfolding and partial misfolding which is connected with the tendency to aggregate. Proteins can acquire a scope of various structural appearances during the aggregation process, and these are enriched in the cross-β structure and include intermediates which vary from unordered amorphous aggregates to ordered fibrils known as amyloid fibrils.*

1.2.3 Protein hydration dynamics

The dynamics of protein solvation are the motions of water in the proteins' hydration shell in aqueous solution. Protein have evolved to perform functions mostly in aqueous environments. Water has been discovered playing a key role in protein structural organization, performing as a driving force that stimulates proteins to equilibrate their desired functional state rather than a mere passive solvent in biological settings. A water cluster was first measured by Pugliano and Saykally with Terahertz Laser Vibration-Rotation-Tunnelling (VRT) spectroscopy in 1992 [164], then the studies at Berkeley by Richard Saykally's group produced much more extensive spectra of the dimer through hexamer clusters, which demonstrated the highest resolution water molecules structures using THz laser technique [165]. In recent years, researchers have defined the water in proteins into three categories, namely internal water, hydration water and bulk water. Internal water refers to the water molecules found in the interior of the protein; the bulk water, also called free water sometimes, is presented as the water that are far away from the dissolved protein; while the hydration water involves water molecules that interact with the surface of the proteins [31]. Many research focused on studying the aqueous protein molecules have highlighted the importance of protein-water interaction, especially the "hydration water" [32], [33]. As a fact that the hydration water is known as existing near a protein surface and relects the perturbation of protein conformations by interacting with the polar groups on the surface of the protein from hydrogen links, as well as surrounding solvent molecules [34].

It has been found that the rate of breaking a hydrogen bond between any two water molecules within the hydration layer is averagely slower than that in the bulk water,

and this rate often differs by more than two orders of magnitude in the timescales. However, as the fact that the protein surface is such a complex and heterogeneous system to study, it is difficult to determine the lifetime of hydrogen bonds within the hydration layer around the protein with a single probing tool or method [35–39]. Although the protein surface extensively disrupts the hydrogen bond, water is still capable of creating chained hydrogen bond network/cluster with two dimensions around the protein [40], [41]. Because of its interaction with the hydrophilic and hydrophobic groups of the protein, this two-dimensional grid itself is controlled. Nonetheless, because of the complexity of the protein surface, the network will have varying rigidity in various surface's regions.

Both the backbone and the side chains, which constitute two different parts of the protein, are vital to the identification of the protein's folded 3-dimensional structural topology [42], [43]. The backbone chain accounts for 60% of the polar atoms of the surface on average, which are composed of polar amide groups. In the case of the folded state, while an inflexible grid is created by the backbone atoms, the side chain will continue mobile in order to supply the structure with flexibility. As highlighted by research studies conducted recently, the protein's side chain being rigid plays a significant role in the dynamics of the hydration layer [37–39]. All the structures and dynamics related to the hydrogen links within the water molecule and between water and protein are affected by the extended hydrogen-linked grid that exists around the protein. The dynamical behaviour of the protein's hydration layer is caused by all these interactions which are also behind the timescale of the side chain's atom motion [40]. So far, there has been no quantitative investigation of these aspects, by using single experimental or theoretical tools [41]. However, it is possible to address certain

features related and simplify the complex dynamics with multitude of techniques and complementary methods [42], [43].

1.3 Interactions between THz wave and protein dynamics

Biological molecules interact with radiation of different frequencies with characteristic differences. Certain energies in the visible and ultraviolet region can cause electrons to be excited to higher energy orbitals. If the energy of a photon is sufficient, e.g. X-rays [44], the molecule may be dissociated or ionized. Ionization is particularly harmful to organic molecules since it creates chemically active radicals, which can result in damage to other molecules [45]. The photon energy in the THz spectral domain is much lower than that of the visible or ultraviolet region and can only excite vibrations in molecules without causing ionization effects in biological systems [46]. Therefore, this low (MeV) energy level gives THz waves the advantage of non-destructive inspection of fragile biological samples such as protein and DNA.

In aqueous conditions, the rotational relaxation time of an average-sized protein is of the order of nanoseconds, faster than that of the dry state [67]. THz wave possesses a period of oscillation of 1 picosecond, which is three orders of magnitude faster than the typical relaxation time of a protein, therefore the alignment of its dipole moment lags the phase of the driving field and each individual dipole moment of a protein is effectively made stationary at THz frequencies [47]. Within the THz range of frequency, there exist a great number of molecular vibrations; such vibrations are inclusive of crystalline phonon vibrations, low-frequency bond vibrations, as well as molecular rotations [48]. THz waves have the ability to supply useful data with respect

to protein structure [49] because they experience interaction with organic molecules' transitions, both rotational transitions and vibrational transitions. In the THz band, both amplitude and phase information can be retrieved from the direct measurements, therefore a variety of coefficients of absorption as well as refractive indexes are likely to be exhibited by differing molecules [50] or by the very same molecules within varying conformations.

There are great potentials for THz spectroscopy to become an indispensable means of comprehending protein dynamics, as well as for the observation of conformational transformations through analysing vibrational modes that are localised and also dependent on the complete framework.

1.4 Application of Terahertz on protein dynamics

In the past 20 to 30 years, many types of research have been applied to the employment of THz spectroscopy to study protein over a broad frequency domain from GHz to THz [51–53]. These measurements were acquired using various THz techniques, including time-domain spectroscopy (THz-TDS) [54] and Fourier transform interferometry (FTIR) [45], [55] in the range up to a few THz, for covering the measurements in a broader frequency range; and spectroscopy employed by narrow-band, but high-power laser, such as free electron laser and the tunable p-Ge lasers [56], to retrieve sufficient information in a more precise manner. All these approaches allow researchers to determine the complex absorption directly from the magnitude of the transmission and phase information. Compared to the well-established analytical techniques on the protein solution study, such as FIR spectroscopies, X-ray scattering

and NMR, THz and sub-THz spectroscopies is still in the early developing stage [57]. However, THz techniques are capable of delivering some specific pieces of unique experimental data for protein analysis from the picosecond to sub-nanosecond order, due to the sensitivity of THz waves to the intra-molecular vibrational modes of proteins.

Many studies have attempted to address the protein-related THz response, which is mainly associated with protein conformation, hydration and ligand binding in solution. Markelz's group was one of the earliest demonstrating that low-frequency vibrational modes of proteins are optically active in the terahertz frequency range [58–60] and has done a number of studies in this research field. They have applied pulsed terahertz spectroscopy and THz-TDS studying the 0.06-2 THz frequency range to identify biomolecular species, conformation states, and mutations in bacteriorhodopsin. In these studies, Biomolecules samples were prepared in either lyophilized powder by freeze-drying, or in hydrated form then pressed into pellets. However, it must be noted that lyophilisation changes protein structure and removes function in many cases [166]. The use of pressed pellets can also be limited in THz measurements, for the measurement of THz dielectric response evolution with biological function, samples must be sufficiently hydrated and allow proteins perform in their comfortable state to monitor the THz response from the bulk of the sample, not only characterize the surface of pellets [54].

THz studies have also been applied to large biomolecules such as DNA and RNA [173], some excellent THz measurements has been performed using FTIR spectrometer in a series studies by T. Globus and co-workers [167-169]. They investigated THz spectra

of single- and double- stranded DNA and RNA in both dry and liquid gel phase, samples were prepared in manners of films that sealed between thin polyethylene films of varies thickness from 50 to 300 μm. It was reported that the THz transmission between native (double-stranded) and denatured (single-stranded) forms of DNA varies and can be determined from four THz absorbance features [169]. It is again noted the extreme sensitivity of the spectra to sample preparation including thickness, rotational orientation and water contents in sample [169].

1.4.1 Protein Hydration

The role of water in protein dynamics has been studied by an increasing number of groups, due to the importance of water playing in the protein function. In the early experiments, many studies have been restricted to dehydrated samples because the THz spectra are mainly featureless due to the high hydrated solutes THz absorbance [61], [62]. The loss of feature is mainly caused by the high THz absorbance emanating from the rotational relaxation within bulk water under uncontrolled thickness [63]. However, most of these measurements results are not related to protein dynamics characterization, mainly due to the thickness is not precisely controlled in these experiments. Meanwhile, protein samples must be sufficiently hydrated in order to perform biological functions on the dielectric response [61].

Xu *et al.,* [64]measured the absorption spectrum of solvated BSA and lysozyme from 0.3-3.72 THz and observed a linear THz absorption change by increasing protein concentration. Vinh *et al.,* [65] studied the lysozyme in the sub-THz region, driven by a vector network analyser based spectroscopy covering 65 to 700 GHz. They discovered that the concentration changes and the dielectric-defined hydration shell of

lysozyme contained 165±15 water molecules. Latterly, the technique of merging molecular dynamics simulations with THz spectroscopy was used by Havenith's group [66] in researching protein–hydration coupling. The researchers have mainly used a p-Germanium laser with a narrow frequency range of 2.1-2.7 THz. By investigating the 2.4 THz absorption of native and denatured ubiquitin and λ-repressor proteins [67], native conformations clearly exhibited a higher THz response for both proteins. This is attributed to the modified water dynamics in the hydration shell of the unfolded state and suggests the major contribution to absorption of THz radiation is the hydration level in the protein. Accordingly, Havenith and co-workers introduced the three-component model (which consists of protein, bulk water and hydration water) into the protein solvation interaction study, and determined the size and absorption of the hydration shell [54], [64], [66], [171]. THz spectrometer was also used as a tool to probe the collective hydration dynamics of ions and proteins, especially the enzymatic hydrolysis, with the help of kinetic terahertz absorption (KITA) spectroscopy [172].

Molecular dynamics (MD) simulations have often been used as an approach to accompany experimental studies on THz providing means for interactions between protein and water [68], [69]. Normal mode analysis (NMA) provides simulation methods on the low-frequency modes of larger molecules [46]. Appropriate MD simulations complemented experimental studies. It relates THz excess of proteins to an increased vibrational density of states of hydration-water over that of bulk water in the frequency domain above 55 cm^{-1} (1.65 THz) [49]. In a combined experimental and MD study of aqueous peptides, Ding *et al.*, [49] used the vibrational density of states to assign and analyze contributions of different structural elements to the absorption spectrum.

Further researches from Havenith's research group [66] were carried on kinetic studies on protein folding by mixing an unfolded protein with a denaturant-free buffer and monitored the THz absorption changes in real time, and further compared the recorded result with fluorescence method, circular dichroism (CD) spectra, small angle X-ray scattering (SAXS) and simulated by MD method, indicating THz can be used as a new approach in tracking protein folding process. For higher ordered tertiary structure of proteins like nanostructures and amyloid fibrils, relative work has not been systematically studied on THz domain except for Liu *et. al.*, [70] using THz-TDS demonstrated a difference in absorbance between the native insulin and the insulin fibrils between 0.2-3.0 THz.

1.4.2 Conformation Change

THz-TDS was used to examine the denaturation of the PsbO protein [71] and two photosynthesis membrane proteins: CP43 and CP47 [72–74]. THz-TDS was proven to be useful for discriminating the different conformational states of proteins with similar structures and for monitoring the denaturation process of proteins.

Regarding the denaturation aspect, the sensitivity of the THz radiation to the conformational state of photoactive yellow protein was observed by Castro-Camus *et al.*, [50]. A clear increasing trend of THz absorption was observed in 0.25 -2 THz range when the native protein turned to unfold partly after being illuminated with blue light at 450 nm. Yoneyama *et al.*, [75] measured BSA in its native and thermally denatured conformations in the terahertz frequency range. The researchers used a membrane device to hold the samples and found that the transmittance amplitude of a

thermally denatured BSA sample was notably higher than that of a BSA sample in the native conformation. Clear phase differences also existed. The results suggested that the membrane device is useful for observing protein conformational changes. Similarly, the thermal denaturation of solid BSA using terahertz dielectric spectroscopy was recently investigated by Li *et al.*, [76].

Havenith's group has conducted important research on protein interactions. Their research showed that terahertz was sensitive specifically to solvation water around proteins [56], [61]. When they investigated the thermal denaturation of human serum albumin and the associated solvation using terahertz spectroscopy in an aqueous buffer solution, a clear correlation was observed between structural changes and changes in the hydration dynamics for the human serum albumin (HSA) protein [77].

1.5 Challenges of THz technologies in protein studies

The major hurdle facing terahertz applications in protein detection is the overwhelming attenuation of the terahertz radiation by water, which is an enormous obstacle. Terahertz technology is not suitable for moisture measurements of high-moisture products with a thickness greater than 1 mm. The inability of terahertz technology to monitor biomolecular interactions in solution is a serious limitation.

The THz signal are sensitive to many environment factors such as temperature and because of the sensitivity to a change in environment factors, the sensitivity of the terahertz system is still lacking. Therefore, improving the instrumental signal-to-noise ratio is one the challenges facing terahertz technology. Temperature and attached water

molecules are the main external parameters that influence the dynamics and thus the functionality of proteins; therefore, controlling these parameters is indispensable for attaining reproducible and reliable results [78].

1.6 Summary

Terahertz spectroscopy provides a unique perspective in biological study in having the capacity to probe long-ranged dynamics of collective networks and displays fast dynamics, which average out in many static or scattering experiments [79]. Currently there are a noticeable number of THz studies on protein hydration dynamics, as THz dielectric response has contributions from vibrational and diffusive motions in the protein and adjacent solvent. Broadband absorption that increased with frequency was observed for lyophilized powder samples of BSA and collagen. Studies on conformation have applied on many proteins such as hen egg white lysozyme (HEWL), different conformations of bacteriorhodopsin (BR), two states of photoactive yellow protein (PYP) [21].

Although the various applications of this technology in different fields show great potential, scientific and technological issues must still be addressed. Some difficulties should be resolved for protein detection applications.

Chapter 2 **Methodology: Biological approaches and THz techniques**

2.1 Materials

2.1.1 Proteins

Six proteins are chosen for carrying on different experiments in this study: bovine serum albumin (BSA), lysozyme, cytochrome C (CYC), myoglobin (Mb) and apomyoglobin (ApoMb) proteins from equine skeletal muscle and the mutated hexametric shell protein PduA*, they were introduced in separate chapters. BSA, lysozyme, and Mb were purchased from *Sigma* Aldrich in the form of lyophilized powder with a purity of 95%-100%; all proteins are purified before use. ApoMb was obtained from the above purified Mb by a heme-group removal step using the 2-butanone extraction method [80]. PduA* DNA was supplied by Dr. Allen Pang from QMUL and cloned into pET14b plasmid vector (Appendix 1).

2.1.2 Buffers

The spectroscopic techniques used in this study have some influence on the choice of buffer. The protein folding studies in Chapter 4 were typically performed in double-distilled (18 MΩ·cm) water in the pH range pH 2 to 7 in 20mM buffer solutions, additional buffers for different pHs are: phosphate buffer (KH_2PO_4/ K_2HPO_4) for pH 6 and 7; Citric acid/sodium citrate for pH 3-5, Hydrochloric Acid-Potassium Chloride Buffer (HCl/KCl) for pH 2, as listed in Table 2-1. Subsequently the pH of the buffer solutions were adjusted by adding a small amount of 0.1M hydrogen

chloride/ sodium hydroxide (<2μL). The nanostructure studied in Chapter 6 was initially prepared in 50 mM Tris, 500 mM or 50 mM NaCl buffer solution at pH 8.5.

Buffer solutions	**Desired pH range**	**Proteins prepared**
Hydrochloric Acid-Potassium Chloride (HCl/KCl)	1.0-2.2	ApoMb at pH 2
Citric acid/sodium citrate	3.0-5.0	ApoMb at pH 3, 4, 5
potassium phosphate buffer (KH_2PO_4/ K_2HPO_4)	5.8-8.0	ApoMb at pH 6, 7; BSA/Lysozyme/Myoglobin at pH 7
Tris-Hydrochloride (Tris-HCl)	8.0-10.2	PduA at pH 8.5

Table 2-1 Buffers used in the thesis.

2.1.3 Sample holder

A Bruker liquid cell (A145), as represented in Figure 2.1, was utilized as a sample holder for the protein solutions. The cell consists of two poly-4-methyl pentene-1 (TPX) plates and a 100 μm polytetrafluoroethylene (PTFE) spacer [174]. The liquid cell is not opened while refilling with liquid samples. For each filling, 50 μl of the solution was injected into the cell with a pipette. Before each measurement, it was firstly filled with distilled water or buffer and pumped out with a syringe, which was repeated twice. Then repeat the fill and empty procedure again with sample solution to equilibrate the cell before the final injection of the sample. This procedure will ensure the thickness of the samples is constant during the experiments.

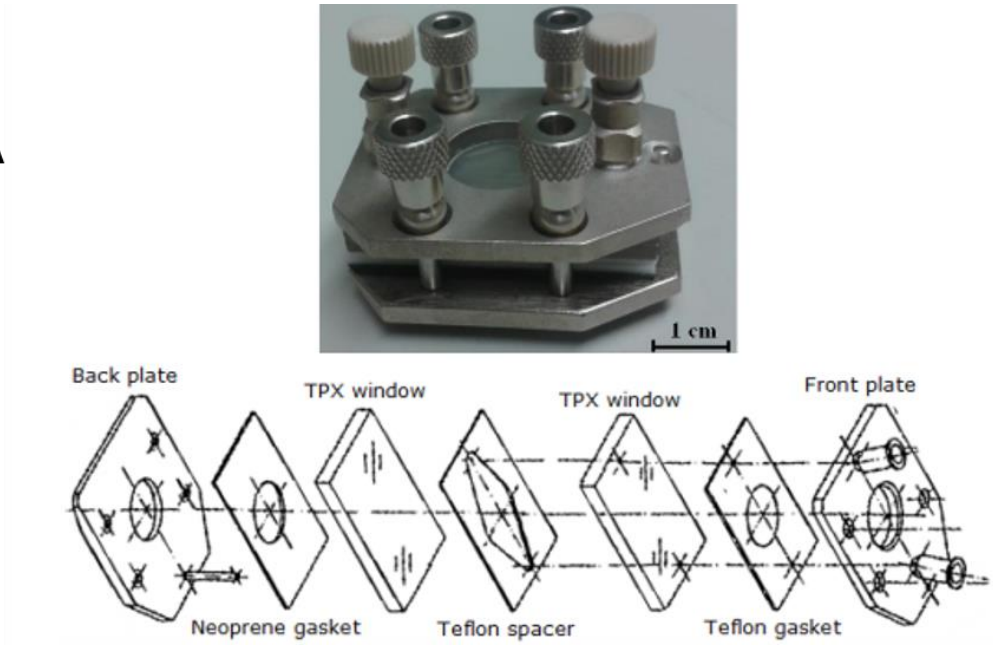


Figure 2.1 The liquid sample holder from Bruker company was used in our THz measurements. (A) The picture of the cell; (B)The alignment of components of the cell.

2.2 Biological approaches

2.2.1 Gel filtration chromatography

This method was performed for protein separation according to their size and shape. The gel filtration column consists of spherical and porous Sephadex beads, allowing for the larger proteins to move through the interstitial spaces and elute first while the smaller protein molecules are retained in the free space in the beads, as shown in Figure 2.2.

In this work, gel filtration chromatography was used as the only protein purification step for BSA, lysozyme, and myoglobin, and the final purification step for apomyoglobin and PduA*. Before application to the gel filtration column, the protein solution was concentrated in a Vivaspin concentrator (Sartorius Stedim Biotech, 5K Da or 10K Da MW cut off) until a final volume of 500 µl to 1ml was reached. The concentrated sample was loaded into a 2 ml loop connected to a Superdex 75 10/300

gel filtration column (GE Healthcare), which was pre-equilibrated with desired final buffer. An automated AKTA FPLC was used to run the chromatography at a rate of 0.5 ml/min with 1 column volume of buffer. The target was collected in 1 ml fraction and analysed by SDS-PAGE. Fraction samples containing the protein of interest were pooled and concentrated for use in further experiments.

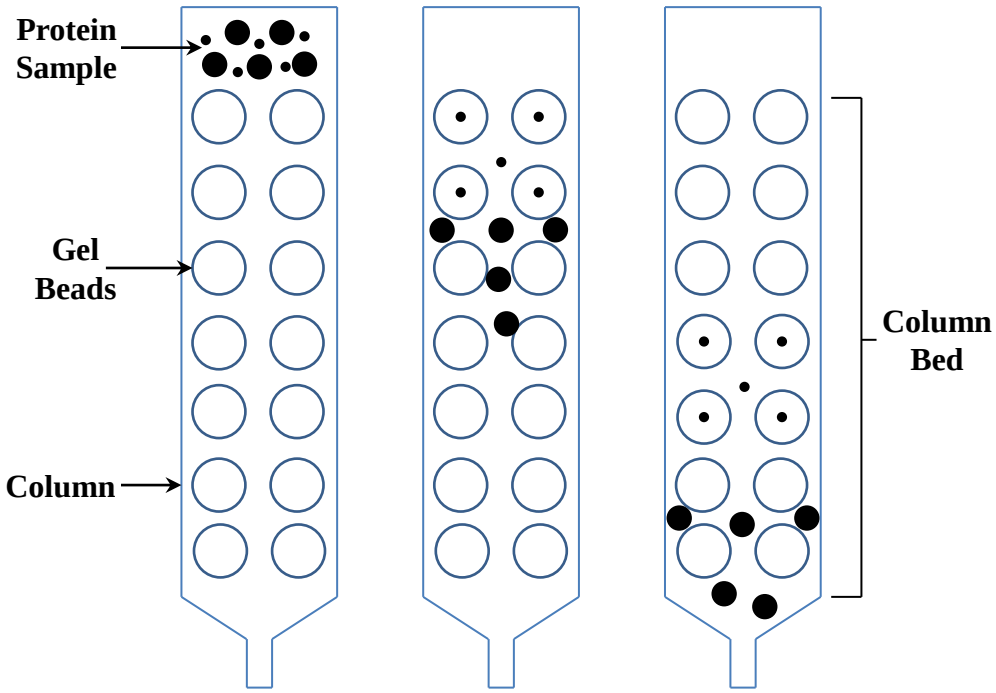


Figure 2.2 A schematic highlighting the separation process by gel filtration chromatography. When a protein sample is added (first column), the larger protein molecules move faster down the column while the smaller protein molecules are retained behind by the beads (second column) and finally the larger protein molecule is eluted first (column three).

2.2.2 Protein analysis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

The purity of a protein sample was analysed using the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) method [81]. SDS is an anionic detergent that denatures proteins by disrupting its non-covalent bonds, causing the

molecule to lose their native conformation [81]. By using this system, protein migration via gel electrophoresis was determined by molecular weight alone.

The Mini-Protean 3 system from Bio-Rad is used for setting up and running polyacrylamide gel electrophoresis. Glass plates were cleaned with ethanol and allowed to dry before assembling into casting frame and casting stand. The 15% to 20% resolving gels and a 5% stacking gel were made and used for all experiments.

Protein samples were supplemented with loading buffer and incubated at 100 ^{o}C for 5 minutes before loading. Samples were loaded with pre-stained Protein Marker, Broad Range (New England BioLabs) to estimate the molecular weight of protein bands. The electrophoresis was run at constant voltage (200 V) until protein bands migrate to the bottom. Gels were viewed after staining and de-staining.

Protein concentration measurement

This technique uses the ultraviolet (UV), 180 to 400 nm, and visible (Vis) spectrum, 400 to 800 nm. It is a particularly useful approach to determine the concentration of proteins. Side-chains belonging to tryptophan (Trp), tyrosine (Tyr), phenylalanine (Phe) and a disulphide bond act as concentration monitors of proteins at 280 nm. Using the Beer-Lambert law (equation 2.1) which highlights the relationship between absorbance and concentration of protein:

$$A = \varepsilon . c . l \tag{2.1}$$

where A is the absorbance, c is the molar concentration of protein; ε is the extinction coefficient of the side-chains and l is the path length in centimetres (cm).

The Hitachi U-3010 UV-Visible spectrophotometer was first calibrated with buffer. A test scan was recorded ensuring that the buffer gave a zero reading. The protein solution, usually at a dilution factor of 50, was measured. As indicated in the Beer-Lambert law, the extinction coefficient must be determined to calculate the protein concentration. This can be obtained by placing the amino acid sequence into PROTPARAM tool of the EXPASY proteomics server (http://www.expasy.ch), which calculates based on the tyrosine and tryptophan residues [82].

As the heme group interacts with the rest of the protein structure to create an absorbance peak at 409 nm, known as the Soret band [83], the change was used to determine the purity of apomyoglobin after heme group removal from myoglobin protein [chapter 4].

2.2.3 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique used to gather information regarding the secondary structure of proteins in solution. As opposed to linearly polarised light, which oscillates only in a single plane, circularly polarised light has a continuously rotating plane of oscillation, forming a helix as it travels. This helix can turn either in a clockwise direction producing a right circularly polarised light or in an anti-clockwise direction to generate a left circularly polarised light [84]. A CD spectrophotometer measures the differential absorption of left and right circularly polarised light.

The far-UV (185-250 nm) range is essential for measuring secondary structural changes of protein, while near UV (>250 nm) CD spectrum is used to acquire

information on the tertiary structure of the protein. Far-UV-CD uses the two electronic absorptions of the backbone amide group, the electronic dipole from the π-π* transition at 190 nm and the magnetic dipole from the weaker *n*-π* transition at 210 nm. These transitions dominate the CD spectrum to produce a trace characteristic of the ψ (psi) and φ (phi) torsion angles, therefore, are characteristic of secondary structure of the polypeptide chain, in a result providing an estimate fraction of a molecule that has α-helical conformation, β-sheet conformation, and random coil, as summarized in Figure 2.3 [84]. Although CD cannot pinpoint where the detected secondary structure is within the molecule, it can be a valuable tool in assessing whether the protein sample is properly folded.

Due to the versatile nature of far-UV-CD, it has been used in this thesis to measure changes in the secondary structure under different pH conditions. Considering the best signal to noise ratio spectrum, a 0.2 ml diluted protein sample with a concentration of about 0.1 mg/ml is placed in a quartz cuvette with a 1 mm path length and subsequently placed onto the Chirascan CD spectrophotometer (Applied Photophysics). Absorbance scans were carried out from 190 to 260 nm at 0.5 nm intervals at a set temperature of 20 $^{\circ}$C. For each sample scan, three trials were taken and then averaged. An absorbance scan of the buffer was also taken and used as a baseline to deduct the absorbance scan of samples. Data acquired from CD were analysed using CDNN (Circular Dichroism analysis using Neural Networks) deconvolution software [84], which interprets CD spectra and employs a complex algorithm to provide predictions of proportions of each secondary structure characteristic present.

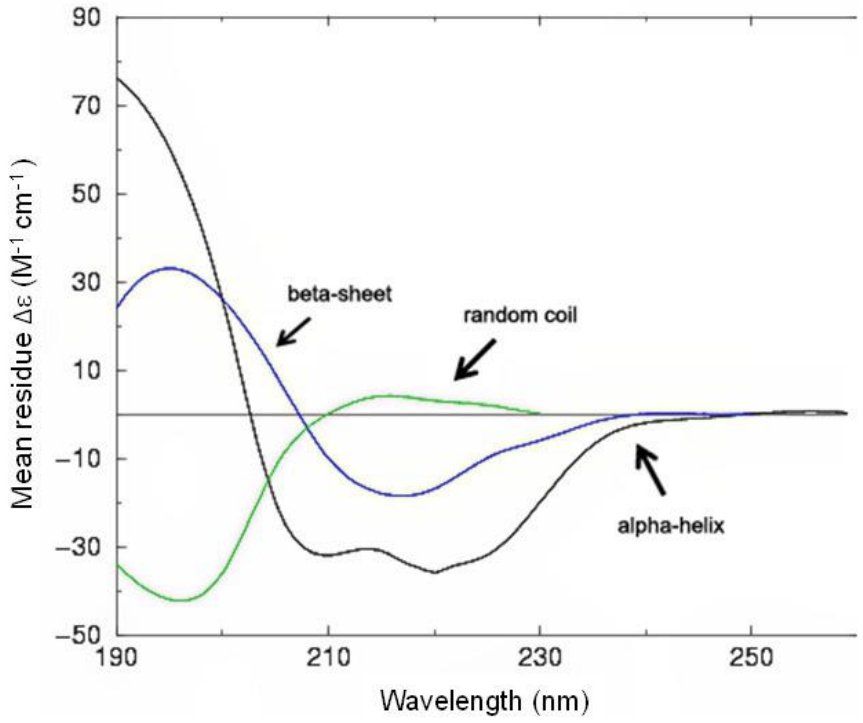


Figure 2.3 Reference CD spectra for various secondary structure conformations. The α-helix (black) has a characteristic defined with two negative bands of similar magnitude in the wavelength 208 and 222 nm. β-sheet (blue) has a CD spectra fingerprint of a negative band at 217 nm. Random coil (green) exhibits a negative band at about 195 nm. (The source is adapted from [84].*)*

2.2.4 Transmission Electron Microscopy (TEM)

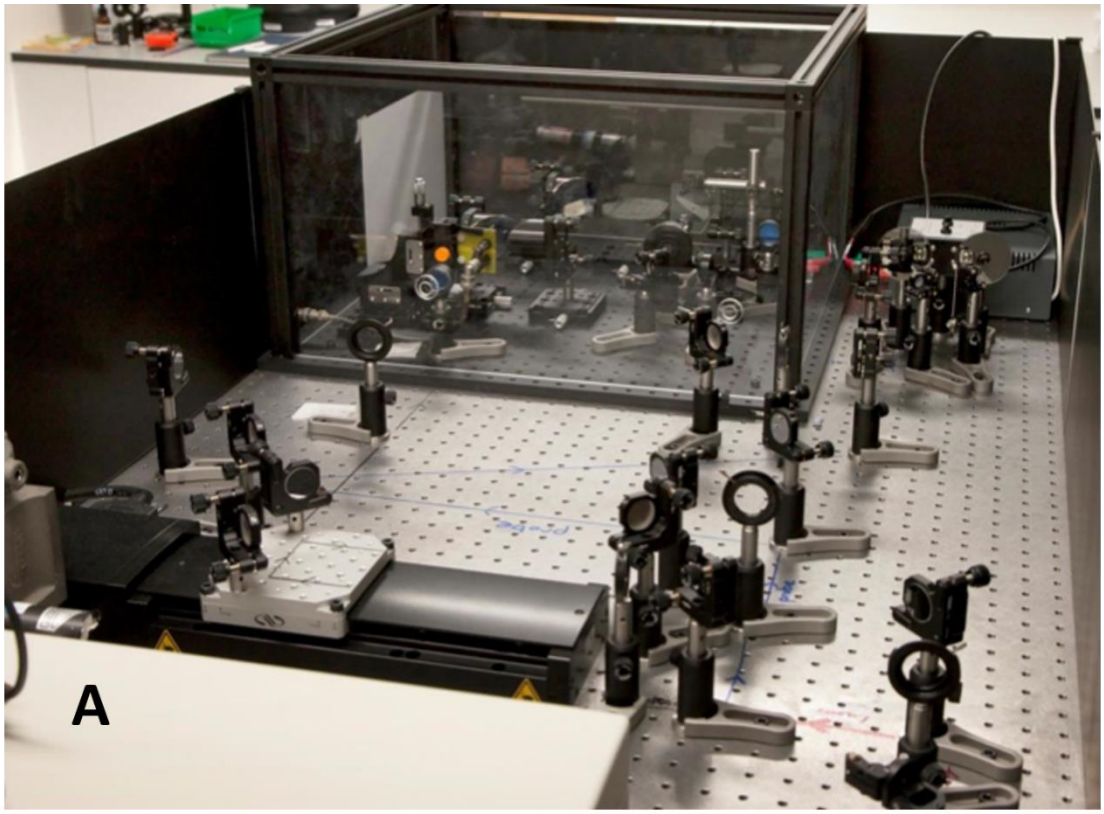
Glow-discharged carbon-coated 300-mesh copper grids were prepared using the droplet method, where 10 µl aliquots of samples from the fibril growth assay were adsorbed for 1 min followed by 10 µl of 1% (w/v) of phosphotungstic acid (PTA) adjusted to the same pH with the proteins, to negatively stain the sample. Images were recorded on a JEOL JEM-1230 electron microscope operated at 80 kV.

2.3 Terahertz approaches

2.3.1 THz-Time Domain Spectroscopy

Compared to infrared spectroscopy, where these instruments are much more mature and commercially used in biological laboratories, THz spectroscopy is now in the progress of transitioning from its initial development by physicists and engineers to being a useful tool for biologists. Among a number of THz radiation sources and systems that are now available, THz-Time Domain Spectroscopy is more common to be used in bio-systems research. A schematic illustration of the setup of a THz-TDS is shown in Figure 2.4B [85].

Our TDS system, which is shown in Figure 2.4A, has its THz pulses generated and detected using 100 fs light pulses from a mode-locked Ti: sapphire laser at a wavelength of 800 nm and a repetition rate of 80 MHz [51]. The beam splitter separates the laser into two beams: pump beam and probe beam; the pump beam propagates through the delay stage and lead the THz emitter to generate the THz pulse [86], the probe beam is used for probing [85]. The THz emission was collimated and focused by a pair of off-axis parabolic mirrors and the beam collected by another pair of off-axis parabolic mirrors. The THz pulse is single cycle oscillation of the electric field with duration of few ps which can be detected by ZnTe (Zinc Telluride) crystal. Gated coherent detection is used in the measurements, whereby an optical delay line adjusts the time delay between the THz-wave generation and detection arms. The empty system are mainly treated as reference and was propagated as a reference signal, and the THz wave propagation through the sample is recorded as sample signal. The signals are then recorded by the lock-in amplifier for further data processing.



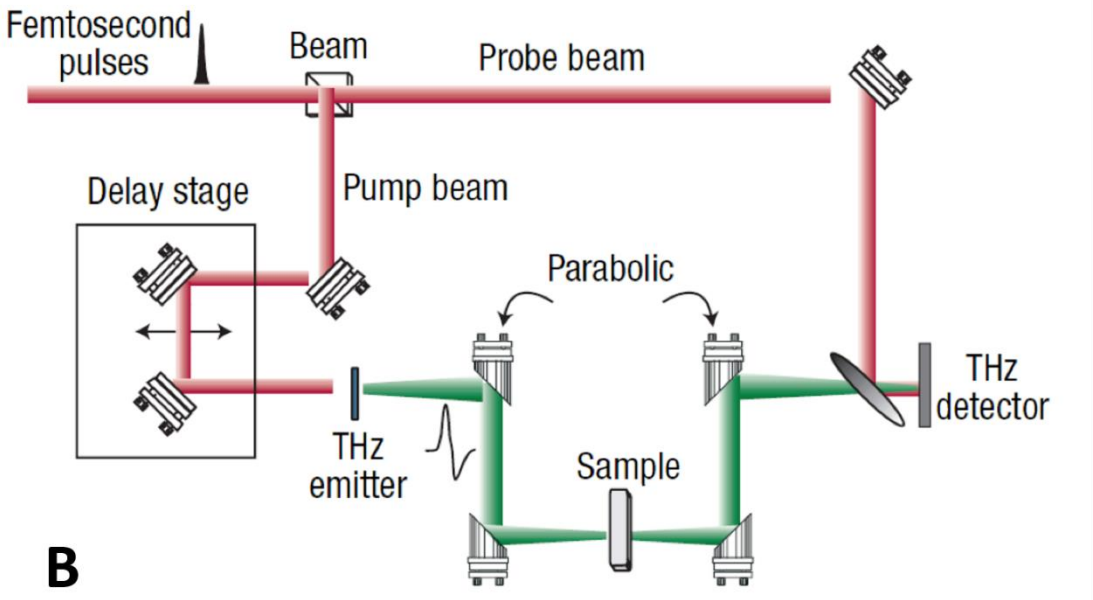


Figure 2.4 The THz-TDS system used for measurements. (A) A picture was taken at QMUL THz lab; (B) A schematic diagram of a common THz time-domain system. The system employs a femtosecond pump laser and is based on THz pulse emission and detection by a photoconductive emitter and a THz detector antenna, adopted form.

2.3.2 Data processing

Since the THz technique measures the field strength, it is possible to invert the complex Fresnel transmission coefficients and thus to obtain the index and absorption of the highly absorbing protein solution samples. In order to extract the spectral information of the sample, a reference measurement with the sample removed will be first done and Fourier transform will be applied to converting the time-domain spectra to frequency-domain spectra. The recorded data consists of a single trace in the time-domain containing the reference pulse and the sample pulse as indicated in Figure 2.5 (a) and the converted frequency-domain spectra was shown in (b) [87].

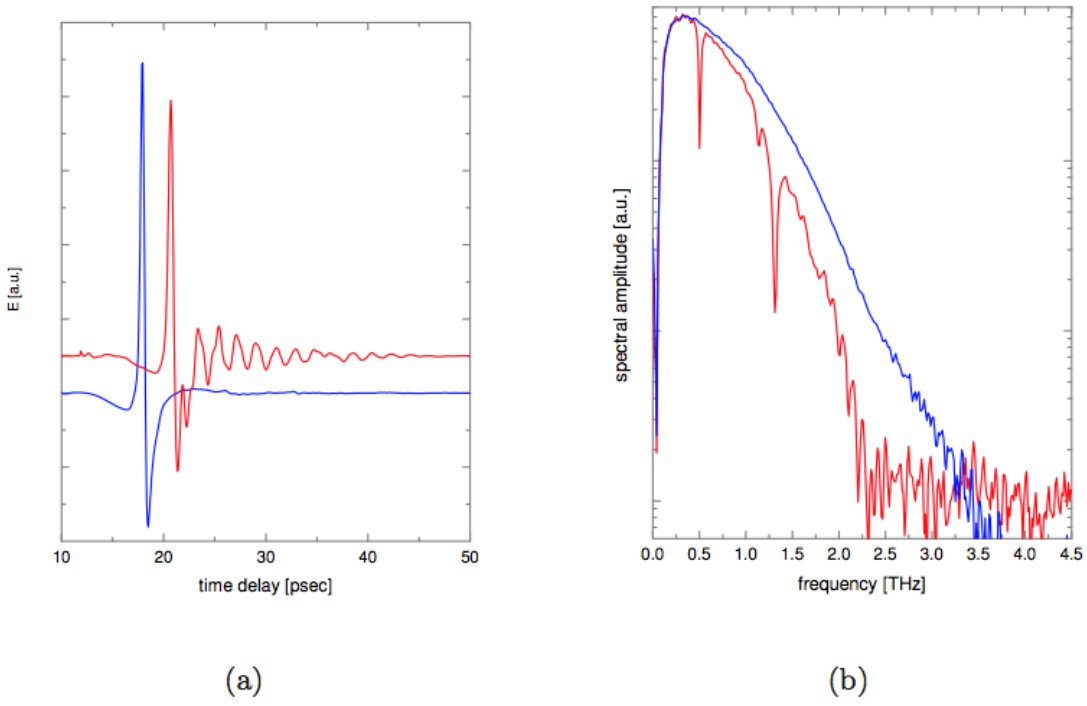


Figure 2.5 An example of the determination of the dielectric function from a reference and sample THz pulse [9]. (a) and (b) represent the time domain and frequency-domain spectra, respectively. Blue line represents the reference signal while the red line is the sample signal.

In the delay stage, a total of $N = 1024$ data points was recorded using 10 μm delay stage steps which resulted in a 10240 μm trace. This single trace is separated into the

two signals $E_{ref}(t)$ and $E_{sam}(t)$ which consist of $N^{'} = N/2$ data points and which then are individually Fourier transformed to the frequency-domain.

The sampling time axis and frequency axis are given by:

$$t_n = 2 * \frac{scan_N}{c},\ f_n = n * \frac{c}{2*scan_N} \tag{2.2}$$

Where $scan_N = n * step\ , n = 1, 2, 3, \ldots, 1024.$

Fourier transforming $E_{ref}\ (t)$ and $E_{sam}(t)$ results in

$$E_{ref}\ (\omega)\ =\ A_{ref}\ (\omega)\ exp[i\varphi_{ref}\ (\omega)] \tag{2.3}$$

$$E_{sam}(\omega)\ =\ A_{sam}(\omega)\ exp[i\varphi_{sam}(\omega)] \tag{2.4}$$

The Fourier transformations are performed as an FFT algorithm in data analysis software MATHCAD giving out values for the amplitudes $A_{ref}(\omega)$ and $A_{sam}(\omega)$ as well as for the phases $\varphi_{ref}(\omega)$ and $\varphi_{sam}(\omega)$ for each frequency value ω.

The two material optical parameters refractive index *n* and extinction coefficient *k*, are formed in the complex refractive index of materials $\tilde{n} = n - ik$ [88]. In this formula, the refractive index is a number describe how the THz radiation propagates through the samples. Extinction coefficient means the THz wave extinction value by the solvation, which related to the wavelength (or frequency). As the THz beam goes through the sample as well as two TPX plates, a transfer function (TF) for three layers have developed to only calculate the solution parameters:

$$TF = T_m \exp\left(-\frac{i2\pi f_m d_w \tilde{n}_w}{c}\right) \exp\left(-\frac{i2\pi f_m d_s \tilde{n}_s}{c}\right) \tag{2.5}$$

in this formula, $\tilde{n}_w$ is the complex refractive index of window material, $\tilde{n}_s$ is the refractive index of sample $\tilde{n} = n - ik$. T_m is the combined Fresnel transmission coefficient, which is associated with the front and back boundary interfaces between sample and host medium to the sample.

As sample is sandwiched by two TPX plates (Figure 2.6), the combined transmission coefficient can be transformed:

$$T_m = \frac{T_1' T_2' T_3' T_4'}{T_1 T_2 T_3 T_4} = \frac{T_2' T_3'}{T_2 T_3} \tag{2.6}$$

The T values take the form as:

$$T_1 = \frac{2\tilde{n}_{w1}}{n_{air} + n_{w1}}, T_2 = \frac{2\tilde{n}_{air}}{n_{air} + n_{w1}}, T_3 = \frac{2\tilde{n}_{w2}}{n_{air} + n_{w2}}, T_4 = \frac{2\tilde{n}_{air}}{n_{air} + n_{w2}} \tag{2.7}$$

$$T_1' = \frac{2\tilde{n}_{w1}}{n_{air} + n_{w1}}, T_2' = \frac{2\tilde{n}_{S}}{n_{S} + n_{w1}}, T_3' = \frac{2\tilde{n}_{w2}}{n_{S} + n_{w2}}, T_4' = \frac{2\tilde{n}_{air}}{n_{air} + n_{w2}} \tag{2.8}$$

Thus the TF function can be further performed as:

$$TF = \frac{T_2' T_3'}{T_2 T_3} \exp\left(-\frac{i2\pi f_m d_w \tilde{n}_w}{c}\right) \exp\left(-\frac{i2\pi f_m d_s \tilde{n}_s}{c}\right)$$

$$= \frac{(\tilde{n}_w + 1)^2 \tilde{n}_s}{(\tilde{n}_w + \tilde{n}_s)^2} \exp\left(-\frac{i2\pi f_m d_w \tilde{n}_w}{c}\right) exp\left(-\frac{i2\pi f_m d_s \tilde{n}_s}{c}\right) \tag{2.9}$$

By using formula$\tilde{n} = n - ik$ and equation (2.7, 2.8) the values of refractive index n and extinction coefficient k can be calculated.

Optimization of apodization functions for the time-domain signals obtained in terahertz (THz) transient spectrometry is also necessary, by enhancing the resolution

and frequency-depedent dynamic range of the THz spectra, as supported by measurements [143].

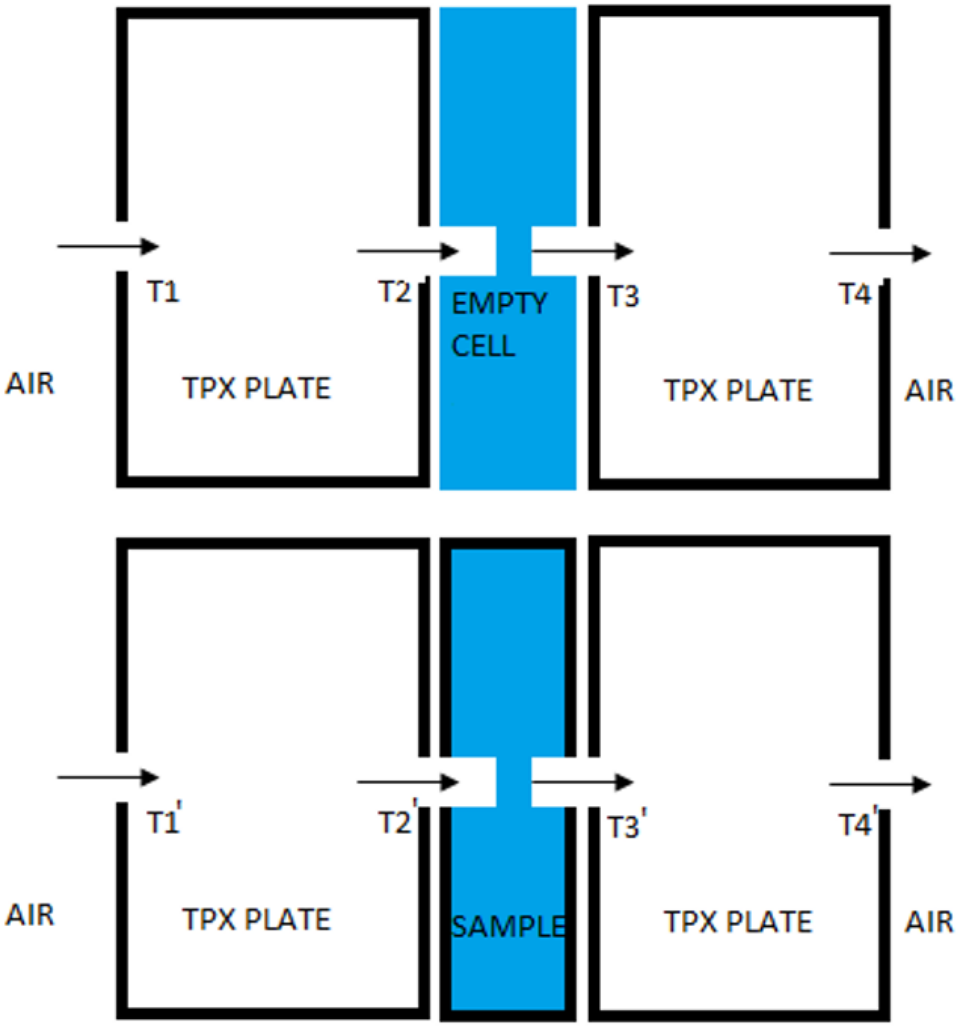


Figure 2.6 Schematic diagram of the sample in the THz-TDS system. THz beam propagates through empty cell (upper panel) or sample (lower panel) sandwiched by two TPX plates. The THz beam follows the direction T1 (air to the TPX plate), T2 (TPX plate to sample), T3 (sample to TPX plate), and T4 (TPX plate to the air). This beam's direction shows four Fresnel transmission coefficients exists [89].

2.3.3 Vector Network Analyzer coupled to quasi-optical bench

While current THz-TDS system can cover the range from 0.2 to 3 THz with a dynamic range of around 80 dB, a vector network analyser-driven quasi-optical bench is

developed to be another useful tool that capable to monitor protein materials from 75 to 325 GHz. The quasi-optical transmissometer used for our work is illustrated in Figure 2.7. It consists of a pair of corrugated horns (H1, H2) for transforming rectangular-wave-guide modes into linearly polarised free space Gaussian beam modes [52]. A pair of parabolic mirrors (F1) to couple the signal beam from H to linearly polarized quasi-collimated beam which passes through the sample (S), and two further parabolic mirrors (F2) couple the THz radiation to focus on the sample. The above setup is driven by a HP 8510C vector network analyser (VNA) in individual waveguide bands (75-110, 110-170, 220-325 GHz are currently available in QMUL lab). In measuring a transmittance, the amplitude and phase of the transmitted signal will be measured with and without the sample (shown in Figure 2.8) in the beam. The ratio of these complex amplitudes represents the insertion loss of the sample.

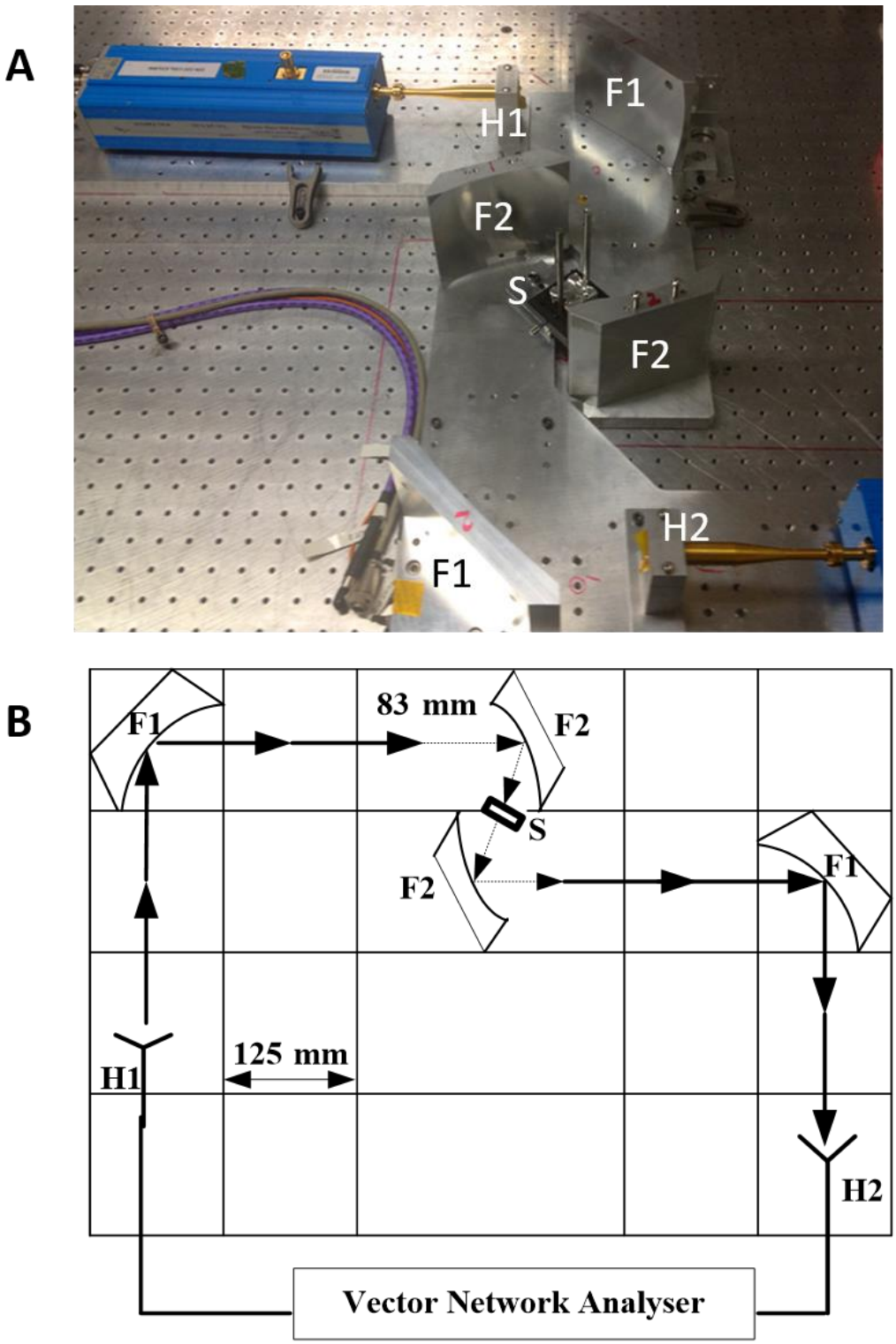


Figure 2.7 The VNA-QO measurement system. (A) The photo of QO measurement circuit at QMUL THz lab. (B) Schematic diagram of the QO-VNA measurement system. H1 and H2 denote the corrugated horns connecting with waveguide bands; F2 are two parabolic mirrors used to couple the THz radiation on the sample while F1 represent another pair of parabolic mirrors between H and F2.

The 220-325 GHz band was used in the measurements of this work, and the complex transmission coefficients S_{21} of the samples were collected using this system. The S_{21} transmission amplitude is used as a measure of absorbance of the samples [175].

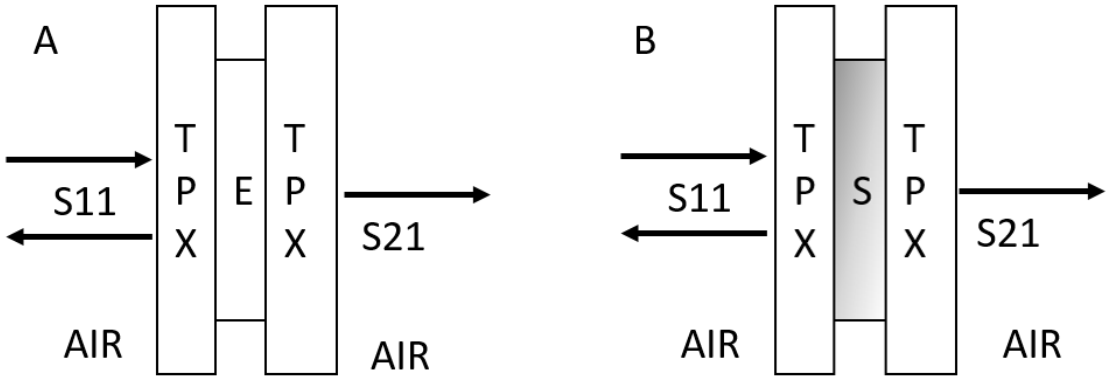


Figure 2.8 Schematic diagram of the sample in VNA system. (A) The empty cell sandwiched by two TPX plates; (B) Sample with two TPX plates. S11 refers to the reflection response and S21 for the transmission response.

The structure of a 3-layer system (TPX plate – sample –TPX plate) can be represented in a corresponding 3-term product ABCD matrix. The transmittance matrix of the assembly is given by:

$$A(\varepsilon, d) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_{TPX} & B_{TPX} \\ C_{TPX} & D_{TPX} \end{bmatrix} \begin{bmatrix} A_S & B_S \\ C_S & D_S \end{bmatrix} \begin{bmatrix} A_{TPX} & B_{TPX} \\ C_{TPX} & D_{TPX} \end{bmatrix} \quad (2.10)$$

d is the material thicknesses; ε, is the materials' dielectric properties; the subscript 'TPX' stands for the front and back TPX plates of the sample cell and 's' for the sample.

By applying wave propagation equations of A, B, C, and D for each layer according to the three layers method from [90], the parameters of reflection (S_{11}) and transmittance (S_{21}) can be obtained:

$$R = \left[\frac{A + B - C - D}{A + B + C + D}\right] \tag{2.11}$$

$$T = \left[\frac{2}{A + B + C + D}\right] \tag{2.12}$$

2.4 Summary

This chapter presents the general overview of both biological approaches and THz measurement techniques used in the project. Different protocols on protein production were applied to individual proteins, the steps of protein purification are discussed in the respective relevant chapters. The biological approaches are started with protein purification; purified proteins were subsequently used for various experiments, including circular dichroism, transmission electron microscopy and THz measurements. In THz techniques section, both Terahertz time-domain spectroscopy and the Vector Network Analyser Deployed with Quasi-Optical measurement system are described.

Chapter 3 **Sub-Terahertz studies of protein in Ionic Environments**

3.1 Background

THz and sub-THz spectroscopy have been applied to the study of solvation dynamics of various bio-molecules ranging from amino acids to complex DNA and lipid layers [5]. THz radiation senses the collective dynamics of molecules and hydrogen-bond networks that are directly related to the properties of a biomolecule, such as conformation state, hydration shell radius and surface hydrophobicity. [66].

Naturally-solvated proteins exist in aqueous environments rich in salt ions, sugars and other biomolecules. In many THz studies, dilute salt buffers are used to stabilize the pH of a solution and protein structure [91], [92]. For instance, Bye *et al.*, have shown using differential scanning calorimetry (DSC), that a NaCl solution actively stabilizes the lysozyme structure above 550 mM [93]. Other studies have already addressed the THz response of varies solvated ions [94]. However, only limited THz studies are available under the influence and stabilizing the effect of various salt buffers on protein stability in solution. THz spectroscopy of bio-solutions is still in its early stages of development and acceptance compared to the established techniques of fluorescence spectroscopy, circular dichroism, calorimetry. However, growing availability of THz sources supported by high-performance molecular dynamics simulations is building a foundation for THz methods and analysis.

In this chapter, a well-designed vector network analyser (VNA) -driven quasi-optical bench is developed to monitor the protein samples from 220 to 325 GHz range. The sub-THz measurement is first verified by monitoring the unfolding process of BSA protein under the influence of strong chaotrope GdmCl, with the aim to investigate the potential of sub-THz spectroscopy to address the salt-protein interaction. By comparing with FTIR and THz-TDS, VNA is able to supply a greater sensitivity in dynamic range with low signal-to-noise ratio.

3.2 Materials and methods

3.2.1 Materials

Three proteins and two salts were chosen for this study: hen egg white lysozyme (PDB: 2LYZ, Mw: 14.3 kDa), myoglobin from equine skeletal muscle (PDB: 1MBN, Mw: 16.7 kDa) and bovine serum albumin (BSA, PDB: 4F5S, Mw: 66.5 kDa); Sodium chloride (NaCl) and guanidinium chloride (GdmCl). All proteins were purchased from *Sigma* Aldrich in the form of lyophilized powder. The purity for lysozyme and BSA is ≥ 98%, for myoglobin 95-100%, essentially salt-free. Proteins are purified with gel chromatography (HiLoad 26/60 *Superdex75*, GE Healthcare) using FPLC (Fast Protein Liquid Chromatography) with 10mM potassium phosphate buffer at pH 6.5. The purity was checked by SDS-PAGE before use. All the proteins were prepared in a 20 mM potassium phosphate buffer (pH 7).

3.2.2 Far-UV circular dichroism (CD)

Circular dichroism (CD) was used to collect the secondary structure changes of proteins, in a quartz cuvette with a 1 mm path length and subsequently placed onto the Chirascan CD spectrophotometer (Applied Photophysics). Absorbance scans were

carried out from 190 to 260 nm at 0.5 nm intervals at a set temperature of 20 ˚C. For each sample scan, three runs were taken and then averaged. An absorbance scan of the buffer was also taken and used as a baseline to deduct the absorbance scan of samples.

3.2.3 Vector network analyser at 220 to 325 GHz

A quasi-optical transmissometer [95] driven by an HP N5244A Vector network analyser (VNA) was used to collect the experimental results. The Agilent VNA (with spanning DC up to 1 THz) is capable of offering a greater dynamic range and signal-to-noise ratio of up to 100 dB. Frequency-multiplier heads coupled to corrugated horns up-convert the frequency to 0.22-0.325 THz. This band of frequency corresponds to a wave period of the order of 1 picosecond. Such a time-scale is attributed to fast relaxation processes of hydration dynamics. The VNA signal power is of the order of 1 mW and is continuous wave radiation. Sub-THz absorption is calculated as the inverse of the transmission coefficient S_{21} integrated over the whole waveguide band.

For measurements, a Bruker liquid cell (A145) with TPX windows was utilized as a holder for the solutions. The thickness of the solutions was set to 100 μm by a polytetrafluoroethylene (PTFE) spacer. Since water is a strong absorber of sub-THz radiation, such thin samples allow measurable radiation through in the desired operating band while still being thick enough to provide sufficient beam-material interaction. For refilling, the cell is removed from the beamline, flushed with distilled water and filled with the next solution. This procedure is done to avoid any thickness deviation which is known to be a major source of uncertainty in absorption measurements [96]. The solutions are initially kept in an ice box, then allowed 30 minutes to equilibration with room temperature before each measurement. Combined

measurement error comprises kinematic uncertainty and instrumental noise and together constitutes ±0.2 % and is invariant for each reading.

3.3 Results and Discussion

3.3.1 Guanidinium Chloride in protein denaturation process

Guanidinium ($C(NH_2)_3^+$, Gdm^+) salts have been widely employed for the denaturation or unfolding of protein molecules in aqueous solutions; its process can be reversible in some cases with careful control of conditions [97]. The denaturing effect to protein on GdmCl has been experimentally studied in Neutron scattering [98] and dielectric spectroscopy [99], concluding that due to the low charge density of Gdm^+, there is no sign of strongly bound solvation shell around the Gdm^+ ions. MD simulation results have predicted a significant ordering of hydration layer around Gdm^+ which means the unfolding changes of protein are also led by the water structure alteration [100]. One of the recent THz studies, Nirnay Samanta *et al.*, [101] has employed THz time domain spectroscopy (TDS) to compare the collective hydrogen bond dynamics in GdmCl and NaCl salts in 0.3-2.0 THz range, and the denaturing effect of GdmCl on HSA (Human serum albumin) protein.

Far-UV Circular Dichroism spectra

Far-UV CD spectra (wavelength 190-250 nm) of BSA and Apomyoglobin protein with GdmCl was collected and used to investigate the structure information of the proteins. As chloride ions absorb strongly below 195 nm, thus the CD signal of proteins in the presence of the GdmCl was plotted at 222 nm (as depicted in Figure 3.1), which is particularly sensitive towards the secondary structure of proteins. It appears from Figure 3.1 that a drastic change is observed in the case of GdmCl for both BSA and

apomyoglobin proteins especially beyond a concentration of 2 M indicating the onset of unfolding of proteins. GdmCl decreases the content of alpha-helix dramatically with a concomitant increase in the random coil conformation.

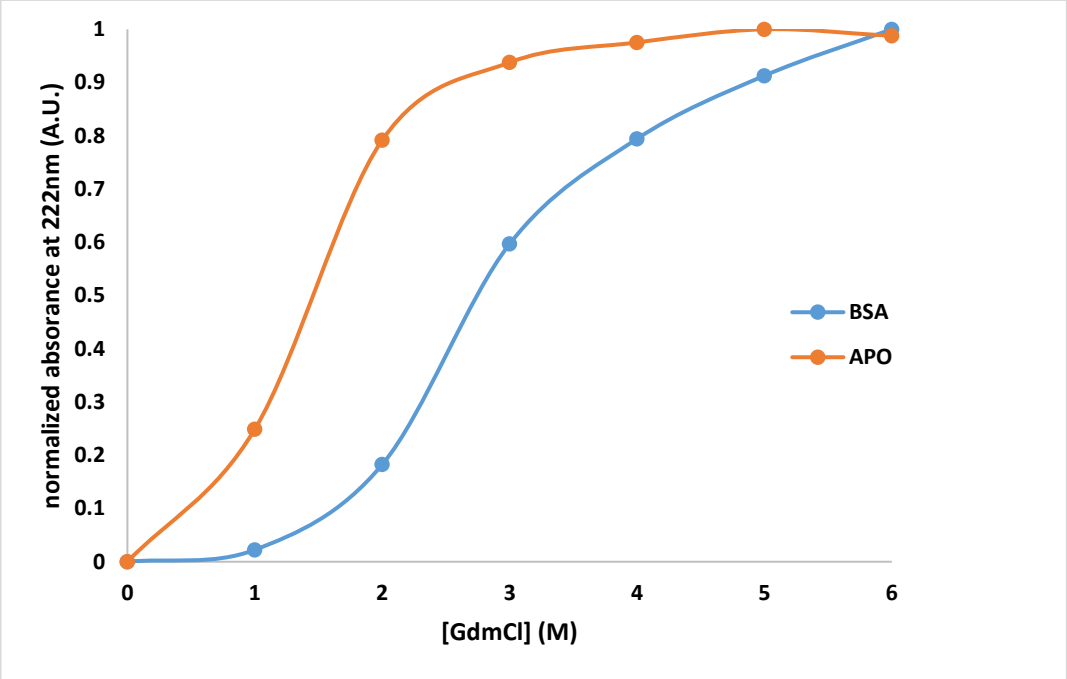


Figure 3.1 Denaturation CD curves of 1 mg/ml BSA and Apomyoglobin protein with GdmCl concentration up to 6 M.

VNA sub-THz measurement

Figure 3.2 shows the relative changes in THz absorption of BSA, lysozyme and myoglobin protein solution at six concentrations of guanidinium chloride (GdmCl). Relative absorption of pure GdmCl is also provided as a reference. GdmCl is a strong chaotrope and is commonly used as a protein denaturant [102]. The dominating mechanism of unfolding is destabilization of protein structure via interaction of charged salt particles with charged and polar sites on the protein surface. Protein solution exhibits a clear decrease in THz transmission, especially Above 1.5-2 M of GdmCl, consistent with THz-TDS study of HSA/GdmCl solution [101]. This concentration marks a transition point where rapid protein unfolding begins. As BSA

molecules unfold and lose their tertiary structure, more hydrophobic residues from the protein core get exposed to the solvent. This process is accompanied by an increase in Gibbs free energy. It is known that hydrophilic and polar molecules possess higher THz absorptivity compared to hydrophobic [103]. Therefore, a protein solution attenuates THz radiation less than GdmCl salt solution as protein molecules undergo denaturing and become more hydrophobic. The critical concentration of GdmCl determined here by THz spectroscopy is consistent with other techniques such as fluorescence spectroscopy [104] and circular dichroism [102]. For instance, enzymatic activity determined by colorimetric methods reveals that glucose oxidase (a dimeric enzyme) starts unfolding at 2 M concentration of GdmCl and 4 M it is completely unfolded [102]. In the same study, other experimental methods as fluorescence spectroscopy, circular dichroism, and size-exclusion chromatography confirm these unfolding trends. This is in agreement with data presented here. The inflexion point of the curve for GdmCl solution can also elucidate specific information regarding the size of the dynamic hydration shell [51].

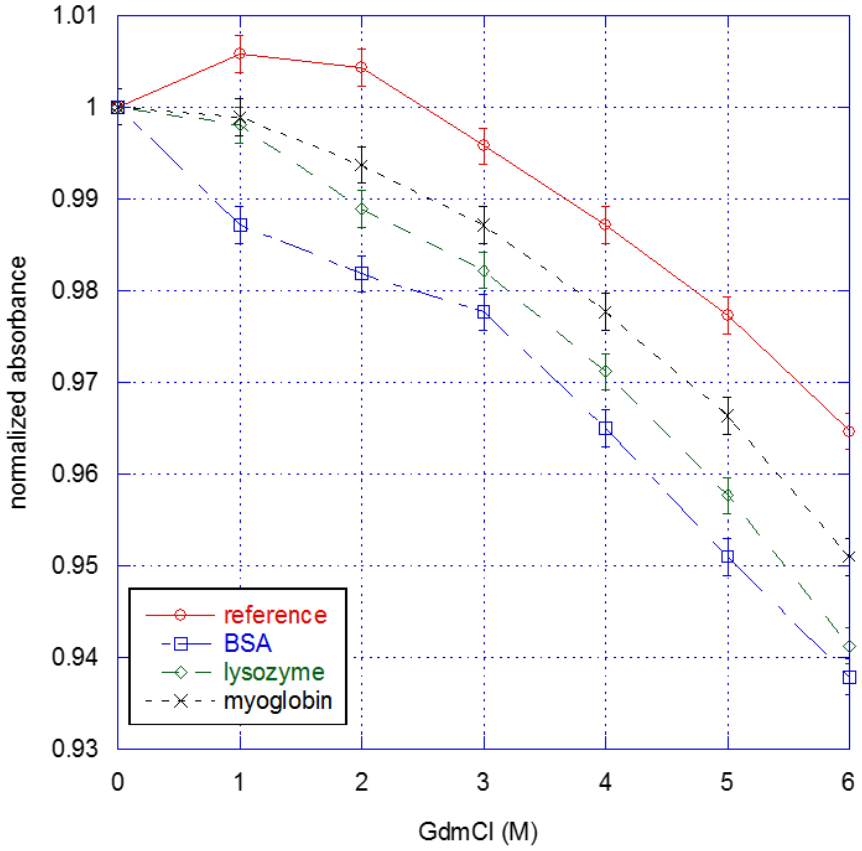


Figure 3.2 The normalized absorbance of 1 mg/ml BSA, lysozyme and myoglobin in GdmCl concentration range (0-6 M), measured in 220-325 GHz range.

3.3.2 Protein in NaCl ion environment

In order to understand the mechanism behind protein stabilisation and solubility, it is important to understand the role salts play. In the early theories, the best known experimentally studies that discussed the ability of ions to influence the protein solubility is by Hofmeister in 1888 [105], [106]. As a general trend, anions and cations can be sorted by their ability to solubilise and destabilise (structure breaker) and precipitate and stabilise proteins (structure maker). It was originally believed that the trend ions play was due to the ability of ions making or breaking water structure around proteins. However, current studies suggest a longer range interaction exist between ions and hydration layer of protein using dielectric relaxation spectroscopy [107],

neutron scattering [108] and molecular dynamic simulation [109]. This leaves the question of what drive salts to act in protein dynamics unanswered and suggests that salt-protein interactions are more complex than previous thought.

Motivated by sub-THz measurement is sensitive to detect absorption change of denaturation in GdmCl solutions, BSA protein was selected to study the THz response in a neutral salt sodium chloride at up to 1 M salt concentration range. The purpose is to investigate the anion stabilization of BSA protein within sub-THz range.

Thermal melts as a measure of protein stability

For most proteins, secondary structure is lost upon unfolding and the far-UV CD spectra of a folded and unfolded protein are thus distinct. The structural stability of the proteins is presented as the temperature where the protein is unfolding at its maximum rate (the T_m value) [110]. The melting temperature is defined as the temperature when the native and denatured states are equally populated at equilibrium. Therefore, it is commonplace to use the variation of melting temperature (T_m) as a faithful measure of stability [110].

BSA protein solution was prepared at 1 mg/ml in different sodium chloride conditions: 50 mM, 100 mM, 150 mM, 200 mM, 500 mM. Protein in 20 mM potassium phosphate buffer was used as a reference. CD spectra were recorded every 2.5 °C from 25 to 90 °C. The molar ellipticity at 222 nm, θ_{222}, as a function of temperature is shown in Figure 3.3. It's not surprising that the melting curves trends are quite similar among different NaCl salt concentration, as Na^+ and Cl^- ions are not strongly interacted with protein or hydration water and will not cause many alterations in protein stability.

However, the melting temperature is getting slightly higher with increasing NaCl salt concentrations up to 500 mM. As greater protein stability will be indicated by higher melting temperatures (T_m values), the protein becomes more stable with increasing NaCl salt concentration from 50 to 500 mM. The reason could be that in low protein concentrations protein structure requires achieving its maximum stabilization with the help of surrounding bindings. The binding of sodium chloride which is stronger than hydrated will help to stabilize the conformation of protein, and the presence of sodium chloride bound to the surface of protein would also alter the hydration layer around the protein and the strength associated with the structure.

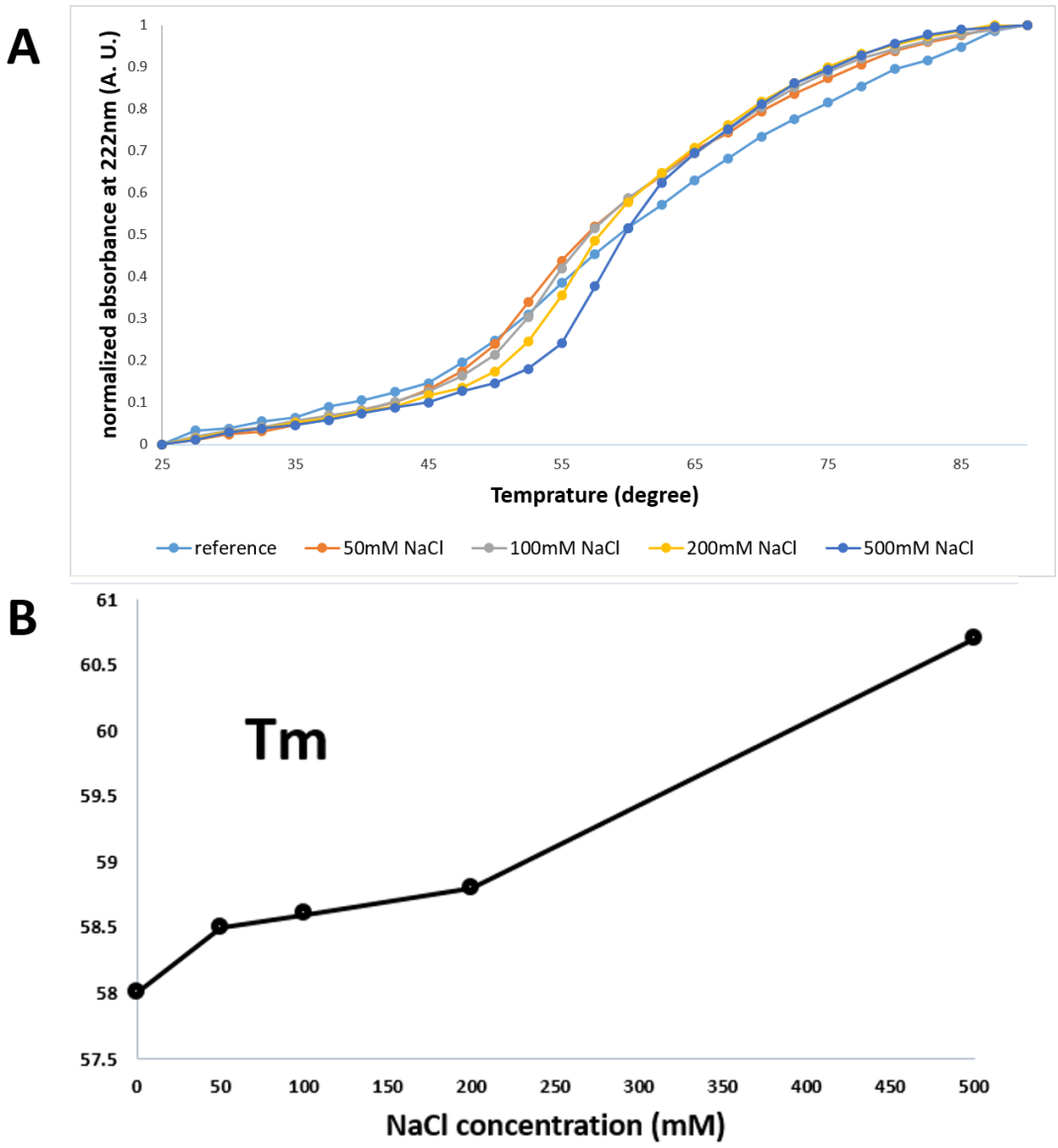


Figure 3.3 (A) Normalized absorbance of BSA protein in different concentrated NaCl solutions (0-500 mM) with 20 mM sodium phosphate buffer performed by CD thermal melts method; (B) The melting temperatures (T_m values) for different NaCl concentrations. The overall pH of each solution was maintained at pH 7. The temperature range was recorded from 25 to 90 °C.

Protein concentration change in 220-325 GHz sub-THz domain

Figure 3.4 shows the relative (with respect to buffer) absorption of BSA proteins in different NaCl salt conditions (0 to 500 mM), compared at different protein concentrations (0 to 100 mg/ml). The spectra start to show a linear decreasing change

after 5 to 10 mg/ml BSA concentration, and a nonlinear trend in absorption is immediately clear at low concentrations.

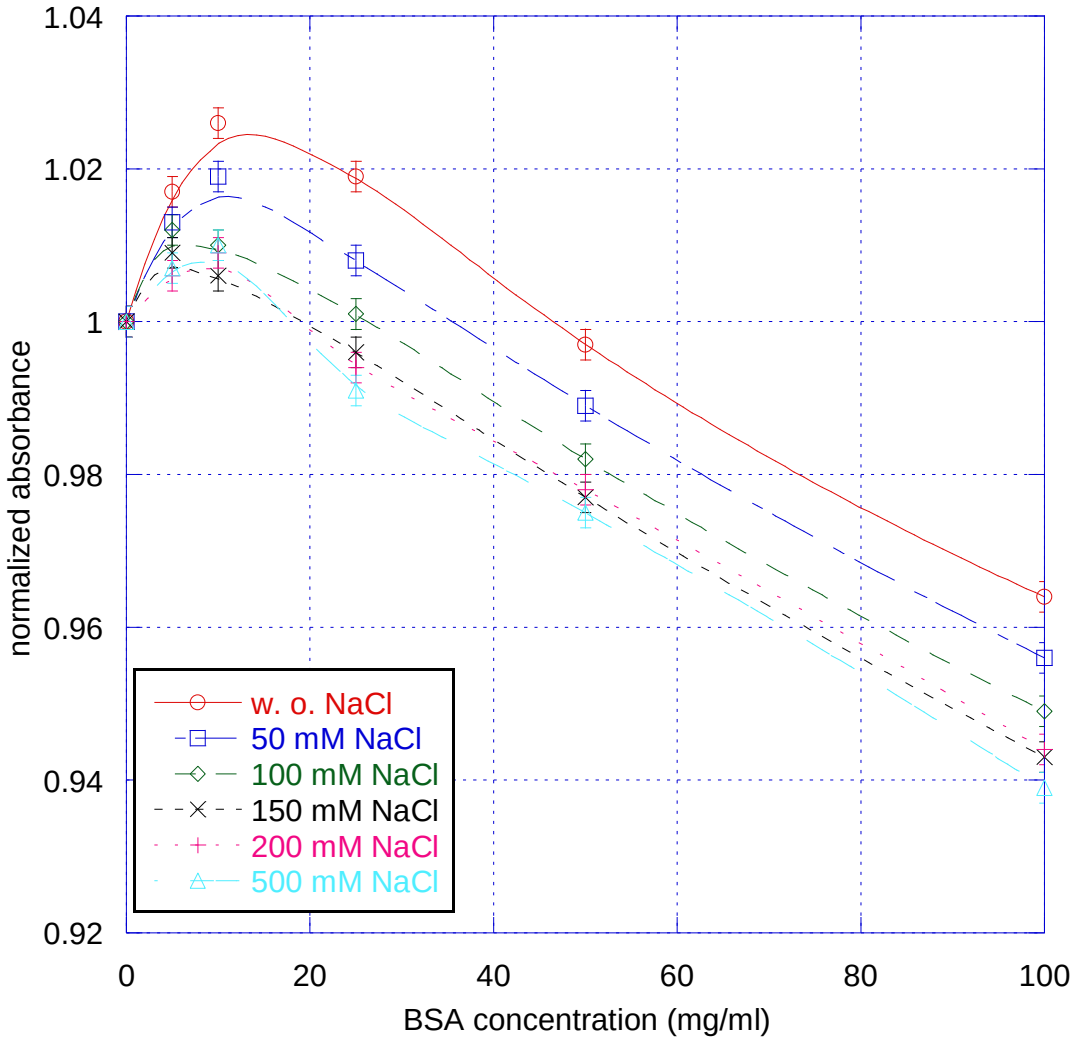


Figure 3.4 Normalized absorbance of different concentration BSA protein at 0, 50, 100, 150, 200 and 500 mM of NaCl. The protein in different concentration of NaCl solutions are compared as a function of BSA concentration. Errors were estimated as a combination of instrument noise and the uncertainty of positioning of the liquid cell in the beam-line. Error bars are invariant at each point in this plot and are equal to ±0.001.

The THz defect (i.e. the smaller electro-magnetic absorption of the solution compared to bulk water) [77], was the expected scenario since protein molecules in the dry-state are known to be less absorbing than water. However, all three protein solutions exhibit

higher absorption as compared to a two-component. Protein in buffer (without NaCl) is seen to have the largest additional absorption with respect to the rest spectra, followed by 50 mM NaCl and 100 mM NaCl. The proteins studied exhibit an initial rise in absorption (THz excess) at low concentrations. The maximum absorbance of solutions occurs at 5 mg/ml and 10 mg/ml BSA concentration, which agrees with the previous study on THz excess in protein and fit the hydration shell size.

THz measurement of BSA protein with NaCl concentration change

Figure 3.5 shows the collected experimental results for NaCl buffer and different BSA protein solutions at 5 to 100 mg/ml concentrations. It illustrates how the absorption of the solution changes with increasing concentration of NaCl. As expected, for NaCl buffer, the absorption rises monotonically until 200 mM, since charged Na^+ and Cl^- ions are more responsive to THz radiation than water molecules, and then saturates. For protein solutions, however, the trend is more complex. Besides 5 mg/ml BSA solutions, most of the spectra exhibit a well-pronounced minimum in THz response at a salt concentration of 100 mM. This feature is interpreted as stabilization of the protein structure by Na^+ and Cl^- ions at this particular concentration. Initially, when the NaCl concentration is increased, ions start to occupy protein binding sites, thereby inhibiting its flexibility. This reduction in action is expressed as a retarded response to THz probe-radiation. However, it is also noted that the normalized absorbance in Figure 3.5 spans within 3% overall, the small variations in absorbance spectra for BSA concentration 25, 50 and 100 mg/ml, could not be clearly attributed to determine the difference between samples.

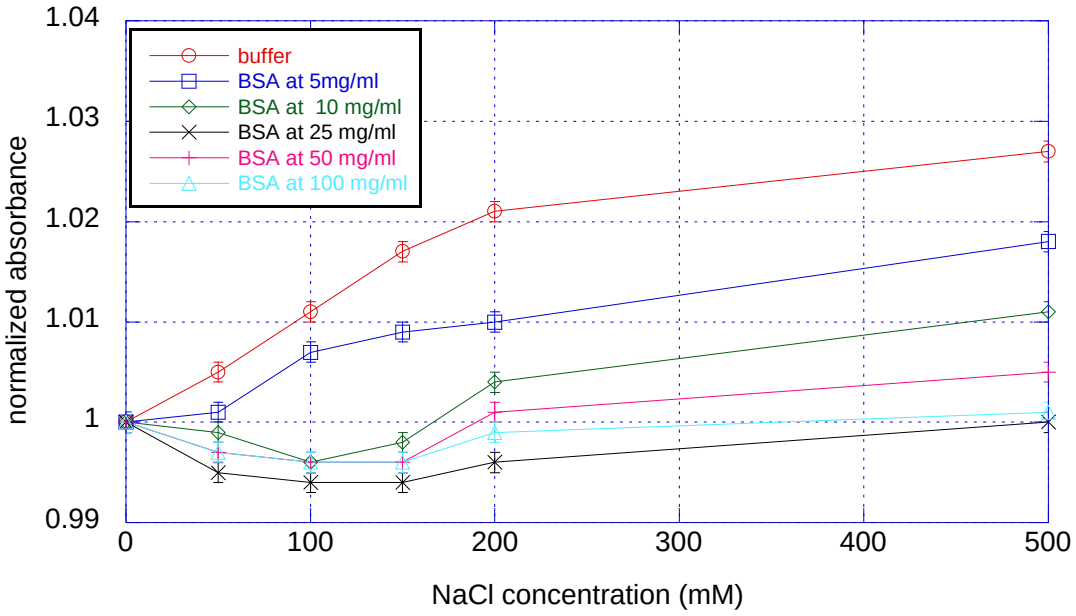


Figure 3.5 Normalized absorbance of BSA proteins, different concentration of BSA solution are compared as a function of NaCl concentration at 0, 50, 100, 150, 200 and 500 mM of NaCl. Error bars are invariant at each point in this plot and are equal to ±0.001.

The evolution of relative absorption is similar from 200 to 500 mM for all three solutions considered. This implies that above 200 mM the additional ions remain mainly in the buffer and do not actively interact with protein molecules as most of the binding sites on the protein surface are already occupied.

Interestingly, the investigation of the thermal stability of BSA by CD is not able to reproduce this feature at 100 mM of NaCl. CD melting curve shows a continuous increase in melting temperature (T_m) of BSA up to 500 mM of NaCl, indicating that thermal stability also rises.

The CD experiments agree with the findings obtained by the same technique for the glucose oxidase enzyme [14]. The T_m of this enzyme constantly grows up 2 M of NaCl

while starting to plateau after 0.5 M. Such differences between THz and CD results might be triggered by distinct physical quantities sensed, as CD measures changes in heat capacity, while THz radiation probes collective long-range dipole reorientation dynamics.

3.4 Summary

This study provides additional evidence that sub-THz spectroscopy is sensitive to molecular conformation changes, illustrated by the unfolding of BSA protein in GdmCl solution. The results also show that sub-THz spectroscopy can be meaningfully applied to identify the buffer concentration that ensures the most stable protein conformation. This is crucial for the study of protein stability, enabling a prolonged storage of protein solutions; and provides insights into protein-salt interaction.

Chapter 4 **TDS-Terahertz Investigation of Myoglobin in Native and Molten Globule-like States**

4.1 Introduction

During the last 20 years, increasing attention has been focussed on the dynamic aspects of protein structure and function. The dynamic properties of proteins are clearly essential for protein folding. Commonly, proteins fold from irregular unfolded conformations to biologically active folded structures in a hierarchical manner. Secondary structure is stabilised by hydrogen bonds between the amide and carbonyl groups of amino acid residues that are close in sequence [111], [112]. Then the tertiary structure is stabilised by both hydrogen bonds and hydrophobic interactions among main-chain and side-chains of more distant segments of the chain [113], [114]. For water-soluble proteins, hydrophobic interactions among non-polar side chains favour the collapse of hydrogen-bonded secondary structure elements into a compact conformation. Frequently, the mechanism of folding is a single two-step process with the stable native and the unfolded states; however, along the folding pathway of some proteins, the intermediate states in the protein-folding process will also be formed. These states have been the focus of close experimental scrutiny because their structure and behaviour may reveal many intimate details of the protein-folding process.

The protein used in this work, Apomyoglobin (ApoMb), is produced from myoglobin by removal of the heme group. Myoglobin (Mb) is a water-soluble globular protein of approximately 17 kDa molecular mass and 153 residues and consists of eight distinct helices named from A-H, without other secondary structure or disulphide bonds. After

the removal of the heme group, ApoMb still retains similar properties to myoglobin, such as the solubility in water, helical content and some of its structural characteristic. However, ApoMb possesses more flexibility and is less compact than myoglobin [115].

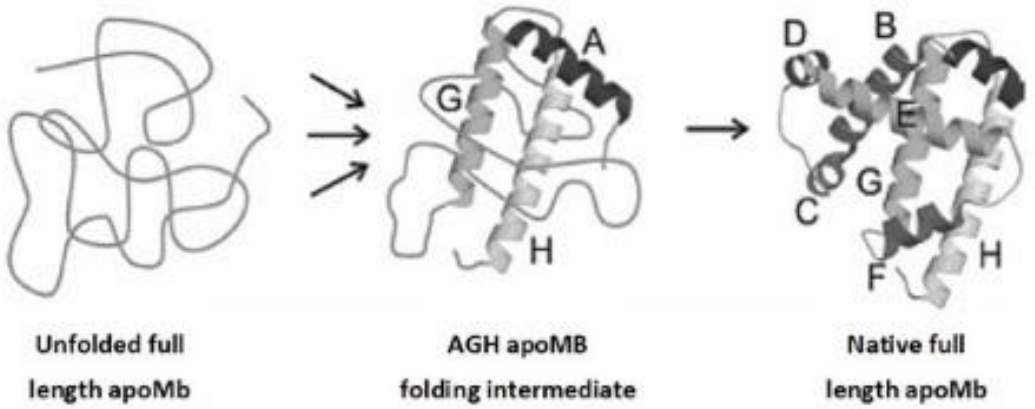


Figure 4.1 Schematic illustration of Apomyoglobin (ApoMb) folding [6].

The folding process of ApoMb undergoes a minimal 3-state-model which includes the molten globule intermediate state. Folding has been extensively studied in both kinetic and equilibrium experiments [116–121] and has used in computational modelling since the 1990s [122], [123]. There is evidence that during low pH-induced unfolding, ApoMb forms an equilibrium acidic intermediate (also called the *I* intermediate) at pH 4.2 with properties of both the folded state (pH 7) and the unfolded state (pH 2) [124–126]. This intermediate was first characterised by Uzawa *et al*. [124] using continuous flow, time-resolved circular dichroism, and SAXS (Small-Angle X-ray Scattering). The results suggested that the intermediate state has a radius of gyration (Rg) close to 23 Å and its helical content is around 33 %. These authors proposed that the helices present in the I-state are helices A, G and H, whereas the remainder of the molecule is disordered, as shown in Figure 4.1.

Recent advances in Terahertz (THz) technology have made the application of THz waves ranging from 0.1-10 THz (3.3-333 cm^{-1}) attractive in the fields of chemical and biomolecular studies. THz waves are particularly sensitive to water absorption which makes it possible to investigate water/biomolecules dynamics [51]; additionally, the sensitive amplitude and nonlinear vibrations of proteins detected in the THz frequency range are considered to be associated with protein function and protein-folding [54]. In particular, THz time-domain spectroscopy (TDS) is capable of obtaining the refractive index in addition to the absorption coefficient.

In this chapter, I determine the THz refractive index and absorption using THz-TDS spectroscopy of ApoMyoglobin in different states, benchmarked against traditional Ultraviolet–visible (UV)-circular dichroism measurements of the helical content of the protein.

4.2 Materials and methods

4.2.1 Materials

1 kilogram of 2-butanone and 1 gram of salt-free Mb from equine skeletal muscle was purchased from Sigma-Aldrich. The Mb was first purified by gel chromatography (HiLoad 26/60 Superdex75, GE Healthcare) using FPLC (Fast Protein Liquid Chromatography) with 10mM potassium phosphate buffer at pH 6.5 as eluent. The purity was checked by SDS-PAGE. Fraction samples containing the protein of interest were dialysed extensively against de-ionised water at 4 °C, and concentrated to 20 mg/ml. ApoMb was obtained from the above purified Mb by a heme group removal step using 2-butanone extraction method [127]. The resulting apoprotein was extensively dialysed against distilled water at 4 °C and concentrated to 20mg/ml,

which is the same concentration as the purified Mb protein sample. Both concentration of Mb and ApoMb were determined using the Beer-Lambert Law (as stated in Chapter 2.2) by measuring the absorption of the protein solution at a wavelength of 280nm. The Hitachi U-3010 UV-Visible spectrophotometer was applied at 20 °C using a molar extinction coefficient, at E409 of 160,000 M-1cm-1. ApoMb was monitored by absorbance at 280 nm, using molar extinction coefficient E280 at 13,700 M-1cm-1. Because the interaction of the heme in myoglobin will result in the Soret band, a strong absorbance peak will form in the visible spectrum at 409 nm. Hence, the measurement of absorbance at 409 nm occurs owing to the exposure of heme to the aqueous polar solvent. In our experiment, there was no absorbance in the 409 nm soret region showing the contamination of heme in the ApoMb solution, indicating the heme group in Mb protein was successfully extracted.

The different concentration solutions (ranging from 1.25 mg/ml to 15 mg/ml) were prepared from the 20 mg/ml Mb and ApoMb protein stocks by diluting with distilled water. Different pH condition solutions were obtained by adding a small amount of 0.1M hydrochloric acid (pH 5 to pH 2) or 0.1M sodium hydroxide (for pH 7); pH 6 was in its native protein solution. Reference buffers for each pH condition were prepared with the same amount of hydrogen chloride/ sodium hydroxide in distilled water.

4.2.2 Far-UV circular dichroism (CD)

The CD measurements were made with a Chirascan CD spectrophotometer (Applied Photophysics) with a 1 mm path length cuvette; absorbance scans were carried out from 190 to 250 nm at 0.5 nm intervals at a set temperature of 25 °C. Concentrations

were 100 μg/ml for all CD protein samples, and three trials were taken for each sample scan and then averaged. An absorbance scan of distilled water was also taken and used as a baseline to deduct the absorbance scan of samples. Data acquired from CD were analysed using de-convolution software, which interprets CD spectra and employs a complex algorithm to provide predictions of proportions of each secondary structure characteristic present.

4.2.3 Terahertz Time Domain Spectroscopy (THz- TDS)

The THz-TDS system used had its THz pulses generated and detected using 100 fs light pulses from a mode-locked Ti: sapphire laser at a wavelength of 800 nm and a repetition rate of 80 MHz [2]. The measurement was made at a controlled temperature 25±0.1 °C with a nitrogen purge that reduced the relative humidity (≤0.3 %) caused by water vapour. A Bruker liquid cell (A145) was utilised as the sample holder for the protein solutions. The cell consisted of two poly-4-methyl pentene-1 (TPX) plates and a 100 μm polytetrafluoroethylene (PTFE) spacer. 50 μl of solutions were injected into the cell with a pipette before each measurement, and a wash and the pump empty procedure (as mentioned in Chapter 2) was applied to each measurement to ensure the thickness of the samples was constant during the experiments.

4.3 Estimation of apomyoglobin secondary structure from CD spectra

Figure 4.2 shows the CD spectra of acid-induced (pH 2 to 7) ApoMb recorded at wavelength 190-260 nm; 222 nm was used to monitor the unfolding process. In Figure 4.2A, spectra at pH 6 and 7 exhibit a typical spectrum for alpha-helical proteins, which is characterised by two negative peaks at around 222 and 208 nm respectively, with a

positive peak around 190 nm. The pH 5 spectra shows smaller negative bands of similar magnitude at 222 and 208 nm, and a lower positive band at 192 nm, which still suggests the presence of a considerable amount of alpha-helical residual secondary structure, but with a decrease in alpha-helical content. pH 2 and 3 spectra have a negative band of great magnitude at around 200 nm and exhibit no characteristic peak at around 220 nm, as expected for a largely unfolded conformation. The CD signal at 222 nm, which is proportional to the alpha-helical protein content, was plotted in Figure 4.2B. It is retrieved from Figure 4.2A as a function of pH decrease to indicate the absorption changes due to unfolding. Overall the plot demonstrates an increase of CD signal with decreasing pH, firstly the CD signal was slightly increased from pH 7 to 6; the molecular transition reflecting a disorganization of secondary structure occurred when pH is below 6 and was followed by a clear increase trend of CD signal from pH 5 to 3, indicating gradually unfolding of apoMb under these pH conditions. Then a complete loss of secondary structure happened under pH 2, showing that apoMb adopted a complete unfolded state.

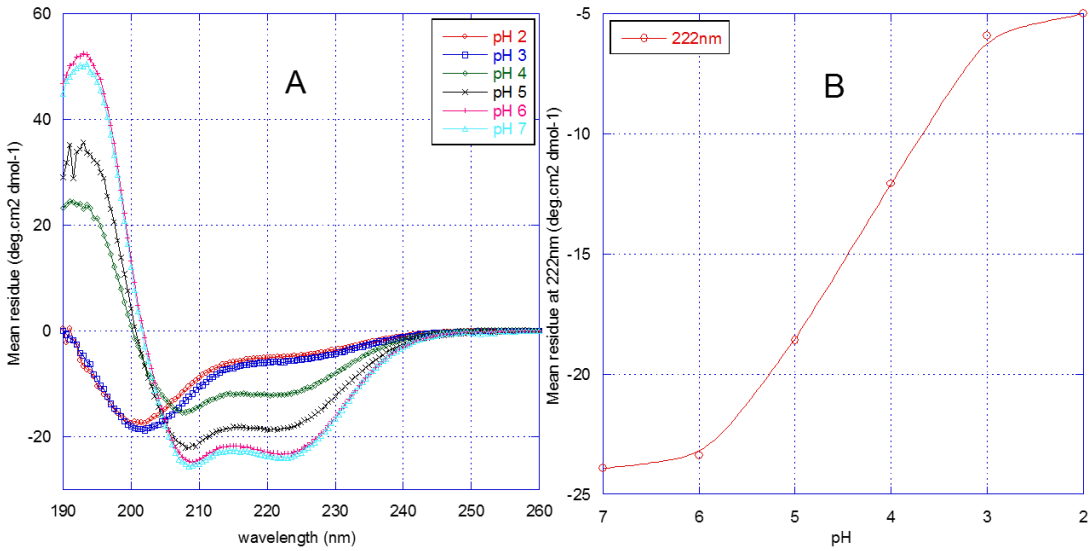


Figure 4.2 The unfolding transition of 0.1mg/ml ApoMb under acidic denaturation. (A) CD spectra of ApoMb as a function of acidic conditions (pH 2 to pH 7) in 190 nm to 260 nm wavelength range; (B) Acid-induced unfolding transitions curves (monitored at 222 nm), to indicate the absorption changes due to pH decrease.

Additionally, CD spectra were deconvoluted using software CDNN [84] to estimate their alpha-helical contents under different conditions. Table 4-1 presents the calculated ApoMb secondary structure composition at different pH values, which agrees with the spectra observed, and correlates with the known unfolding trend of ApoMb alpha-helical structure, although the percentage of helical content differs slightly from previous measurements of sperm whale apoMb secondary structure composition [83]. The percentage of α-helical contents was calculated using data in the ranges of 190-260 nm, 195-260 nm and 200-260 nm separately.

pH	190-260nm (%)	195-260nm (%)	200-260nm (%)
7	84.8	85	85.6
6	83.4	82.8	81.8
5	68	69.9	72.5
4	45	44.6	43.8
3	20.2	19.9	22.1
2	18.2	17.9	18.5

Table 4-1 The pH dependence of α-helical content of ApoMyoglobin using the de-convolution software CDNN [22].

The CD spectra in Figure 4.3, made using the same conditions of the native states of Mb and ApoMb, show spectral differences in the wavelength range 200 nm to 220 nm. The spectra of ApoMb decreases less than Mb at negative band 208nm, which suggests ApoMb contains less alpha-helical content than Mb. Although the ApoMb has approximately the same conformation as Mb, removal of the heme group could cause a change in the F-helix, which is stabilised by contacts with the heme [2].

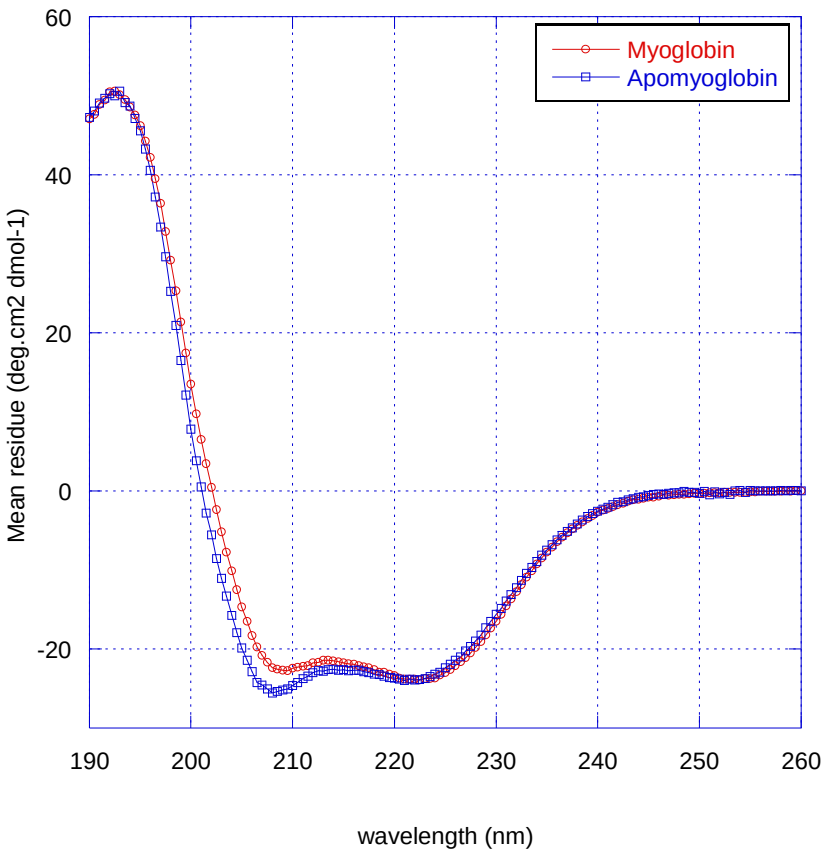


Figure 4.3 CD absorption spectra of Mb and ApoMb at pH 7 under the same condition (0.1mg/ml in 20mM potassium phosphate buffer at 25°C). The change was observed between 200 nm and 220 nm range, and the maximum shift is at 208 nm.

4.4 THz spectra of pH-dependent ApoMb

THz absorption and refractive index spectra for ApoMb were measured at six pH values and two concentrations (Figure 4.4). Absorption spectra are shown in panels A and C, and the refractive index is shown in panels B and D, for 10 mg/ml and 20 mg/ml ApoMb. Corresponding baselines were measured using buffer only.

According to figure 4.4, the spectra show approximately linear increase in absorption in 0-1.2 THz and 1.5-2.2 THz range with increasing THz frequency, the refractive index spectra is generally a decrease trend in the frequency range from 0.2 to 2.2 THz. Both spectra trend of the THz absorption and refractive index show quite similar

comparing the figures of two concentrations (A and C, B and D), which suggests it is not sensitive enough to be able to detect the concentration-dependent difference in this THz spectra result. The result may because that the strength of the signals in VNA is relatively weak and water interactions in the hydration layer surface could limit the detection of variation in different concentration conditions. The background noise is averagely less than 1 % in these measurements.

The blue line in figure 4.4 represents spectra of buffer and the red line for protein. Comparing the protein sample with buffer reveals no distinctive feature or trend that can be assigned to the presence of protein, except in pH 6 the refractive index spectra of buffer is slight higher than that of the protein. However, it is not convenced the THz spectra are sensitive to detect refractive index variation in this pH conditions.

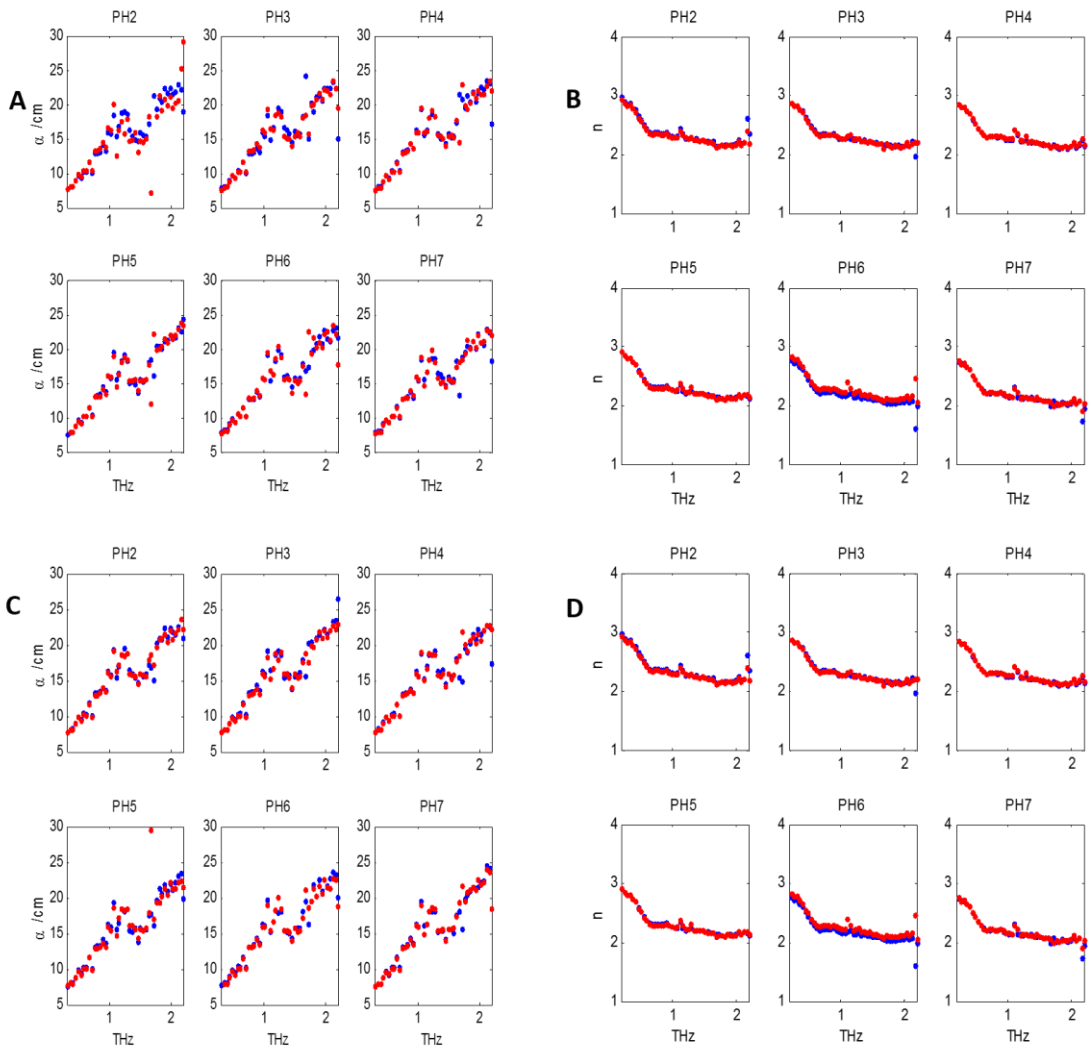


Figure 4.4 THz transmission spectra of 6 acidic ApoMb solutions (pH 2 to pH 7) compared with its relative buffers in frequency range 0.2-2.2 THz. The blue line is for buffer and the red line for protein. (A) and (B) show THz absorption coefficient and refractive index for 10 mg/ml protein concentration conditions respectively; (C) and (D) for 20 mg/ml.

The THz fractive index spectra of both proteins and buffers from Figure 4.4B and 4.4D have been reploted in figure 4.5 as a function of pH. Figure 4.5A presents the refractive index spectra of ApoMb protein with 10mg/ml concentration sample measured in a range of pH from 2 to 7, while figure 4.5B shows the spectra with 20mg/ml protein sample. The spectra of corresponding controls which are buffers balanced at the same

pH as the samples are presented in Figure 4.5C and 4.5D.

The similarity of the spectral changes for Figure 4.5B and the control Figure 4.5D suggests that the change in refractive index is due to the quantity of acid added to the solution rather than to some protein-solvent interaction underlining the significance of water in determining the THz signal.

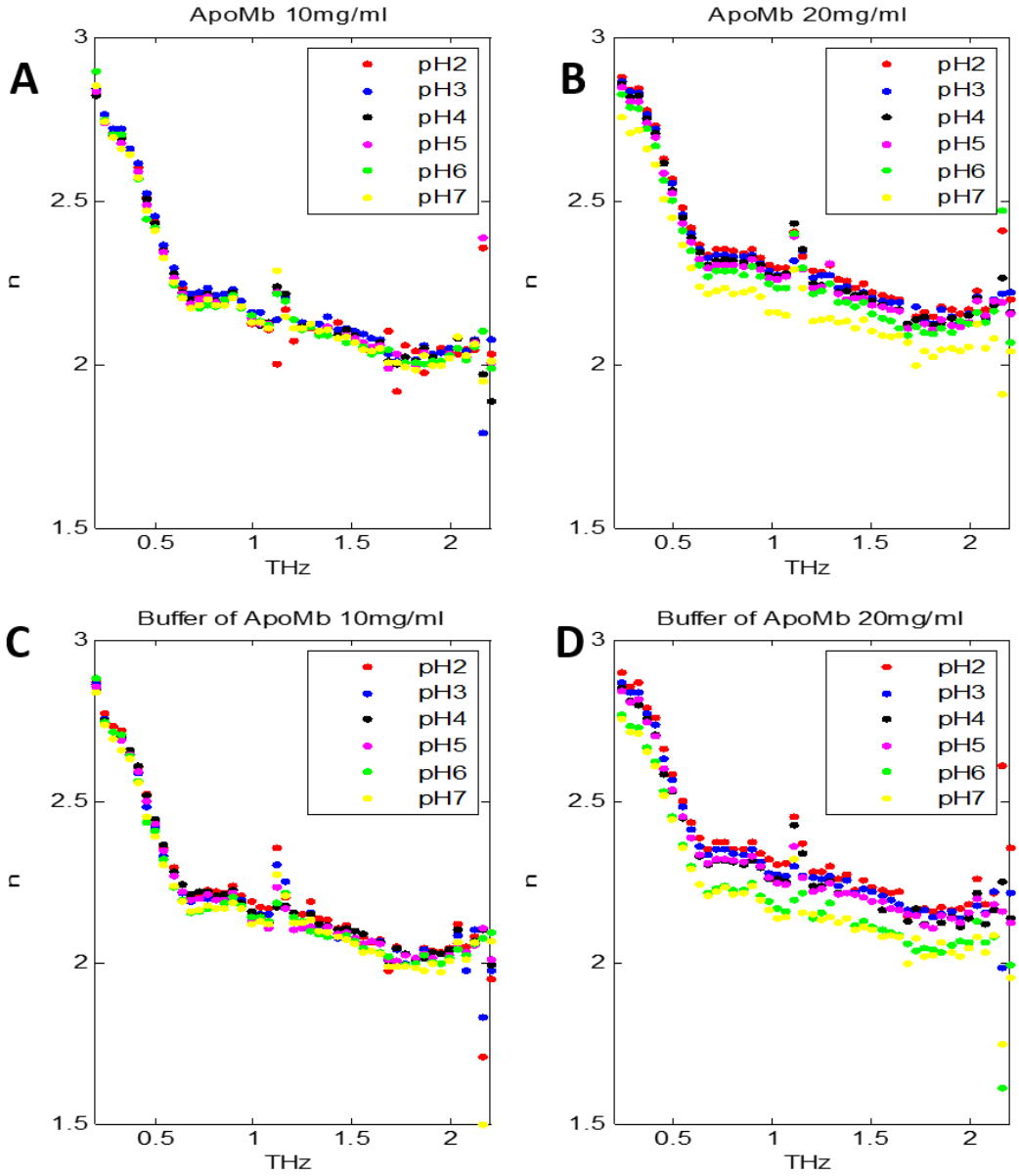


Figure 4.5 THz refractive index spectra of pH-dependent ApoMb protein as a function of pH measured in the 0.2-2.2 THz frequency range, with two concentrations 10 mg/ml (A) and 20 mg/ml (B). Panels C and D are the corresponding controls.

When the two protein measurements (20 mg/ml and 10 mg/ml) are compared (Figure 4.6), no significant difference in absorption can be discerned, but there is a difference in the refractive index of the two samples at lower pH values. The difference can be clearly seen at pH 2 and decreases as pH increases to pH 7 where the difference is no longer visible (Figure 4.6D). A similar effect can be seen using buffer only. It should be remembered that the buffer was prepared by adding the quantity of HCl necessary to pH the protein/buffer solution. The major effect again appears to be due to the quantity of HCl added, and the effect in turn has on the amplitude and phase of the THz signal. At pH 6, there could be a protein-solvent effect as the corresponding panels in Figure 4.6D and 4.6B are not identical.

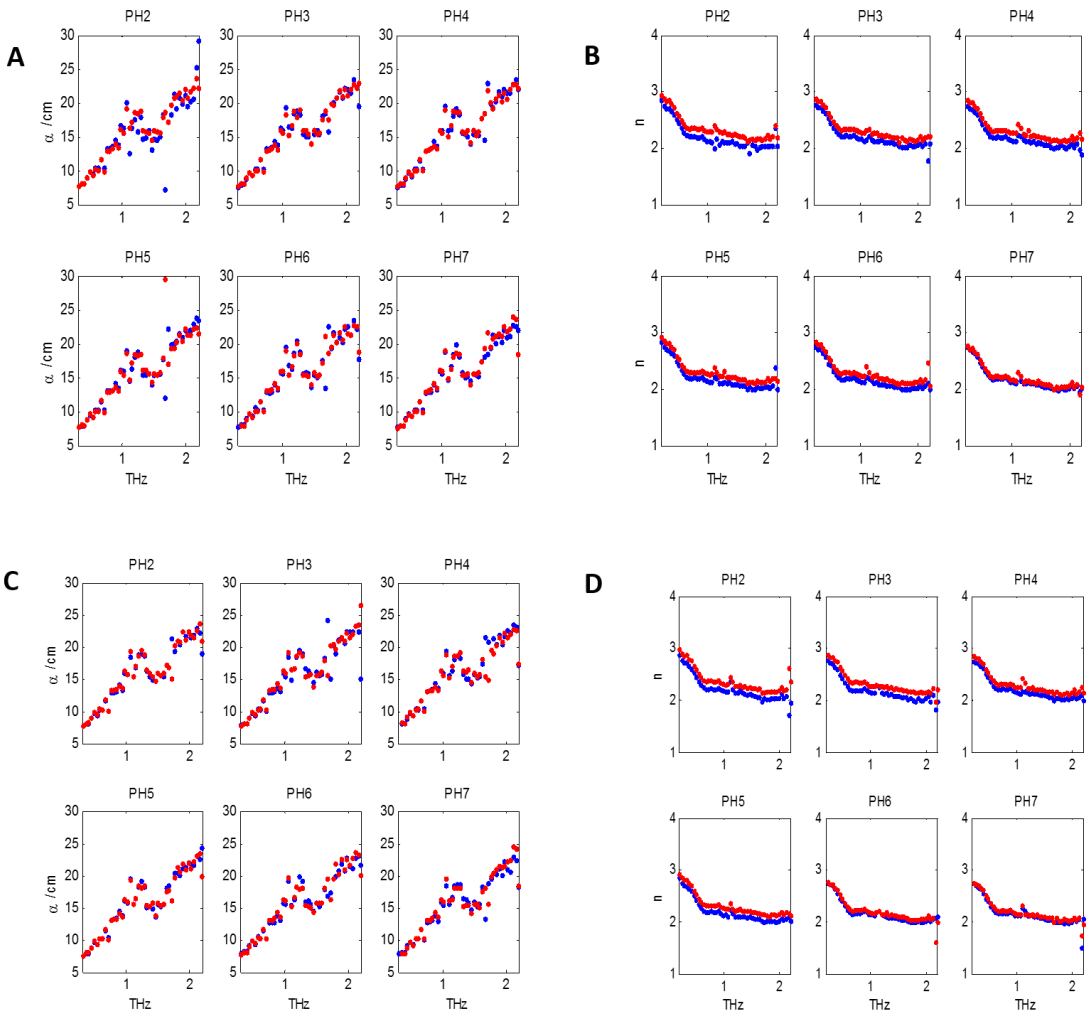


Figure 4.6 THz transmission spectra of two ApoMb concentration, 20 mg/ml (red line) and 10 mg/ml (blue line), under different pH conditions from pH 2 to pH 7 in 0.2-2.2 THz frequency range. (A) and (B) shows THz absorption coefficient and refractive index for buffer; (C) and (D) for ApoMb.

4.5 THz spectra of concentration dependent Mb and ApoMb

Measurement of the refractive index of Mb and ApoMb at different concentrations and pH 6 suggests a difference in the amplitude and phase response of these two proteins.

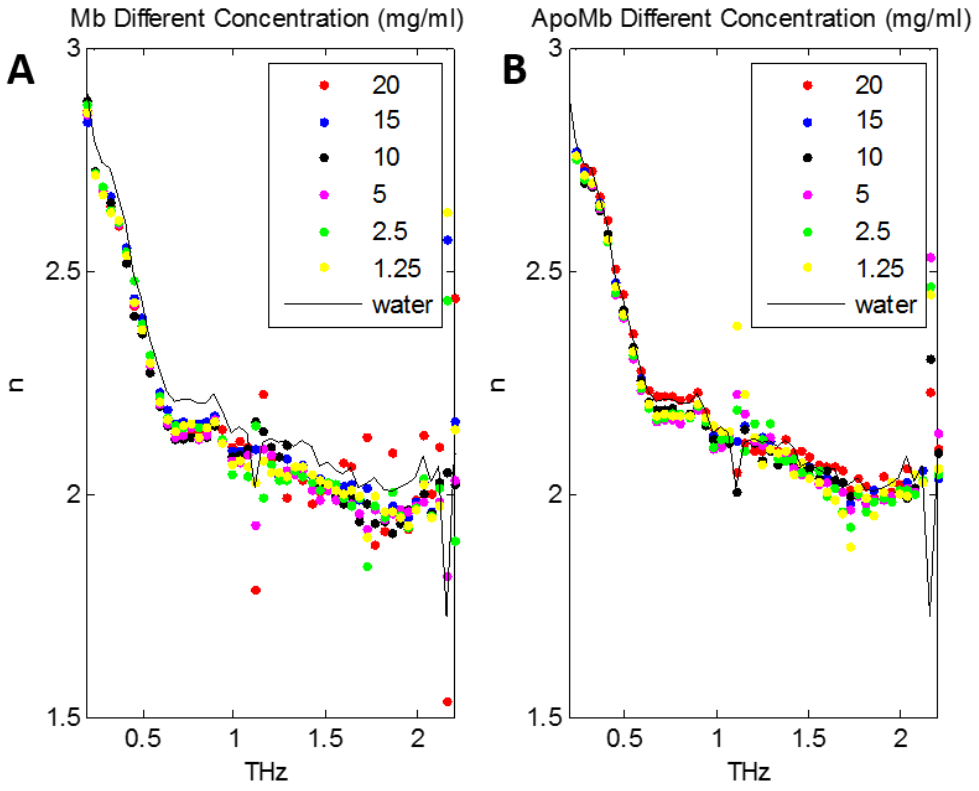


Figure 4.7 THz refractive index spectra of solvated protein Mb (A) and ApoMb (B) under concentration variations from 1.25 mg/ml to 20 mg/ml. The spectrum of water is presented as a black line. All measurements were at pH 6.

There is no clear evidence that the refractive index varies as a function of protein concentration. However, when the spectra of proteins are compared to water, as presented in Figure 4.7, it can be clearly seen that ApoMb is consistent with the water spectra while the refractive index of Mb is lower than that measured for water in all concentrations. As the difference between Mb and ApoMb is the removal of the heme group in ApoMb, this could be attributed to the presence of the heme. As the THz signal is dominated by the behaviour of water molecules, it is likely that this is not a

direct consequence of the presence or absence of heme, but rather an effect of the removal of the *heme* on the flexibility and dynamics of the protein. The presence or absence of the *heme* can apparently be detected at protein concentrations of 1.25 mg/ml.

4.6 Summary

I was expecting to able to probe the molten globule state and unfolded states of apomyoglobin using THz time-domain spectroscopy. These states were successfully detected using UV-circular dichroism spectroscopy as described above.

THz absorption and refractive index spectra were measured successfully for a range of solutions for ApoMb (different pH values and different concentrations), but it was not possible to detect a difference in the response of the solutions with protein from those without protein. The importance of the quantity of added HCl on the THz response of the water was the dominant feature revealed in these initial studies. There is an indication of protein involvement in the response (pH 6.0), but the evidence is not overwhelming.

In the last experiment, which involved comparison of the refractive index of Mb and ApoMb, a difference was detected. The refractive index of Mb was different from that of water and that of ApoMb, which was essentially the same as water. This measurement offers some hope for the future that differences can be measured using these techniques – differences that may provide information on the dynamics of Mb compared to ApoMb.

Chapter 5 *In vitro* production of bacterial micro-compartment nanotubes

5.1 Introduction

Compartmentalisation derives its significance from the fact that it enables biochemical pathways to be segregated; also, in synthetic biology, compartmentalisation may serve as an important tool that could be used to generate defined microenvironments for the production of particular products [128]. Bacterial microcompartments (BMCs) are produced when bacteria are grown on a particular substrate; they exist in a large number of bacteria [129–131]. These polyhedral cellular assemblies were first found in cyanobacteria and chemotrophic bacteria by Transmission Electron Microscopy (TEM) [132]. To date, four types of bacterial microcompartments have been extensively characterised and well-established functionality studied, including carboxysome, ethanol utilisation (Etu), ethanolamine utilisation (Eut), and 1,2-propanediol utilisation (Pdu) bacterial microcompartments [128]. A common feature of the BMCs is the proteinaceous shell which encapsulates functionally related enzymes, substrates, toxic and labile intermediates [133–137].

These shell-proteins can be easily recognized by their conserved protein fold known as the bacterial microcompartment (BMC) domain [138], which constitutes the major part of the shell protein subunit. A subunit may involve one or two domains; the majority of shell-proteins consist of a single BMC domain (Pfam 00936) in each subunit and self-assemble into a hexamer, while some shell proteins have a tandem

repeat of the Pfam 00936 and lead to the formation of pseudo-hexameric trimers [23], [139], [140].

The propanediol utilisation (Pdu) microcompartment shell is built from thousands of hexagonally shaped protein oligomers made from seven different types of protein subunits: PduA, PduB, PduJ, PduK, PduN, PduT, and PduU [128], [141], [142]. In previous studies, it was found that not all of these seven *pdu* genes encoding shell-proteins are necessary for the formation of heterologous microcompartments in *Escherichia coli* [143], among these, PduA, PduB, PduJ, PduK and PduN are the minimum shell components to form a non-aberrant empty Pdu microcompartment [143]. Among these major shell components, PduA is reported to play a crucial role in coordinating with other shell-proteins as indicated by the protein-protein binding studies, allowing the newly synthesised shell-protein to interact with other subsequently translated shell components [143]. Moreover, by deleting PduA from a construct harbouring the minimum number of genes for the formation of empty BMCs, elongated filamentous structures will form, and the construct no longer resembles bacterial microcompartments [143]. This suggests that PduA could potentially act as a scaffold for the assembly of the microcompartment [128]. Interestingly, PduA is found to be able to form filamentous structures when overproduced *in vivo* under observation by transmission electron microscopy (Figure 5.1). A soluble and more stable construct from PduA, the PduA*, obtained with a mutated C-terminal residue and extra C-terminal 23 residues [23], shows a greater density of production of tube-like structures within the *E. coli* cell and brings us a clearer and greater visual impact in micrographs [144].

The crystal structure of PduA from *Salmonella enterica*, (PDB: 3NGK) [145], revealed that the PduA subunits pack closely together to form a biologically authentic hexamer. Of the ~20,400 $Å^2$ total surface area of the isolated PduA subunit, ~9600 $Å^2$ is buried in the hexamer. The native PduA hexamers in the crystal have 6-fold rotational symmetry (Figure 5.2) and tile within the crystal lattice (space group P622) with adjacent hexamers separated by 67.2 Å (the *a* axis of the cell). This interaction is repeated to form a sheet of molecules [128].

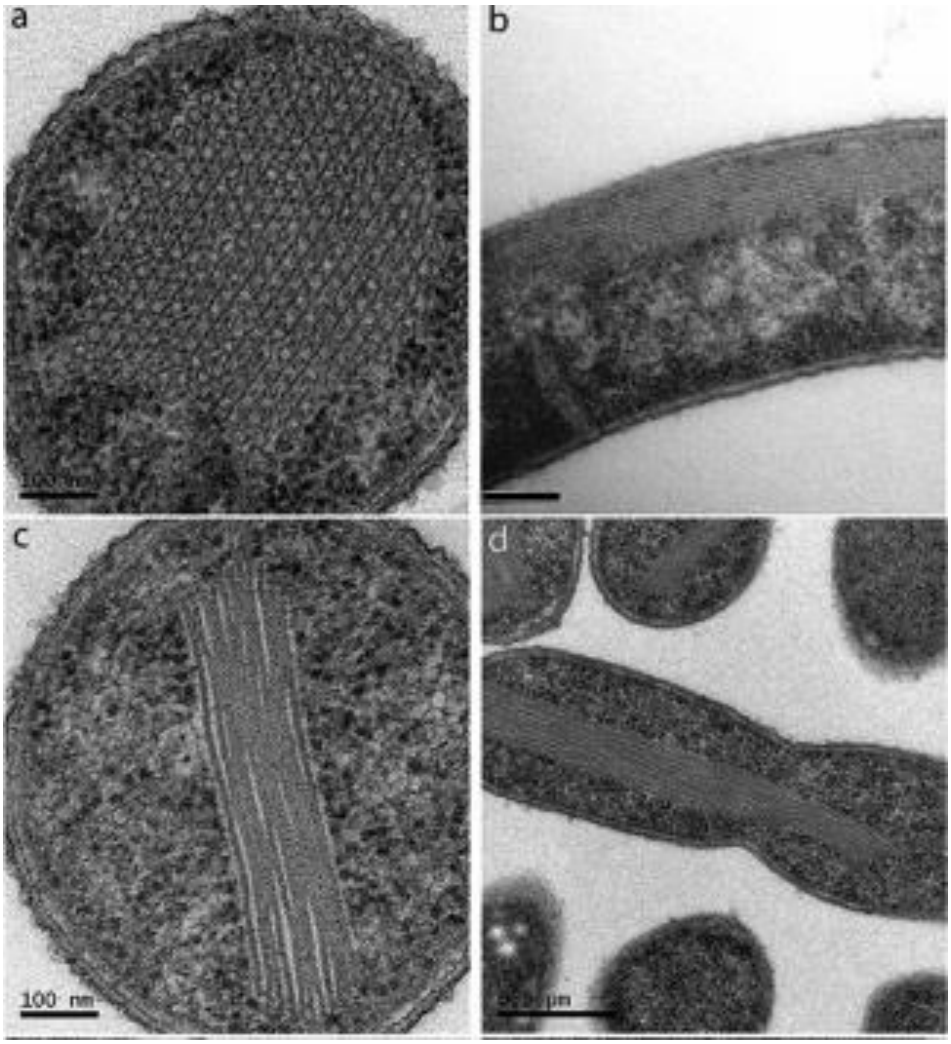


Figure 5.1 Transmission electron microscopy of native PduA (panels, a and b) and hexamer-hexamer interaction mutants (c and d) produced in E. coli cells. Images on the left show transverse sections, and images on the right show longitudinal sections through the cell (adapted from [24]*). The scale bar for panels a and c is 100 nm; panel b and d 500 nm.*

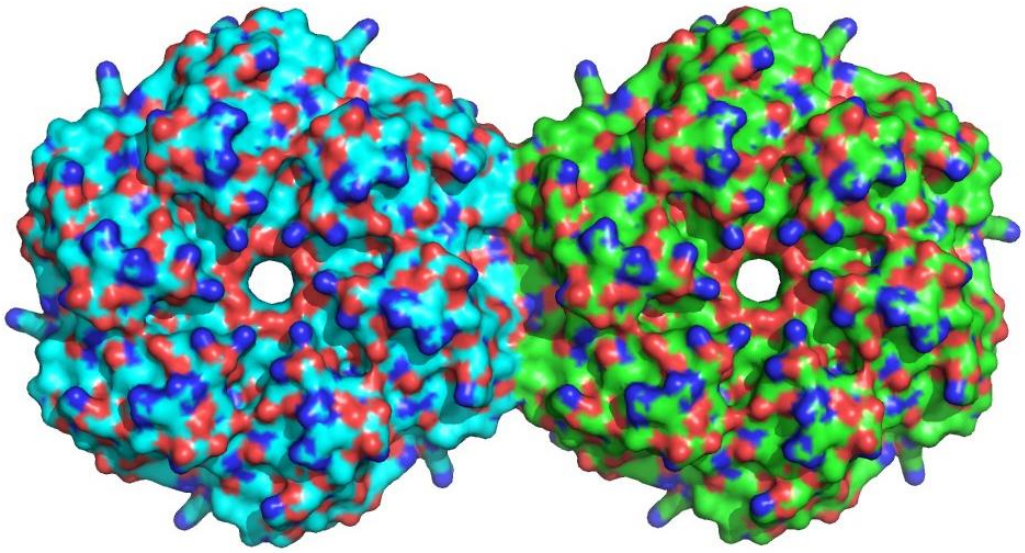


Figure 5.2 Crystal structure of native PduA (Protein Data Bank accession code 3NGK). Figure is adapted from [24]. Six PduA subunits form a hexamer with a six-fold rotational symmetry, and adjacent hexamers tile a surface. A surface representation of two such PduA hexamers, one with cyan carbon atoms and one with green, one hexamer-hexamer interface is represented [24].

In this Chapter, it is shown that PduA* hexamers can self-assemble *in vitro* into protein nanotubes. The PduA* nanotubes produced were investigated using the VNA and quasi-optical bench THz system, to explore the potential of using THz probe radiation to study nanostructures.

5.2 Protein production

5.2.1 Generation of PduA protein and PduA K26D mutants*

The cloning of PduA* into pET3a (PduA* is a more soluble form of *Citrobacter freundii* PduA with 23 additional C-terminal residues; sequences are shown in Appendix 1. The correct sequence of the constructs was verified by DNA sequencing

(Eurofins). PduA* K26D was sub-cloned into pET14b to facilitate PduA* K26D overproduction as a fusion protein with an N-terminal His_6 tag in *E. coli.* The cloning and expression procedures was contributed from the work of Dr. Allan Pang. The construct was supplied from the QMUL lab before being transformed into BL21 cells for protein expression.

E. coli BL21(DE3) pLysS cells containing chloramphenicol were transformed with PduA* and the PduA K26D mutant constructs, the starter cultures were prepared by inoculating bacterial glycerol stock (mix equal amounts of bacterial culture and 80% glycerol) to 10 ml Luria Bertani (LB) media containing ampicillin with final concentration at 100 μg/ml. Cells were allowed to grow at 37℃ overnight. 1L LB media supplemented with ampicillin was inoculated with the starter culture and shaken at 37℃ until an OD_{600} of 0.6-0.8 is reached. IPTG (isopropyl β-D-thiogalactoside) with final concentration at 400 μM was added to induce protein production. The bacterial culture was left shaking overnight at 16℃.

The bacterial cultures were harvested via centrifugation (Avanti J-30I, Beckman Coulter) with rotor speed of 8,000 rpm (G force at 15,900 x g) for 20 minutes. Collected cell pellets were re-suspended in lysis buffer (50 mM Tris-HCl, 50 mM NaCl at pH 8.5, 10 mM imidazole) then lysed using a sonicator at 20 watts for 20 seconds with a resting time of 1 minute between repetitions of this procedure. This procedure was performed for 6-8 times. It is important that cell lysis procedure is kept on ice to prevent overheating as a result of sonication of bacterial cells. Following lysis, the cell debris was sedimented by centrifugation (Avanti J-30I, Beckman Coulter) at 18,000

rpm rotor speed (G force at 47,900 x g) for 30 minutes. The supernatant containing the protein of interest was collected for purification.

A histidine tag attached to the proteins was used for purification by immobilised metal affinity chromatography (IMAC). An PD-10 column was filled with 5ml Chelating Sepharose Fast Flow resins. IMAC method is designed for purifying native and recombinant protein and peptide that have an affinity to metal ions, therefore will selectively retain proteins with exposed histidine. The resins were washed with deionised water to remove ethanol, then subsequently charged with 100 mM nickel sulfate. Charged resins were acclimatised to the binding buffer before use. After loading the supernatant onto the column containing charged resins, an increasing concentration of imidazole (50 mM, 100 mM, 200 mM, 400 mM) was applied to remove non-specific binding. The samples were collected in 400 mM imidazole elution buffer. The proteins were more than 90 % pure, as determined by polyacrylamide gel electrophoresis (Figure 5.3). Proteins eluted out from the immobilised metal affinity chromatography consist of a high NaCl and imidazole concentration; thus, buffer exchange is carried out via a desalting column (HiPrep 26/10) to remove high imidazole and salt. PduA* and PduA K26D was further purified, and equilibrated with 50 mM Tris-HCl, pH 8.5, 50 mM NaCl.

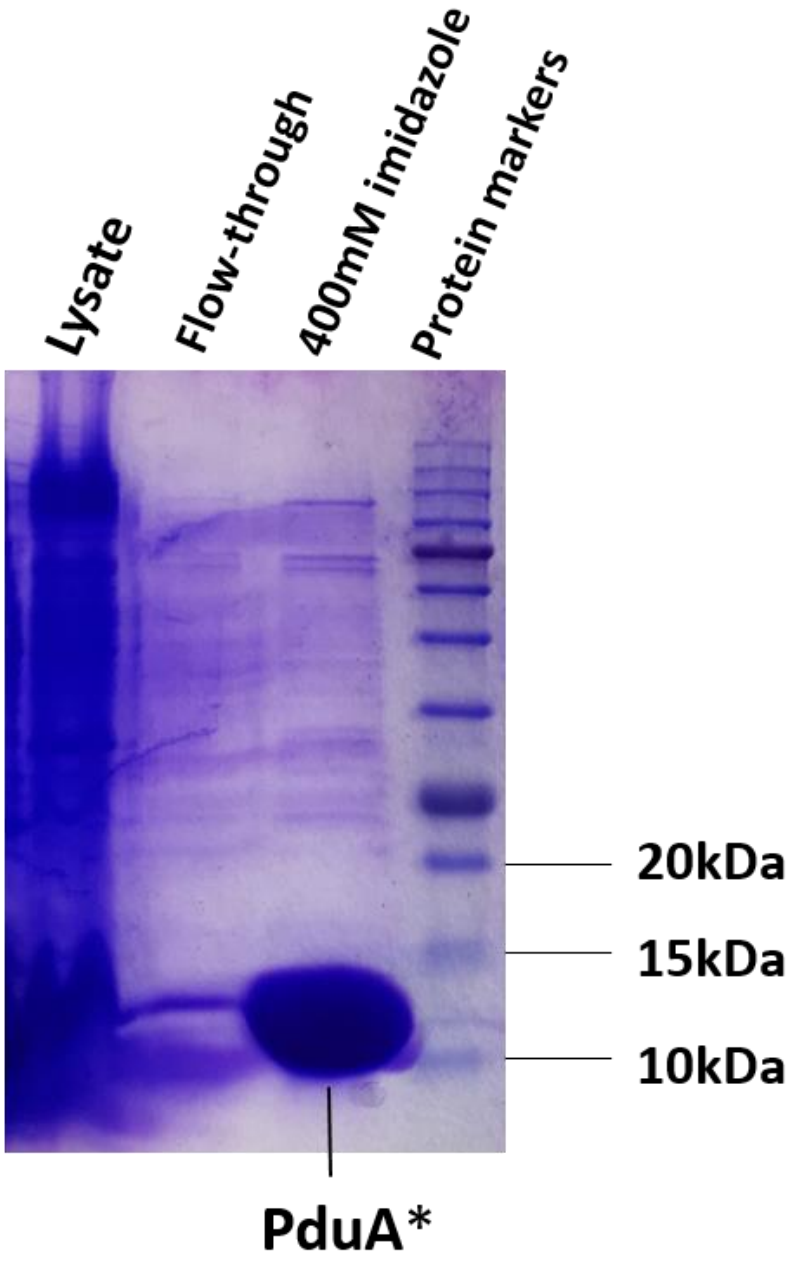


Figure 5.3 SDS-PAGE (15% gel) results from protein purification obtained with 400mM imidazole elution by IMAC. The flow-through contains proteins without binding, leading to enrichment of PduA protein in the 400 imidazole gel lane. The thick band of PduA* was found at around the size of 12 kDa that corresponds to the subunit size of PduA*(11.9 kDa).*

5.2.2 Generation of PduA nanostructures in vitro*

With the help of TEM micrographs, initially, very few tubes-like structures were captured forming on the TEM grids. The PduA* protein was purified via nickel affinity purification, and accommodates in 50 mM Tris, 500 mM NaCl, 400 mM imidazole buffer solution. A clear appearance of tubes started to be observed after buffer exchanged to low salt condition then concentrating to a higher concentration. Here glow-discharged carbon-coated 300-mesh copper grids were prepared using the droplet method, where 10 μl aliquots of PduA solution samples were adsorbed for 1 minute followed by 10 μl of 1 % (w/v) of phosphotungstic acid (PTA) adjusted to the same pH with the proteins, to negatively stain the sample. PTA stain was observed within the lumen of the nanotubes, suggesting that these structures were nanotubes. The outer diameter of the extracted tubes was 25.95 ± 50 nm (n = 50), and an inner diameter of around 25 nm which has an approximate value (n = 50). These values suggest the nanotube walls are possibly composed of a single layer of hexamers. Tube lengths varied, ranging from a few hundred nanometers to tens of microns. Extracted tubes were observed either as bundles (Figure 5.4A), individually (Figure 5.4B), or in layers and disassembled forms (Figure 5.4C). As a comparison, the mutant protein PduA-K26D was captured not forming tube-like structures, which is in agreement with previous TEM observation within *E. coli* cell.

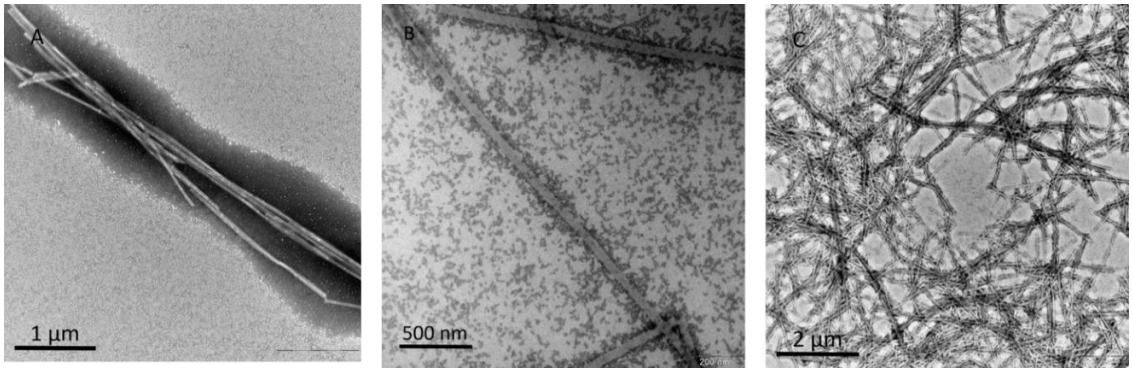


Figure 5.4 *Negatively stained TEM micrographs PduA nanotubes. (A) shows nanotubes formed in a bundle; (B) depicts individual nanotubes, which were observed less frequently than nanotube layers and bundles; and (C) shows nanotubes in layers in varied forms, over a wide scale.*

5.2.3 Concentration effect on nanotube formation

It was realized that the protein concentration itself and the salt condition might present two distinct key factors influencing nanotube formation. The impact of concentration was firstly evaluated by a simple concentration and dilution method. Negative stains of preparations of PduA* samples in concentrations ranging from 0.1 to 7.2 mg/ml revealed a progressive accumulation of assembled nanotubes as PduA* concentration increased, as shown in Figure 5.5 A-D.

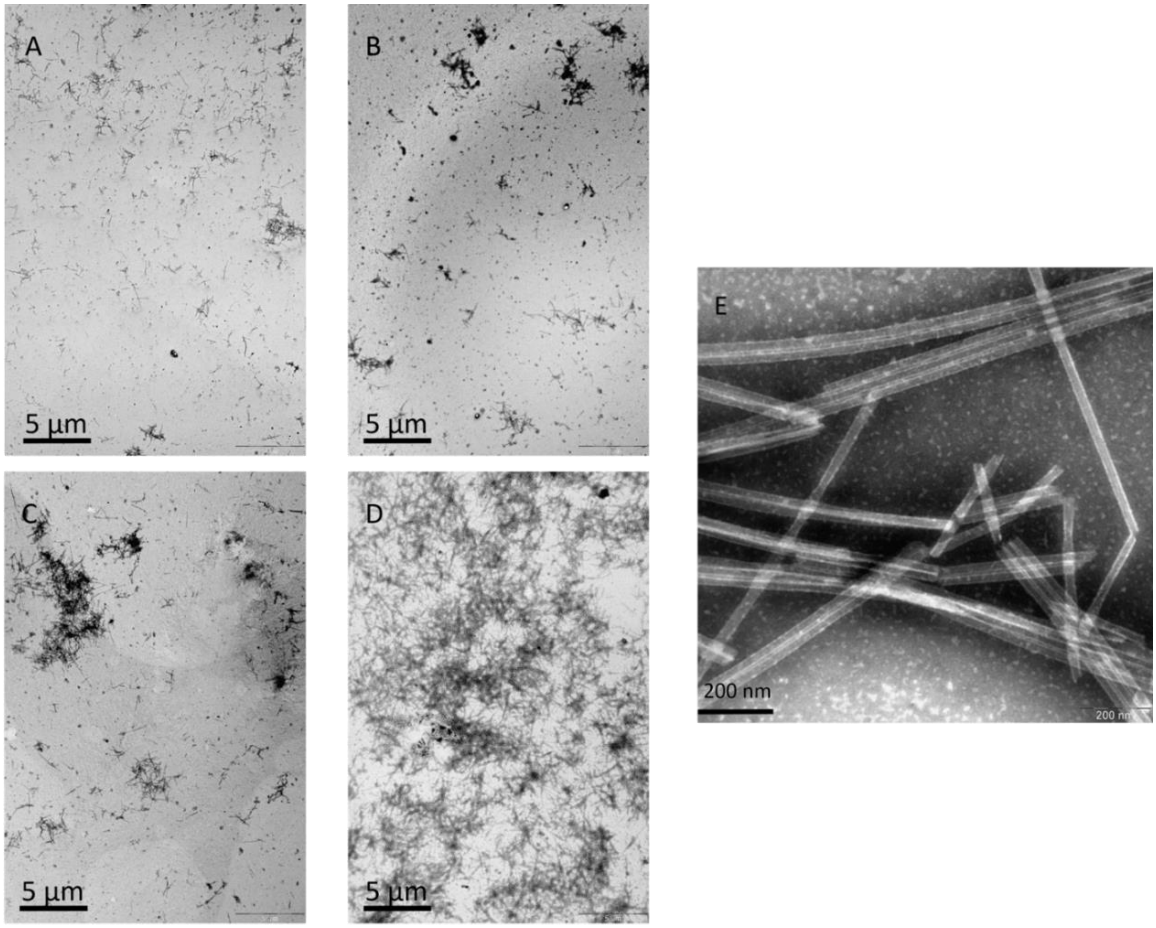


Figure 5.5 Progression of PduA nanotube formation as visualized by representative TEM micrographs. The four images A-D shows protein solutions prepared on grids in concentrations of 1 mg/ml, 3 mg/ml, 5 mg/ml and 7.2 mg/ml, respectively. The progression shows PduA* nanotubes assembled with an increase of concentration. E presents the existence of nanotubes under diluted solution (0.1mg/ml) from D (7.2mg/ml).*

Additionally, a 0.1 mg/ml nanotube sample in Tris-HCl buffer (50 mM Tris, 50 mM NaCl, pH 8.5) was concentrated to 7.2 mg/ml with the same buffer, resulting in nanotube formation. By diluting the sample back to 0.1 mg/ml, the resulted nanotube was still visible (Figure 5.5E), suggesting that the nanotubes produced are soluble and stable once the tube has formed.

5.2.4 The effect of ionic strength on nanotube formation

Purified nanotube integrity was also examined as a function of ionic strength by changing the NaCl salt concentrations and imidazole conditions, for which the nanotubes displayed various degrees of order, (see Figure 5.6). All these samples were collected from the same PduA sample solution with nanotubes already formed (0.1 mg/ml concentration, 50 mM Tris-HCl buffer, pH 8.5), and were only treated differently with varying concentrations of salt and imidazole. In summary, stable nanotubes assembled stable in low salt in either low or high imidazole condition, rather than high salt conditions. This suggests the increase in ionic strength of NaCl salt may cause disassembly of PduA* tubes. Collectively, this accords with recently reported observations made for isolated BMC shells in response to buffer conditions [146], [147].

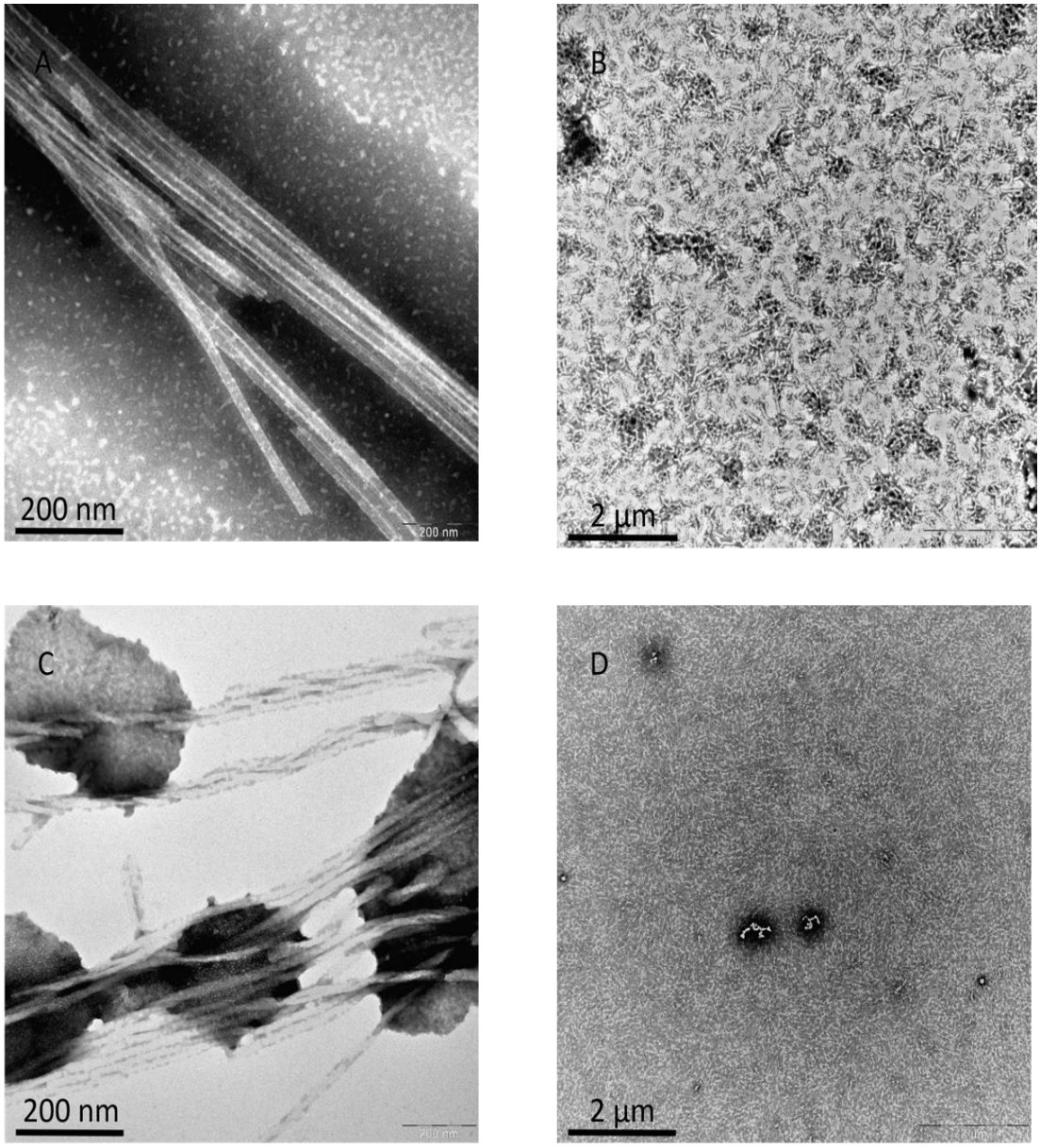


Figure 5.6 TEM micrographs of PduA nanotubes under different NaCl salt and imidazole conditions, showing the effects of ion strength on PduA assembly into nanotubes. (A) Nanotubes in 50 mM Tris-HCl, 50 mM NaCl, 50 mM imidazole, pH 8.5 (low salt, low imidazole), nanotubes exhibited bundling, and individual tubes were stable, mainly long length; (B) Nanotubes in 50 mM Tris-HCl, 50 mM NaCl, 400 mM imidazole, pH 8.5 (low salt, high imidazole), nanotubes are characterised as short and chopped; (C) Nanotubes in 50 mM Tris-HCl, 500 mM NaCl, 400 mM imidazole (high salt and high imidazole condition), the nanotubes are mostly formed in longer twisted bundles; (D) Nanotubes in 50 mM Tris-HCl, 500 mM NaCl, 50 mM imidazole (high salt and low imidazole), nanotubes were almost entirely disassembled and barely to be found.*

5.2.5 Determining the position of hexa-histidine tag in the nanotubes

When the hexamers tile the surface, for instance, the surface of a cylinder to form a nanotube, there is likely to be a single or preferred orientation of the hexamer. The consistency of the diameter of the tube suggests the orientation is consistent, but which side faces out? The His-tag is on one side of the hexamer, so if it can be convincingly located to the inside or outside, this will determine the orientation of the hexamer. Chelating Sepharose Fast Flow resins, charged with nickel sulphate, were used to bind to the hexahistidine tag.

100 μl Chelating Sepharose Fast Flow resins were firstly charged with 1 ml nickel sulphate, washed with deionized water and acclimatized to the binding buffer. 0.5 ml of 0.1 mg/ml PduA* solution containing nanotubes was loaded and incubated for 15 mins at 4 °C. The beads were further washed twice with washing buffer (50Mm Tris-HCl, pH 8.5) which containing 10 mM imidazole (both washing buffers were kept for examination), and finally eluted with 100 μl elution buffer (50mM Tris, 500mM NaCl, pH 8.5) containing 400 mM imidazole. As a comparison, the eluted sample was tested with both high salt (500 mM NaCl) and low salt (50 mM NaCl) in the sample buffer (50 mM Tris-HCl, pH 8.5).

With the assumption that, if the hexamer-histidine in PduA* nanotubes are externally presented, we would expect that the imidazole side chain of the hexahistidine tag in the PduA* protein to bind with the nickel-charged resins and the nanotubes will be eluted only in the final highest concentration imidazole. In contrast, if the histidine-

tags in PduA* nanotubes are internal, nanotubes will be washed away easily and left in the washing elution buffers.

BioRad assay [148] was used to examine whether the PduA* exists in the buffers. It is a protein determination method that involves the binding of protein-dye for quickly checking the global protein concentrations). The dye will perform as protonated red cationic form under acidic conditions (red colour), and convert to a stable anionic form (blue colour) when binding with proteins [33], [149]. The blue protein-dye form can be detected at 595 nm in the assay using a spectrophotometer or microplate reader. In our experiment, the red colour, which is the cationic form was shown in the 50 mM and 100 mM imidazole solutions, suggested there were no proteins binding to the two washes. In 400 mM imidazole elution buffer, the blue protein-dye form was observed, suggesting that proteins were washed off and existed at the 400 mM imidazole elution buffer, as shown in Figure 5.7. A further check applied the eluted samples with 400 mM imidazole in low/high salt conditions as TEM grids, and tubes was found in 400 mM imidazole buffer conditions, while short and chopped tubes are formed in low salt condition, low and coiled tubes in higher salt condition. Both results provide evidence that the hexahistidine tag is external, in agreement with the proposed model of the nanotube making use of the wedge-shape of PduA hexamers.

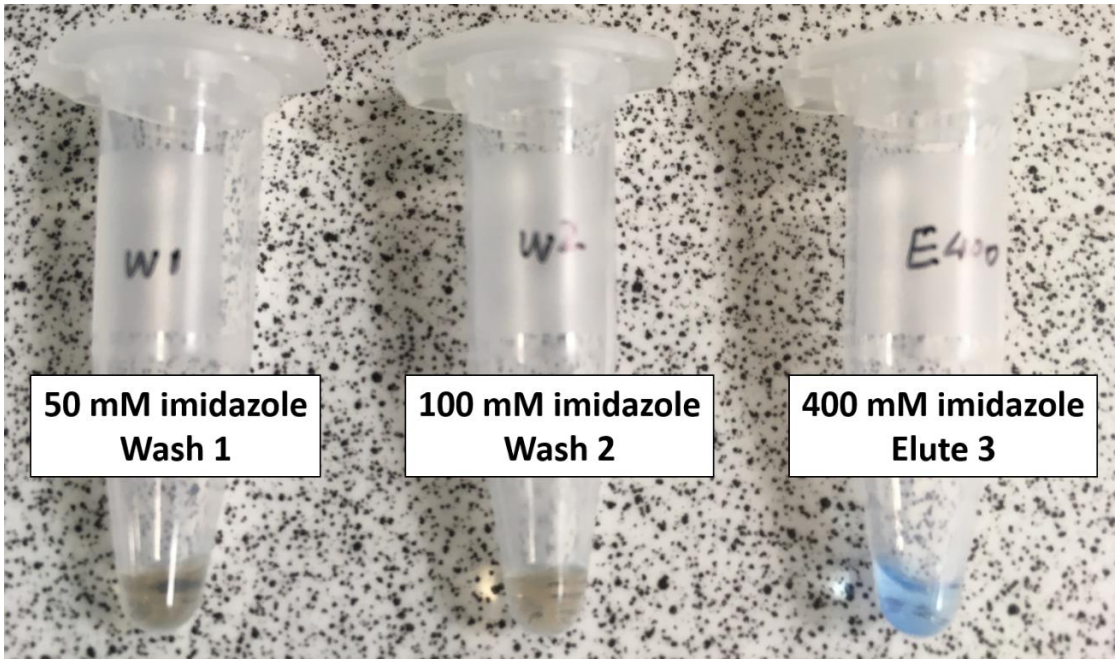


Figure 5.7 Apply BioRad protein assay to determine the existence of PduA nanotubes. From left to right: the first wash buffer with 50 mM imidazole (wash 1), the second wash buffer with 100 mM imidazole (wash 2), the final elute buffer containing 400 mM imidazole (elute 3). The red colour was shown in the first two washes, and blue colour in the elute wash indicating the existence of protein.*

5.3 Sub-THz frequency Measurement on PduA* nanotubes

A transceiver pair of W-band (220-325 GHz) corrugated horns was used for measurements of the complex transmission coefficients S_{21} of the samples located within a quasi-optical (QO) circuit driven by a vector network analyser (VNA). The setup and configuration of this system are described in detail in chapter 2. The S_{21} transmission amplitude was used as a measure of absorbance of the samples.

Theoretical Debye Model and ABCD matrix for 3-layer sample measurement

Water molecules have dipole moments, and the electromagnetic response of a dipole is best described with a Debye relaxation model [18]. Liquid water is a weak hydrogen-

bonded network of molecules, which is easily energised by electromagnetic fields. Previous experiments on water established two Debye relaxation modes in the THz frequency domain [14], [19]. As the vibrational excitations of water exist in the THz frequency domain, we would, therefore, expect that the parameters of the double Debye model will change as the dynamics are altered by the surrounding hydration of protein molecules in solution.

The structure of a 3-layer system (TPX plate-sample-TPX plate) can be represented in a corresponding 3-term product ABCD matrix. The transmittance matrix of the assembly is given by:

$$A(\varepsilon, d) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_{TPX} & B_{TPX} \\ C_{TPX} & D_{TPX} \end{bmatrix} \begin{bmatrix} A_S & B_S \\ C_S & D_S \end{bmatrix} \begin{bmatrix} A_{TPX} & B_{TPX} \\ C_{TPX} & D_{TPX} \end{bmatrix} \quad (5.1)$$

d is the material thicknesses; ε, is the materials' dielectric properties; the subscript 'TPX' stands for the front and back TPX plates of the sample cell and 's' for the sample.

By applying wave propagation equations of A, B, C, and D for each layer according to the three layers method from [19], the parameters of transmittance (S_{21}) can be obtained:

$$\mathrm{T} = \left[\frac{2}{A + B + C + D}\right] \quad (5.2)$$

Based on the Debye model, the theoretical permittivity characterising the dielectric response of a sample is obtained from the transmission coefficient. Both the real and imaginary parts of the permittivity are plotted in Figure 5.8.

It is observed that the dielectric permittivity decreases from 220-325 GHz. This trend is similar to that of water (dashed lines) for both real and imaginary parts. Further, the trend agrees with the previously published dielectric constants for pure water for the frequency band 0.1-1 THz, as shown in Figure 5.9 [150]. It is therefore assumed that the dielectric response shown is due to relaxation mechanisms, i.e. dipole polarization of the protein-bounded water rather than bulk water. As both the real and imaginary parts of the permittivity are experiencing an increasing change with increasing concentrations of PduA* nanotubes, it is assumed that by adding more PduA* nanotubes to the water, the intra-PduA* motions are detectable by significant differences in permittivity, especially in the in-phase real part of the permittivity.

Figure 5.10 presents the measured amplitude and phase, coupled with the theoretical transmittance data. The amplitude spectra, which related to the absorption of the sample, is noted to decrease with increasing changes in concentration. This is due to the overall sample solution volume being occupied by the solvation shell surrounding the nanotube formed PduA* proteins, bulk water and non-bulk-like buffer. Adding more proteins therefore results in replacing strongly absorbing solvation shell water by more weakly absorbing protein.

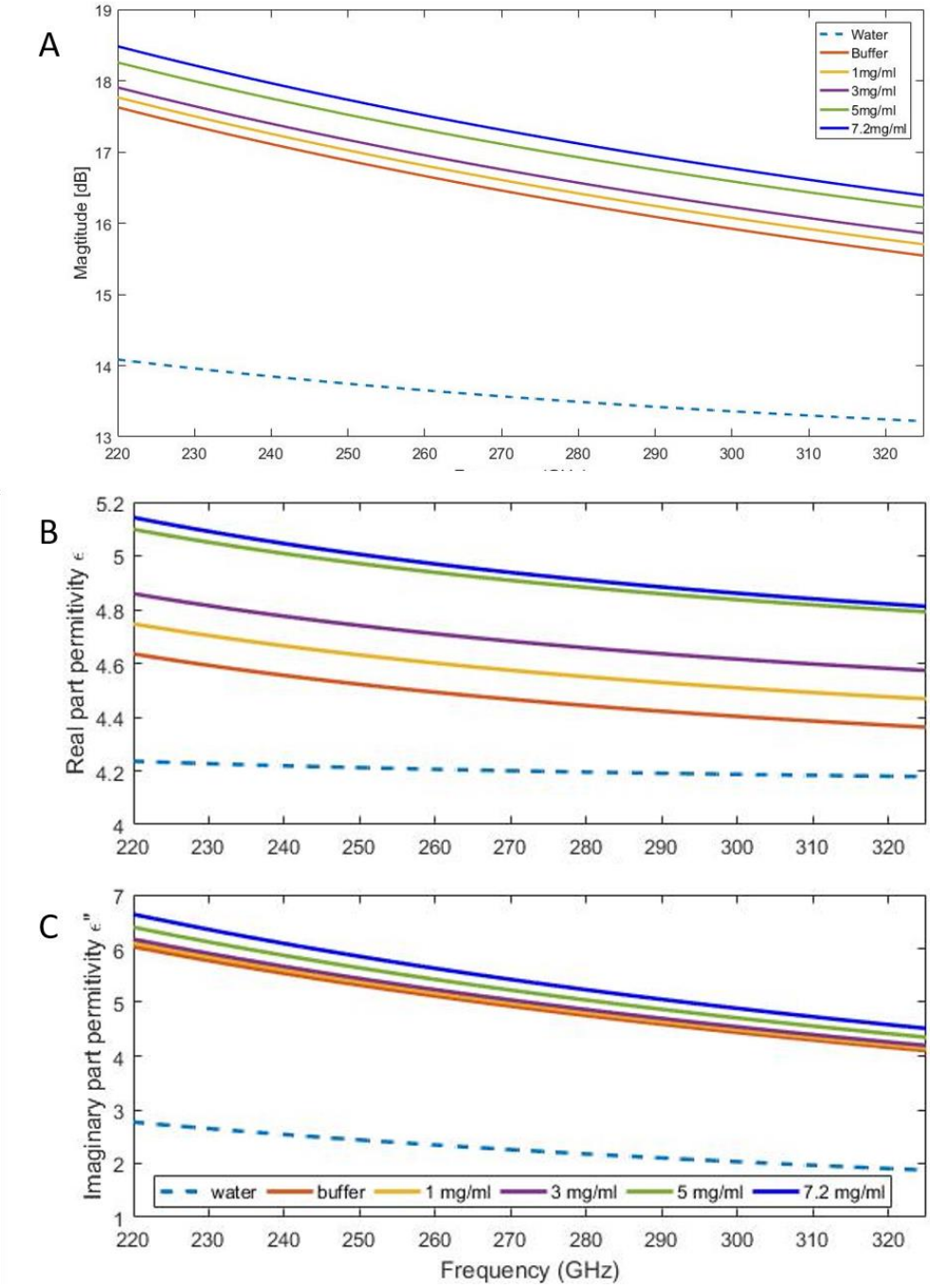


Figure 5.8 Permittivity plots of PduA nanotubes at four concentrations (1 mg/ml, 3 mg/ml, 5 mg/ml, 7.2 mg/ml) with respect to their reference buffer and water. (A)Magnitude; (B) Real part permittivity; (C) Imaginary part permittivity. By adding more PduA* nanotubes to the water, the intra-PduA* motions are detectable by significant differences in permittivity, especially in the in-phase real part of the permittivity.*

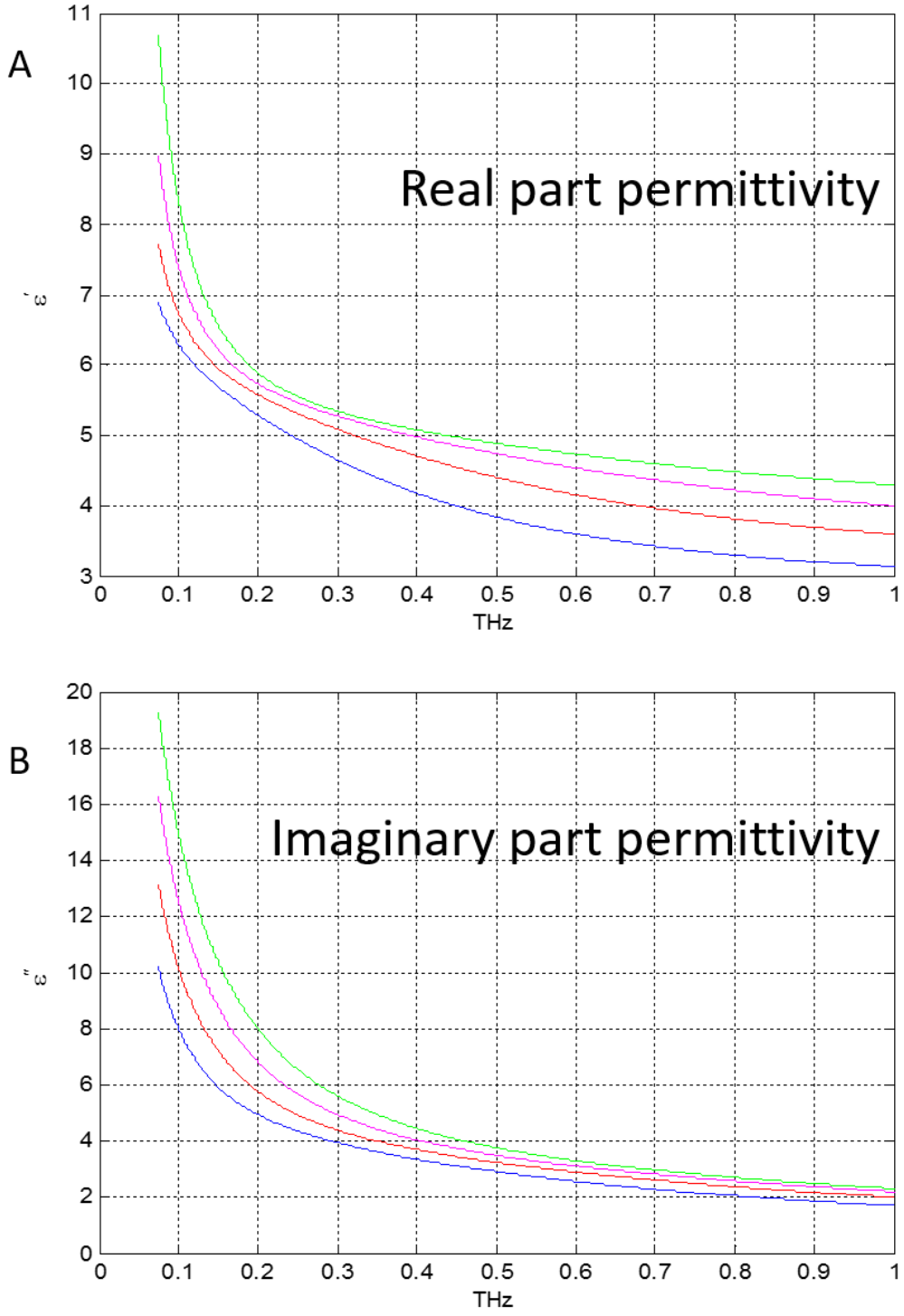


Figure 5.9 Re-plot of the temperature-dependant dielectric constants of pure water at frequency range 0.1-1 THz. The source is adopted from [150]. *A is for the real part of the permittivity; B is for the imaginary part. The blue trace is at 0°C, red at 10 °C, cyan at 20 °C, and green is at 30 °C.*

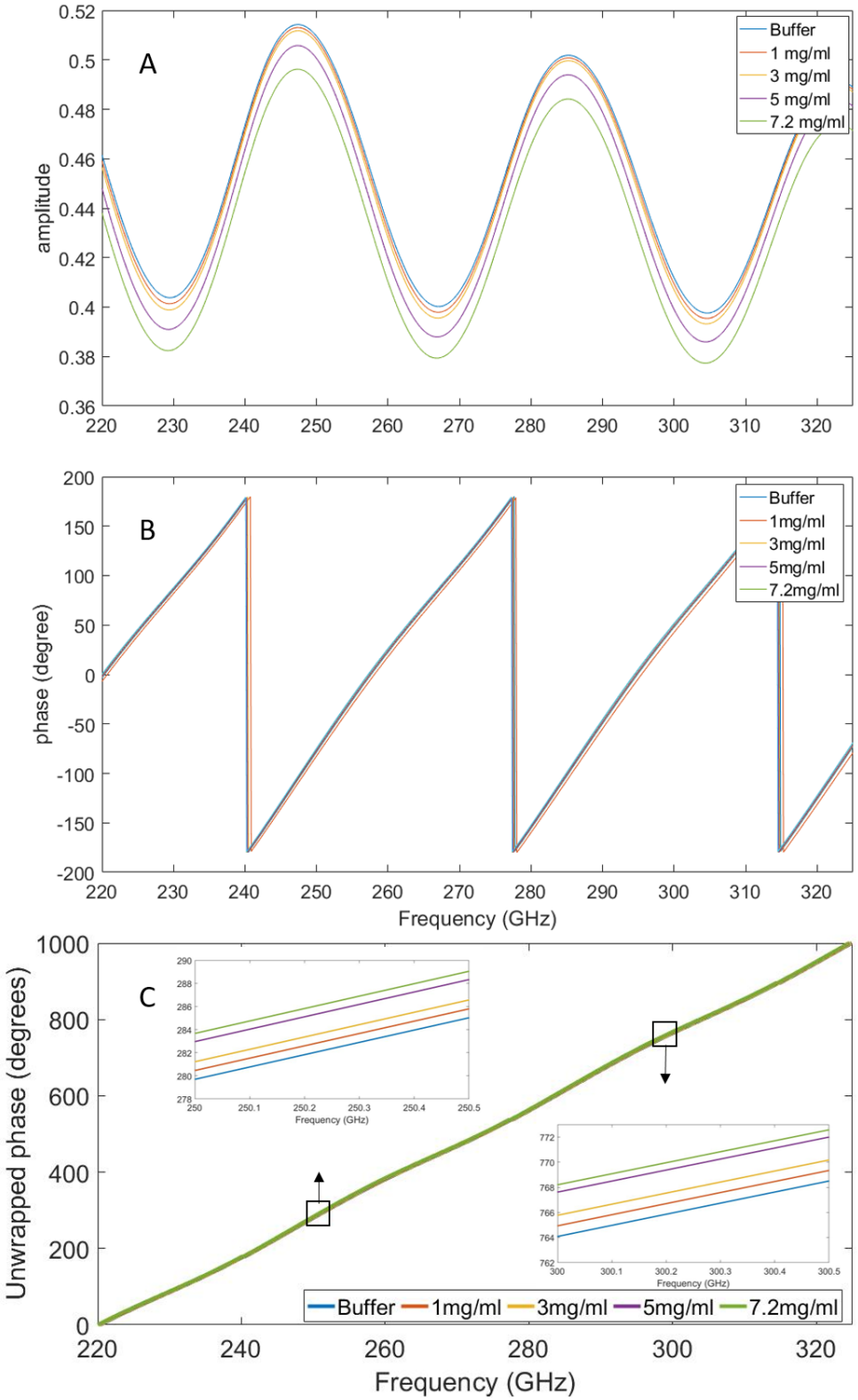


Figure 5.10 *Transmittance spectra of four concentrations of PduA nanotubes and the reference buffer at 220-325 GHz. (A) Amplitudes; (B) phases; (C) Unwrapped phase, with zoom in spectra in frequency range 250-250.5 GHz and 300-300.5 GHz displayed in two inset figures.*

5.4 Summary

In this chapter, *in vitro* production of nanotubes assembled from purified PduA* hexamer has been described. This is the first time PduA* nanotubes have been assembled *in vitro.* Previous reports are of *in vivo* assembly and purification from bacterial cells. Formation of these nanotubes requires the removal of salt and the concentration of the protein. The *in vitro* assembly reveals that this is a self-assembly process, independent of any cellular processes or molecules.

The current model of these nanotubes is one where the hexamers tile the surface of a cylinder. The question is, which side is of the hexamer external and which internal? Initial studies suggested the His-tag is external, consistent with the original model of the nanotubes proposed in Figure 5.11 by Pang *et al*. [24].

After successfully producing the above nanotubes, sub-THz measurements were carried out to check if coherent THz radiation can probe differences between different concentrations of PduA* nanostructures. The PduA* samples with different concentrations have been measured in the sub-terahertz spectral domain of 220-325 GHz. The measurements determined that their permittivity can sensitively detect the concentration changes. Furthermore, transmittance amplitude data, indicates that it is possible to use the dielectric response as a measure to characterize intra-molecule motion of the nanotube.

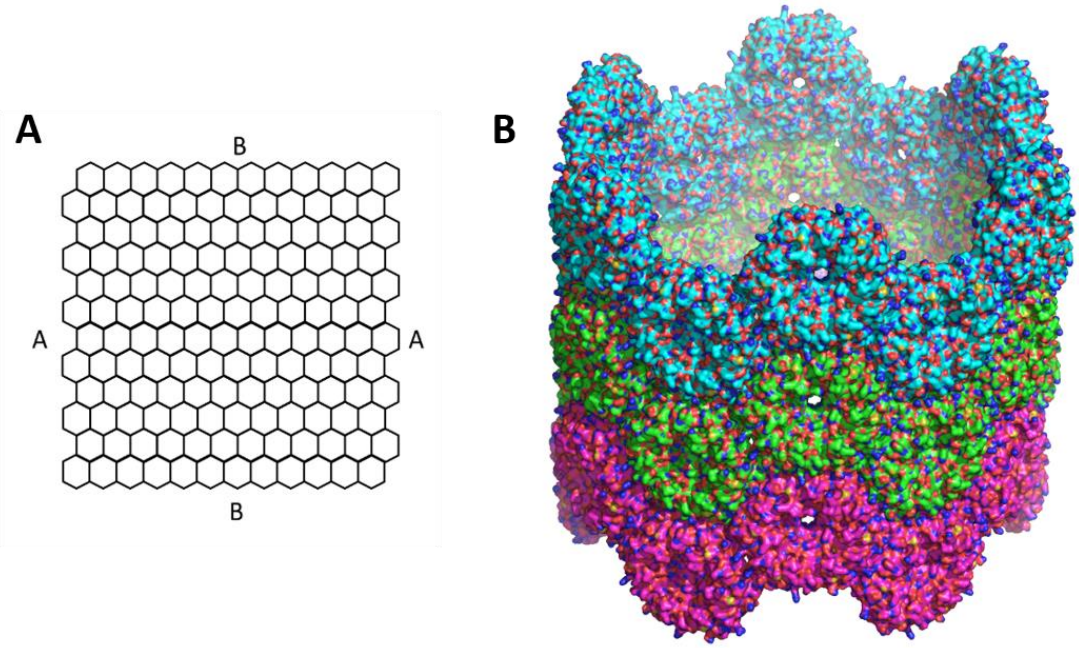


Figure 5.11 A model of the native PduA derived nanotubes seen for native (and V51A mutant) PduA. [adapted from [24]*]. (A) A hexagonal sheet can be rolled into a tube in one of two ways, either 'A' can be connected to 'A', or 'B' to 'B'. (B) In this model of diameter 20 nm, the sheet is rolled so that A and A are brought together and there are 12 hexamers per ring of the protein nanotube. The histidine tag in each hexamer is external.*

Chapter 6 **Quasi-optical Terahertz Circular Dichroism spectroscopy: system and methods**

6.1 Introduction

According to Zhang *et al.*, [151], frameworks without a mirror symmetry plane, having a ubiquitous property and covering a range including crystals, polymers and molecules, are referred to as chirality. Life's building blocks, for example DNA, and almost all biomolecules are chiral, and for this reason, this is of major significance in the fields of medical and biological science. Woody [152] describes circular dichroism as the result of chiral molecules and polarised light interacting with each other. It is stated that almost every biological molecule is chiral [153]. As many as 19 out of the 20 familiar amino acids which generate proteins, are, for example, chiral themselves; and so are a large number of other biologically significant molecules, and also the higher protein structures, DNA and RNA [154]. The chemistry of biological molecules, which is immensely chiral, is particularly suitable for examination using circular dichroism, and the principal application of this method is research into biological molecules. Contained in the understanding of the greater order structures of chiral macromolecules, for example DNA and proteins, is a substantial subset regarding the utilisation of circular dichroism within biochemistry [153]. This is because the CD spectrum of a DNA molecule or of a protein does not represent the total of each separate base or residue in the CD spectra, but it is, however, considerably affected by the macromolecule structure itself, which is three-dimensional. The particular circular dichroism signature, which is possessed by every structure, is able to be utilised as a

means of discovering structural elements, and also as a method of following the transitions within the chiral macromolecular structure [153].

The ellipticity shown by the optically active peptide bonds within proteins is subject to change on the basis of the molecule's local conformation. It is possible to examine the secondary constructions of proteins by utilising far-UV (190-250 nm) area of light. Random coil conformations together with ordered α-helices, β-sheets, β-turn, possess distinguishing spectra. The foundation for the analysis of the secondary structure of proteins is formed by these unique spectra concerning CD, the fact that it is possible to establish only the relative fractions of residues within each conformation, and is not possible to establish the precise location of each structural feature within the molecule is a matter that ought to be noted. It is essential to change the information to a standardised value which is not dependent on molecular length, when CD information for sizeable biomolecules is delineated. This is accomplished by dividing the molar ellipticity by the number of monomer units or the number of residues within the molecule [152].

The CD can also be used to address conformational transitions within molecules. It is possible to utilise CD to establish the level of similarity between a mutant protein and a wild-type protein or to demonstrate the range of denaturation with a changed chemical environment or a temperature change. Additionally, data concerning changes in structure on ligand binding can be supplied by CD. It is essential to establish the native conformation spectrum if any of this data is to be understood [152], [155].

It is possible to utilise near-UV spectroscopy in order to ascertain certain data concerning the proteins' tertiary structure. The cause of absorptions which are in the range of 200 nm to 300 nm is the orientation of the dipole and also the bordering environment of the tryptophan, aromatic amino acids, tyrosine, phenylalanine, and also the residues of cystine which has the ability to generate disulphide bonds. Also, structural data concerning prosthetic groups within proteins can be supplied by the utilisation of near-UV methods [153].

Using visible CD spectroscopy, proteins which contain metal can be studied. The d-d changes to metals within chiral environments is excited by visible CD light. CD light cannot be absorbed by free ions in a solution. Therefore it is possible to establish the pH dependence on the stoichiometry and also on the metal binding [156].

IR light is utilised by vibrational CD (VCD) as a means of establishing 3D frameworks of carbohydrates, nucleic acids, and short peptides. Use has been made of VCD to indicate the number and shape of in A-, B-, and Z-DNA. VCD, which promises to be an extremely strong tool, remains a comparatively modern method. It is necessary to use substantial ab initio computations and also large concentrations in order to solve the spectra. These need to be conducted within water, but the molecule is likely to be forced into a conformation which is not native [157].

There is a good understanding of circular dichroism which is well established within the frequency scope of violet, visible, near- and mid-infrared and is an indispensable part of present-day biophysics involving many exceptionally good turnkey tools which are obtainable on the commercial market. Nevertheless, there have been few documented measurements of terahertz circular dichroism.

In additionally, the results of theoretical research show consistency with the major significance of the collective dynamics of the inter-domain within the biopolymer framework and function. The research additionally implies that many collective modes of THz are present [46], [58], [158–160]. As charged macromolecules, the absorption of proteins should be coupled with circular dichroism feature since the collective vibrational modes of a folded, chiral biomolecule would fundamentally behave differential absorption between left and right handed circularly polarized radiation in THz region. It would be expected that an estimation of the THz CD's size and the connection between the structure and the signatures could be attained by the biomaterials' mechanical properties and the distribution of charge.

The combination of a VNA and QO bench was an approach to studying the aqueous proteins with linearly polarised absorbance, which has been introduced in the previous chapters. In this chapter, we used it for the circularly polarized measurements because the special capability of the VNA to make precise and stable phase measurements allows us to transform directly measured complex co-polar and cross-polar transmittances respectively, into complex circular-polarization transmittances. Therefore, a sub-Terahertz circular dichroism spectrometer is developed and built based our VNA+QO system at 220-325 GHz sub-THz range (VNA-CD), to explore the potential significant absorption feature between VNA-CD and our aqueous protein samples.

6.2 Materials and methods

6.2.1 Materials

For examining the relationship between the biopolymers' (protein) structural change and THz-CD absorbance, four proteins are selected as a simple physical representation that captures the essential features of biological macromolecules: BSA, myoglobin, Apomyogobin (ApoMb) and Cytochrome C (CyC). They were initially prepared in 20 mM phosphate buffer solution. Myoglobin and ApoMb are globins and similar in helical structure. Cytochrome C is a four-helix bundle topology. These proteins are all rich in alpha helical structure so their similarity or difference in THz-CD spectral change could relate to the detection of a secondary structure alpha helix.

All chemicals, BSA, myoglobin and Cytochrome C are purchased from *Sigma* Aldrich in the form of lyophilized powder and were purified with gel chromatography (HiLoad 26/60 *Superdex75*, GE Healthcare) using FPLC (Fast Protein Liquid Chromatography); ApoMb was obtained from myoglobin with the removal of the heme group using the 2-butanone extraction method, as mentioned in chapter 2 and chapter 4. The same Bruker liquid cell with two TPX plates and a 100 μm (PTFE) spacer was used as the liquid sample holder.

6.2.2 Spectrometer configuration for sub-Terahertz VNA-CD

The VNA [161] in use has a 10 MHz to 43.5 GHz operating frequency domain; special millimetre-wave frequency extension heads cover the waveguide bands of 220-325 GHz. Rectangular waveguide outputs of the extension-heads are terminated with high-gain corrugated horns. As shown in Figure 6.1, with an ellipsoidal mirror (F1) to direct the radiation from the transmitting horn, the ellipsoidal reflector (F2) receives the

signal beam diverging from the VNA head-unit (H1) and transforms it into a linearly polarised quasi-collimated beam which passes through the kinematically-located sample (S). The transmit beam is propagated to a similar ellipsoidal reflector (F2) and a further ellipsoidal mirror (F1), efficiently couples the beam into the receiving horn of the VNA head-unit (H2).

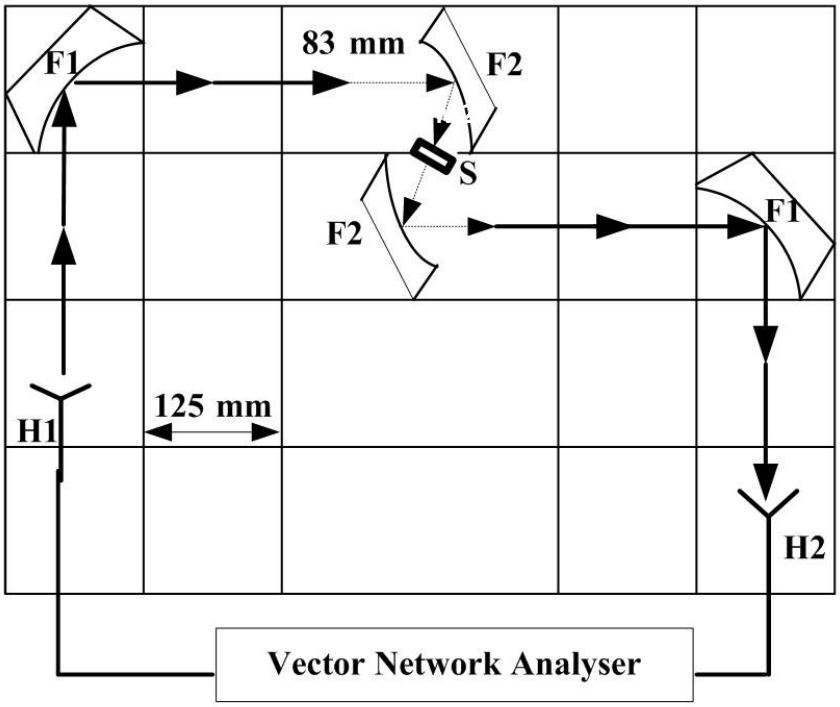


Figure 6.1 A schematic diagram of the quasi-optical transmissometer driven by a Vector network analyser. H1 and H2 denote the corrugated horns connecting with waveguide bands; F1 and F2 are a confocal pair of mirrors used to couple the THz radiation on the sample (F1 focal length is 250 mm; F2 focal length is 83 mm); S is the sample under test.

In the earlier work [95], the circularly polarised method was developed using the VNA driving a QO circuit to measure a set of hexaferrite ceramics plates in order to characterise the magneto-optical performance. The methodology and system design are proven and stable. This analyser is adapted; to customise the spectrometer to realise THz CD by modulating the incident circular polarizations between left (ALCP) and

right (ARCP), and measuring the differential absorption of these two circular senses of polarization in the conventional sense:

$$\Delta A = \text{ ALCP - ARCP} \tag{6.1}$$

ARCP and ALCP will be denoted as $t_{\pm}(\omega,d)$ ('+' = ALCP and '-', ARCP), relating the transmittance (t) to probe frequency ω and the sample thickness d. The complex circular-polarization transmittances, $t_{+}(\omega,d)$ and $t_{-}(\omega,d)$ therefore can be calculated and transformed by directly measuring the co-polar and cross-polar linear transmittances, using [95]:

$$t_{\pm}(\omega,d) = t_{co}(\omega,d) \mp i \times t_{cr}(\omega,d) \tag{6.2}$$

6.2.2.1 Method-1 VNA-CD configuration

Initially, the VNA-CD configuration was set up for measuring co-polar and cross-polar linear transmittances, as shown in Figure 6.2. The configuration is similar to the case of linear transmittance shown in Figure 6.1, only that the planar horizontal and vertically-aligned wire-grid polarizers (G1 and G2) (arrays of freestanding fine wires to reflect polarisation parallel to the wires and transmit polarisation perpendicular to the wires), are included in the quasi-optical circuit so that circularly polarised light can be generated based on the component linear-polarisation transmittances. The measured co- and cross-polar linear transmittances are superposed to form circularly polarised transmittances.

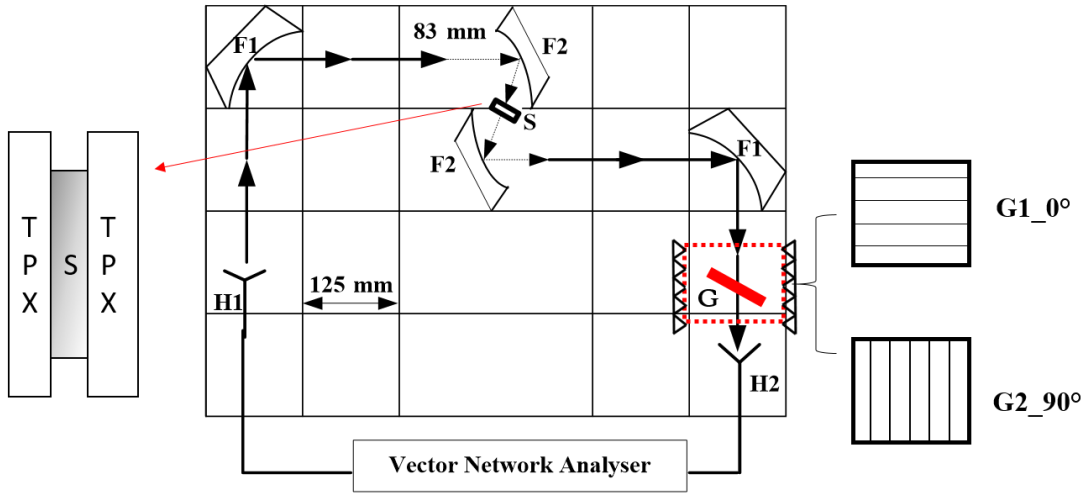


Figure 6.2 Schematic diagram of the Quasi-Optical VNA transmissometer for the method-1 configuration. H denotes a corrugated feed-horn; F, an ellipsoidal reflector; S a sample under test; C a quasi-optical directional coupler and G is a wire-grid polarizer.

The VNA head units (H1 and H2) and wire-grids (G1 and G2) are required to be adjusted accordingly in the measurement procedure in order to get co-polar and cross-polar transmittances separately. The measurement steps are described in Appendix 2 in details.

For material measurements in the millimetre-wave and sub-millimetre wave bands, amplitude and phase variations are very sensitive to mechanical disturbances such as cable movement or lateral displacements (equating to phase-shifts). Based on the previous configuration, the wire-grid at the output arm has to be changed by 90˚ and the receiver horn has to be rotated similarly to make the cross-polar measurements simultaneously. Additionally, as the cross-polar transmittance is comparably weak, and may be heavily affected by the system, the resultant calculation will be dominated by the co-polar transmittance component.

Therefore, we have improved the stability of the system by bringing in the advanced configuration of method-2.

Method-2 VNA-CD configuration

The optimised quasi-optical (QO) measurement configuration is shown in Figure 6.4. Compared to method-1, the method-2 enhances the following:

- A ferrite planar element is incorporated ahead of the sample cell, and an air gap is left between the ferrite piece and the three-layer sample. The application of a magnetically hard, thin, planar ferrite has been reported in the literature to enhance CD by preferentially emphasising the weaker cross-polar component [162]. The inherent optical properties of gyrotropic ferrite materials mean that they support circularly polarized propagating waves rather than linearly polarized waves. Therefore, as presented in Figure 6.3B, the circularly polarised signal was generated prior to passing through the sample.

- By introducing an independent 45° wire grid (shown in Figure 6.5) into the new measurement configuration removes the need to rotate the receive horn for cross-polar transmittance measurements, enabling removal of a systematic error traced to cable movement arising during the measurement circuit re-configuration between co- and cross-polar measurements.

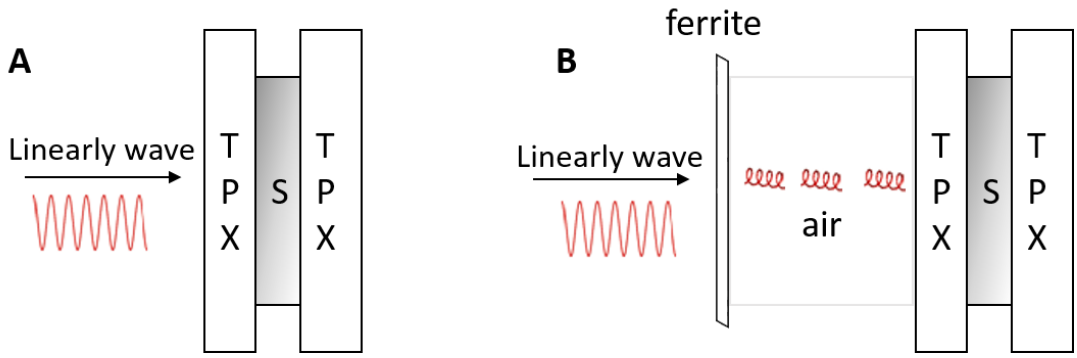


Figure 6.3 Schematic diagram of the sample used in method-1 (A) and method-2 (B). (A) The three-layer sample in method-1 is formed by the sample being sandwiched between two TPX plates. (B) The sandwiched-sample in method-2 is preceded by a ferrite plate. A linearly polarised wave propagating through the ferrite suffers rotation. In post-processing in constructing circular polarising transmittance, the effect of the presence of the ferrite is to enhance the cross-polar signal beam amplitude. The air-gap is used to give enough space for this transform happen before it passes into the three-layer sample.

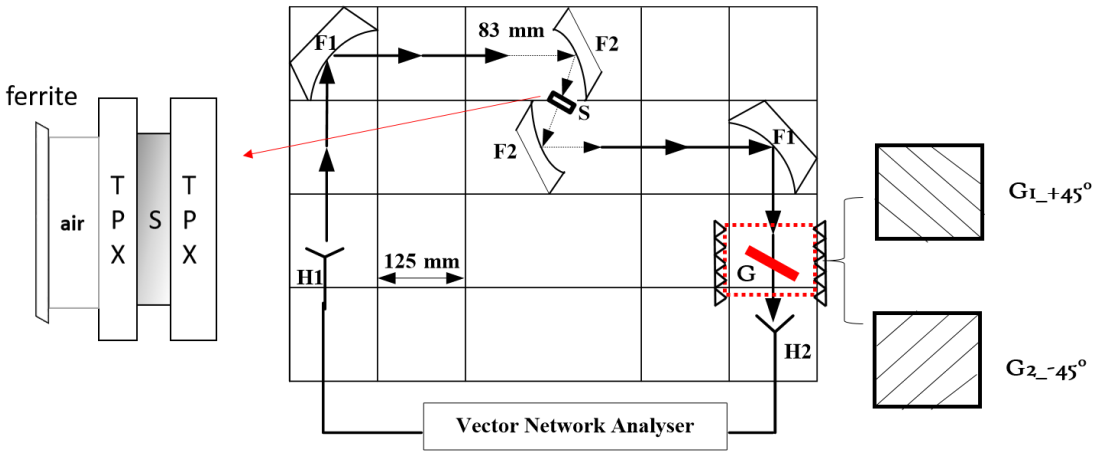


Figure 6.4 Schematic diagram of the QO transmissometer in method-2. H again denotes a corrugated feed-horn; F, an ellipsoidal reflector; S the liquid sample with leading ferrite plate sticking G, a vertical wire-grid polarizer and; G_45°, a 45° wire grid.

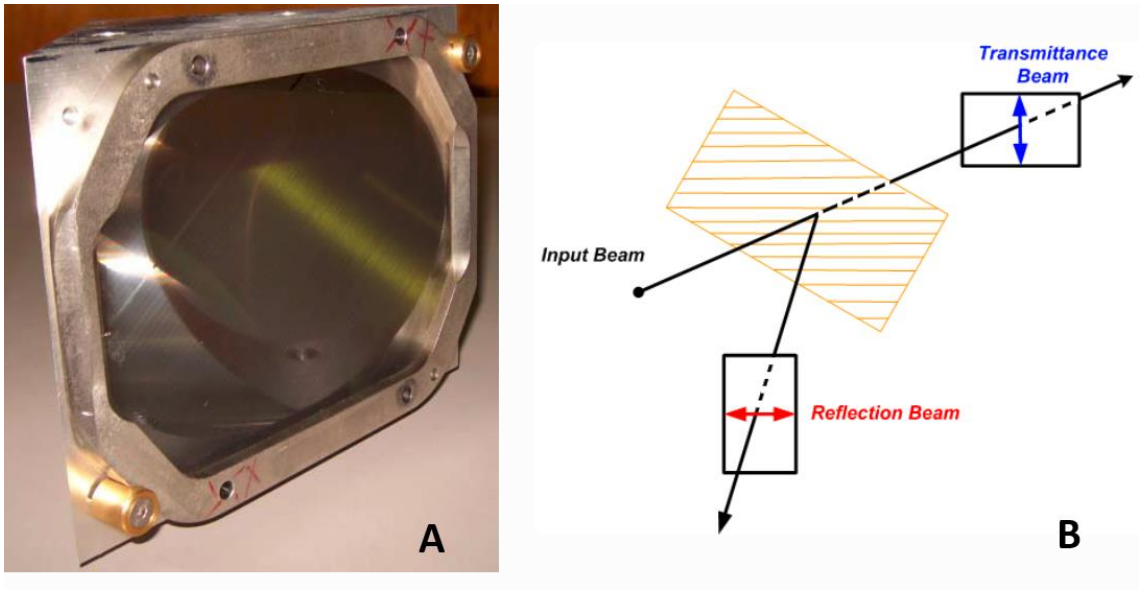


Figure 6.5 The 45° wire grid used in the VNA-CD measurement. (A) Photograph of the 45° wire grid; (B) the working principle of a 45° wire grid; the decomposed component parallel with 45° wire grid is reflected, and the component perpendicular is transmitted.

Instead of the directly measured co- and cross-polar transmittance obtained in method-1, we recorded three measurements for each sample in method-2; T_{CO}, $T_{1,45^\circ}$ and $T_{2,45^\circ}$, which refer to the co-polar transmittance, the 45° '+' transmittance and the 45° '-' transmittance, respectively. The measurement procedures are explained in detailed steps in Appedix 2.

With the known co-polar transmittance (T_{CO}), the cross-polar transmittance (T_{CR}) can be obtained using the measured $T_{1,45^\circ}$ and $T_{2,45^\circ}$ data. θ_1 is the changed polarized angle when the wave propagates through the ferrite. As θ_1 is unknown, however, $\theta 2$ can be determined by the measurements $T_{1,45^\circ}$ and $T_{2,45^\circ}$. The relationship between $\theta 1$ and $\theta 2$ is $\theta_1 = \pi/4 \pm \theta_2$, as shown in Figure 6.6.

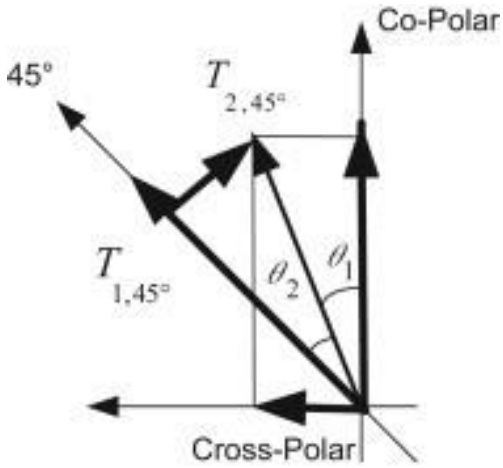


Figure 6.6 Picture showing the principle for calculating cross-polar transmittance by using the measured co-polar transmittance T_CO and two 45° direction transmittances T_(1,45°), T_(2,45°).

The complex circular-polarization transmittances, $t_+(\omega,d)$ and $t_-(\omega,d)$, which are marked as T_P and T_n in the following expression, was then calculated following equation 6.2 as a measure of the spectra from VNA.

Using the above two configurations for a VNA-CD spectrometer, their performance was firstly verified by calculating the difference between two repeated cross-polar transmittance spectra of the same distilled water solution at 220-325 GHz, employing measured data in method-1 and calculated data in method-2 (Figure 6.7). The two repeated measurements used in the plot are individually recorded with an interval of 30 minutes for both method-1 and 2. The liquid sample cell was repositioned, and the system was normalized with the background before each measurement to include any system and sample-handling error. It is clear to see that the differential cross-polar transmission in method-2 is much lower than that of method-1, indicating the uncertainty level of method-2 is significantly improved comparing with method-1. The transformed co- and cross-polar in method-2 is plotted in Figure 6.8. The trend is in

good agreement with the previous results from Yang *et al.*, measuring the transmission of hexaferrite materials only at 75-110 GHz [162].

Additionally, by checking the repeatability of the measurements using method-2, the VNA system error alone is of the order of 10^{-3}. A data set is the average of 30 repeated measurements for the same sample with continuous tracking over a 30-minute period. Recognising that the VNA is a high-spectral-resolution system with high signal-to-noise performance, and the fact that the total recording time for the data used in producing the measurements was within a few minutes, we would expect method-2 to be sensitive to changes in a protein sample.

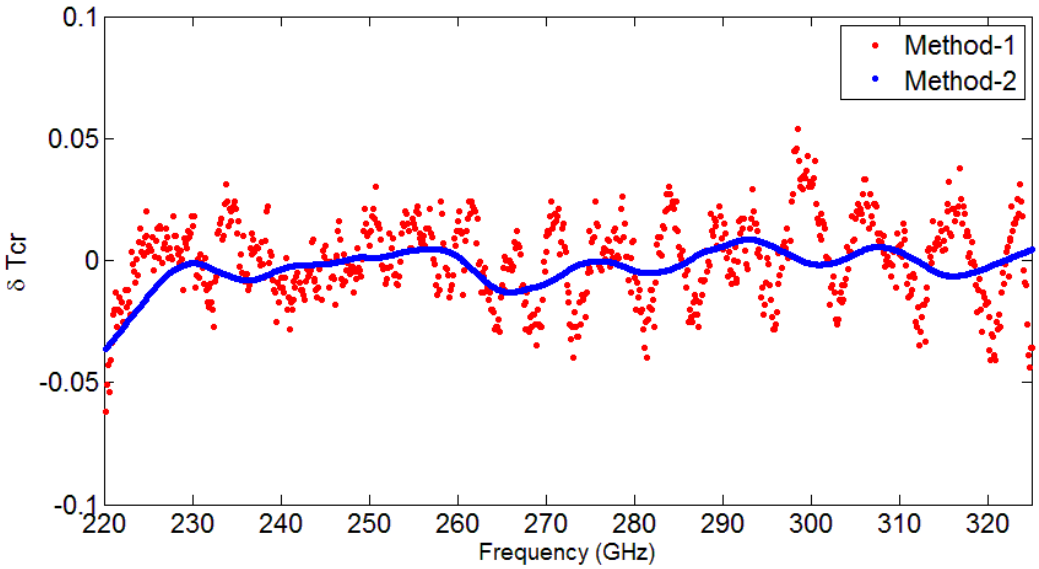


Figure 6.7 The difference between two repeated cross-polar transmittances of water for both method-1 and method-2, as a measure of the uncertainty. The modulus of the amplitude was plotted, red dots for method-1 and blue dots for method-2.

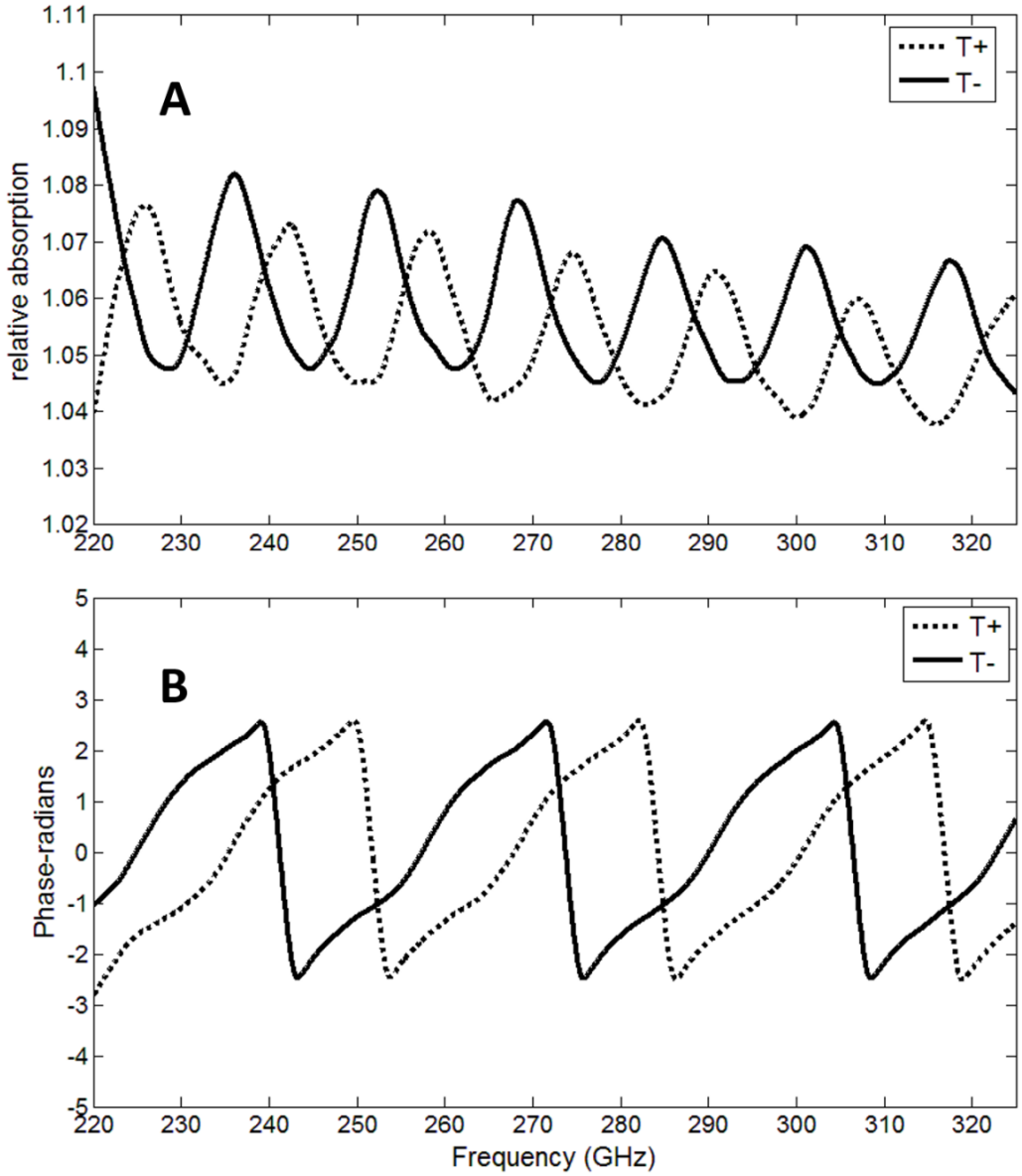


Figure 6.8 Measured complex circular-polarization transmittances of the system background (empty cell with ferrite) from 220 to 325 GHz. (A) The amplitude and (B) the phase. The dash line T_+ refers to the left-handed circular polarization transmittance and solid line T_- for the right-handed (or clockwise sence when looking in the direction of the wave-vector).

6.3 The VNA-CD measurements

Using method-2, the transmission of left/right circularly polarization (T_+, T_-) and their difference (ΔT) were measured with three proteins (myoglobin, apomyoglobin,

Cycrome c) under various conditions. They were prepared in the same buffer, with identical concentrations at 2 mg/ml and 20 mg/ml. The individual proteins were firstly compared in two concentrations, taking BSA as an example in Figure 6.8. The three proteins at the same concentration and buffer conditions are compared in Figure 6.9. For detecting denaturation change in protein by VNA-CD, apomygolobin was treated with acidic environment and GdmCl solution, prepared at pH 2, pH 4 and pH 6, and in GdmCl concentrations at 2, 4 and 6 M, shown in Figures 6.10 and 6.11 respectively. By comparing $T_+, T_-$and ΔT, the protein spectra differences (ΔT) on VNA-CD were hard to distinguish for both changes in protein concentrations and denatured conformation (Figure 6.8-6.11C), and the small variations in either T_+ or T_- could not be clearly attributed to variation in protein conformation, as shown in Figure 6.8-6.11 panels A and B. The reason appears mainly to be because chiral activity in the THz frequency domain is significantly weak compare with the strong conformational rigidity of proteins in the infrared. A channel spectrum is supposed existing in both left/right circularly polarization (T_+, T_-) spectra, due to the fact that the complex circular polarization transmittance is integrated by co-polar and cross-polar linear transmittance and would potentially contain informative spectra and coupled with circular dichroism features in the mixed electric field vectors. However, although the VNA-CD experimental methodology has been carefully designed to suppress systematic error and the CD signal de-biased against the co-polar component by the introduction of the ferrite, the THz spectral response at W-band could still be largely

affected by the strong water absorption loss.

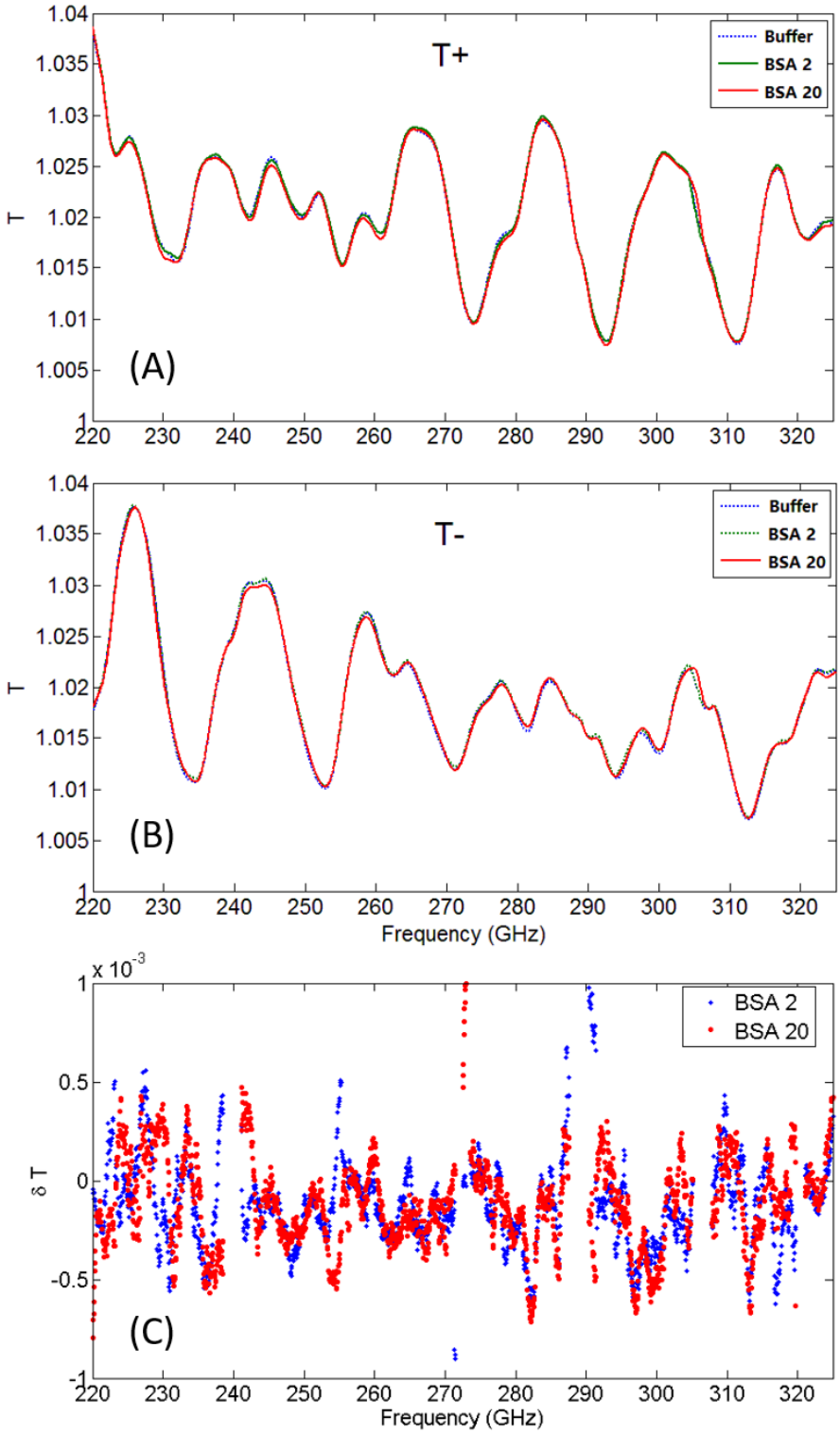


Figure 6.9 The circularly polarized transmittances of BSA protein comparing two concentrations: 2 and 10 mg/ml. (A) Left handed circular polarization; (B) Right handed circular polarization; (C) The differential between left and right, after subtracting the relevant buffer.

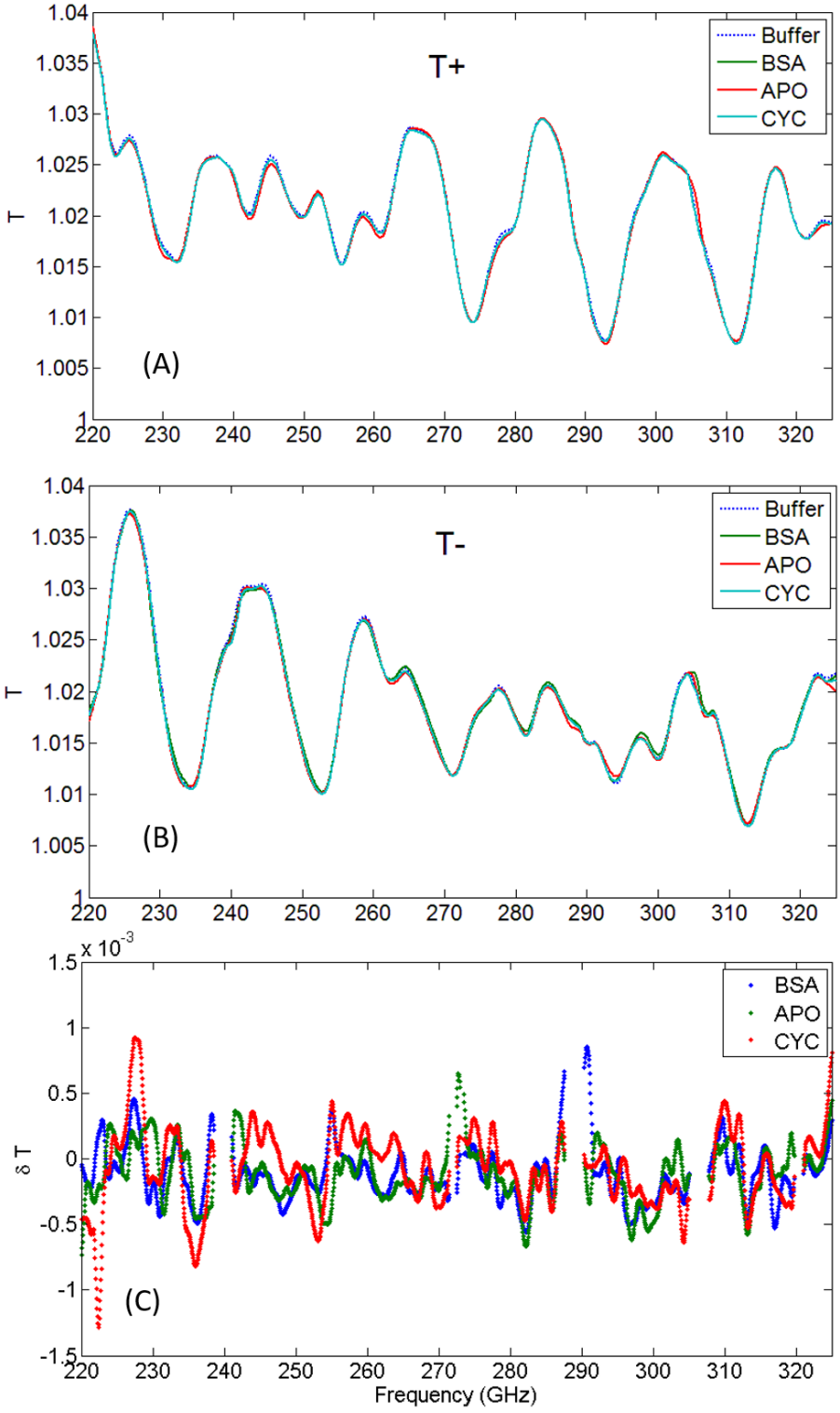


Figure 6.10 The circularly polarized transmittances are comparing BSA, ApoMb and Cytochrome c proteins at the same concentration (2 mg/ml) and buffer conditions (20 mM phosphate pH 6.5). (A) Left handed circular polarization; (B) Right handed circular polarization; (C) the differential between left and right, after subtracting the relevant buffer.

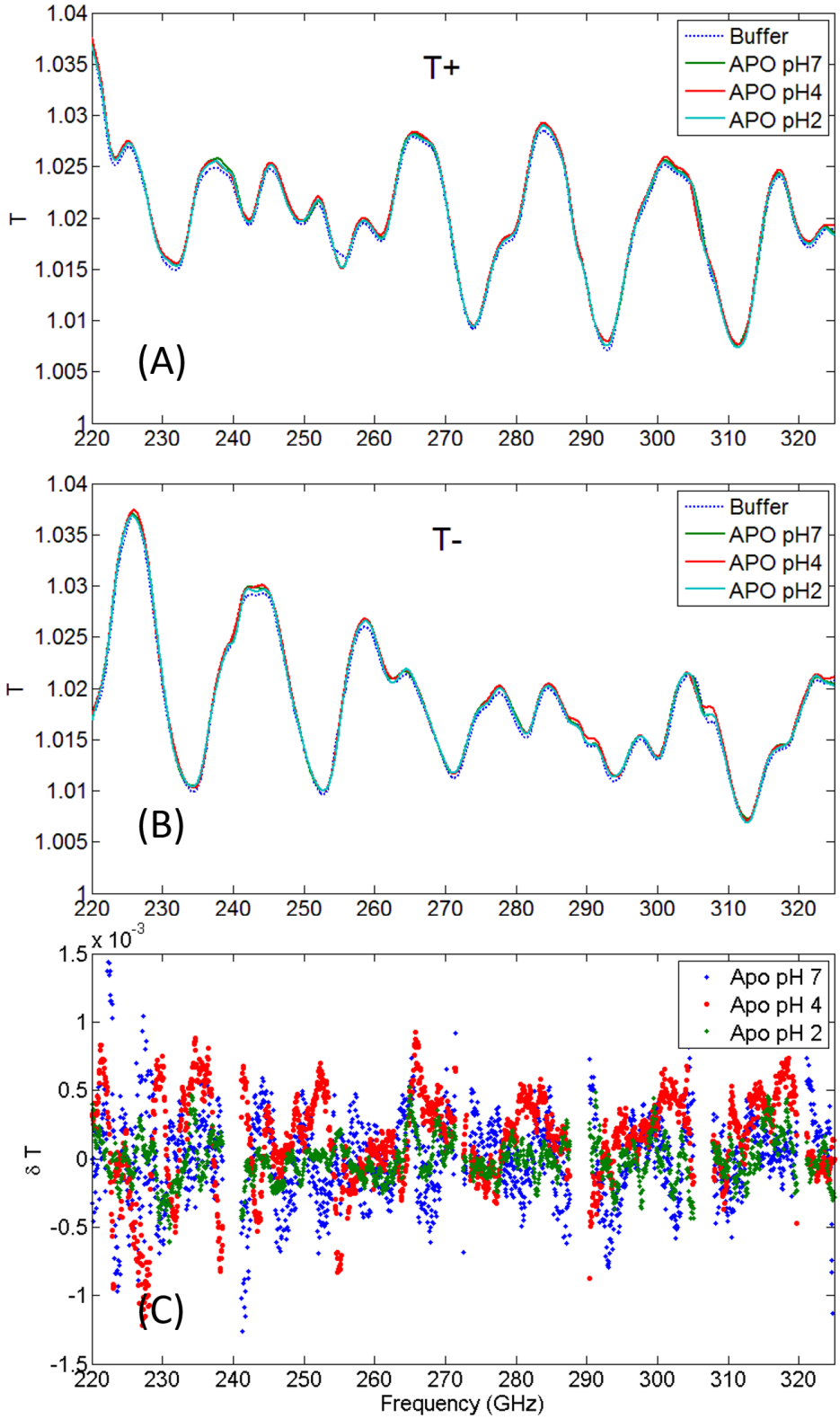


Figure 6.11 The circular polarized transmittances comparing ApoMb at 20 mg/ml under different pH conditions (pH 2, pH 4, pH 7). (A) Left handed circular polarization; (B) Right handed circular polarization; (C) the differential between left and right, after subtracting the relevant buffer.

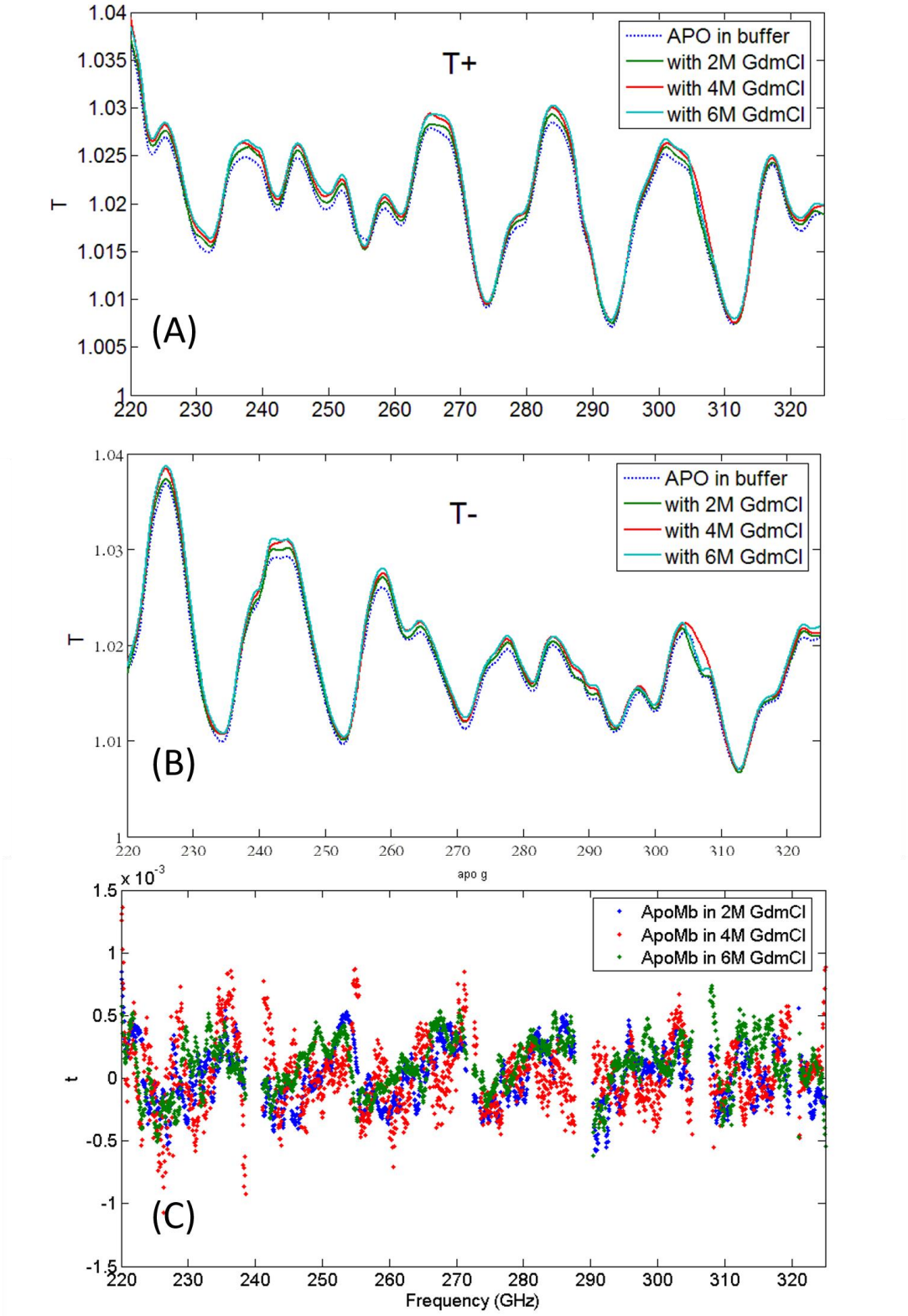


Figure 6.12 The circularly polarized transmittances are comparing ApoMb at 20 mg/ml by treated with different GdmCl conditions (2 M, 4 M and 6 M). (A) Left handed circular polarization; (B) Right handed circular polarization; (C) the differential between left and right, after subtracting the relevant buffer.

6.4 Summary

Terahertz circular dichroism provides a potential means for detecting group vibrational action of biomolecules that are theorised to be responsible for protein function. The analysis suggest that water interactions in the hydration layer surface may be mutinginter-molecular action, since the strength of the signals is relatively weak and VNA-CD was limited to just one particular waveguide band. Further work will probe for VNA-CD activity from 0.325-1.1 THz via VNA+QOs.

Table 6.3 provides a list of errors in repeatability for CD VNA studies under 10 specified frequencies from the 220 to 325 GHz range. Each error in repeatability was averaged from 5 repeated measurements. By comparing the theoretical limits and practical data, it suggests that at the current level of uncertainty in experimental results, there is no discernible dependence in differential transmission on long-range energetics in protein conformational variation. This may indicate that the current instrument is either not sensitive enough or that activity lies in other THz sub-bands.

Frequency (GHz)	**Error from VNA System**		**Error from sample**	
	T_{co} (10^{-3})	T_{cr} (10^{-3})	T_{co} (10^{-3})	T_{cr} (10^{-3})
230	4.31	1.89	1.51	3.95
250	4.41	0.81	2.75	0.64
270	1.11	2.48	4.27	2.76
390	2.50	5.77	7.44	1.56
310	4.66	9.33	5.73	10.93

Table 6-1 A list of errors in repeatability for CD VNA studies under 5 specified frequencies (230, 250, 270, 290 and 310 GHz) from the 220 to 325 GHz range. The uncertainty for CD VNA system are mainly from systematic error (which is due to environment change in background, cable movement and re-position of the sample) and signal loss from hydrated sample. Errors were averaged from five repeated CD VNA measurement for both co-polar (T_{co}) and cross-polar (T_{cr}) transmittances. The theoretical value of error should not excess 1×10^{-3}.

However, the instrument affords coherent sensing of complex amplitude with high spectral resolution and repeatability. Incorporation of a ferrite element has served to enhance the sensitivity of the methodology further. In the light of these preliminary VNA-CD results, the newly trialled VNA-CD instrument is reasonably anticipated to yield unambiguous analysis with a stronger source and better control of bulk water effects.

Publications and conferences contributions

CONFERENCE PAPERS:

J. Qiu, B. Yang, O. Sushko, R. W. Pikersgill, and R. S. Donnan, “Comparing Terahertz transmission response on pH-dependent apomyoglobin proteins dynamics with circular dichroism,” in RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio), 2014 IEEE MTT-S International Microwave Workshop Series on, 2014, pp. 1–3.

O. Sushko, J. Qiu, R. Dubrovka, R. W. Pickersgill, and R. S. Donnan, “Sub-terahertz spectroscopy as a probe for protein stability in an ionic environment,” in Infrared, Millimeter, and Terahertz waves (IRMMW-THz), 2015 40th International Conference on, 2015, pp. 1–2.

J. Qiu, Y. Zeng, B. Yang, and R. S. Donnan, “Exploring Cell-Scaffold Interactions By THz Frequency Domain Spectroscopy,” in Infrared, Millimeter, and Terahertz waves (IRMMW-THz), 2016 41th International Conference on, 2016, pp. 1–3. (submitted)

OTHER CONFERENCE CONTRIBUTIONS:

J. Qiu, O. Sushko, R. S. Donnan, R. Pickersgill, “A sub-terahertz spectroscopic study of Guanidinium Chloride in protein denaturation process” Nano-Spectroscopy & Bio-Imaging, 2015, poster presentation.

J. Qiu, O. Sushko, R. S. Donnan, R. Pickersgill, "Exploring the potential for detecting conformational changes in protein structure and organisation using terahertz spectroscopy" Application of Terahertz Spectroscopy to Protein Chemistry Workshop, 2014, oral presentation

J. Qiu, O. Sushko, R. S. Donnan, R. Pickersgill, "Exploring changes in protein conformation using THz spectroscopy" Queen Mary EECS research showcase 2014, poster presentation (second prize).

R. S. Donnan, B. Yang, A. McIntosh, O. Sushko, J. Qiu, "Towards Developing a Dielectric Metrology for the Life Sciences" PIERS 2013, Stockholm, Sweden – oral presentation.

O. Sushko, R. Dubrovka, R. Pickersgill, J. Qiu, R. S. Donnan, "Potential of THz Radiation for Examination of Dissolved Proteins" Today's RF Tomorrow's Medicine, London, UK, 2013, poster presentation.

Appendix 1

The PduA and PduA* DNA sequence (Panel A) and protein sequences (Panel B) from *Citrobacter freundii* using ClustalW2 [163]:

A

```
PduA            ATGCAACAAGAAGCGTTAGGAATGGTAGAAACCAAAGGCTTGACTGCAGCCATAGAGGCC 60
PduA*           ATGCAACAAGAAGCGTTAGGAATGGTAGAAACCAAAGGCTTGACTGCAGCCATAGAGGCC 60
                ************************************************************

PduA            GCAGATGCAATGGTGAAGTCAGCCAATGTAATGCTGGTCGGCTACGAAAAAATTGGTTCG 120
PduA*           GCAGATGCAATGGTGAAGTCAGCCAATGTAATGCTGGTCGGCTACGAAAAAATTGGTTCG 120
                ************************************************************

PduA            GGGCTGGTAACAGTCATTGTCCGCGGCGATGTTGGCGCAGTCAAAGCAGCAACAGATGCA 180
PduA*           GGGCTGGTAACAGTCATTGTCCGCGGCGATGTTGGCGCAGTCAAAGCAGCAACAGATGCA 180
                ************************************************************

PduA            GGTGCCGCCGCAGCACGTAATGTGGGAGAAGTGAAAGCCGTACACGTCATCCCACGCCCT 240
PduA*           GGTGCCGCCGCAGCACGTAATGTGGGAGAAGTGAAAGCCGTACACGTCATCCCACGCCCT 240
                ************************************************************

PduA            CACACCGATGTAGAAAAAATCTTACCGAAGGGAATTAG-CTAA----------------- 282
PduA*           CACACCGATGTAGAAAAAATCTTACCGAAGGGAATTAGACTAGTAAAGGATCCGGCTGCT 300
                ************************************ ***

PduA            ---------------------------------------------------
PduA*           AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA 351
```

B

```
PduA              1 MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGD     50
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
PduA*             1 MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGD     50

PduA             51 VGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILPKGIS-------     93
                    |||||||||||||||||||||||||||||||||||||||||||||.
PduA*            51 VGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILPKGIRLVKDPAA    100

PduA             94 ----------------     93

PduA*           101 NKARKEAELAAATAEQ    116
```

Appendix 2

The measurement procedures for the VNA-CD method-1:

(1) For co-polar transmittances, the VNA head-units H1 and H2 are both set to transmit/receive vertical polarization, and the wire-grids G1 are set to vertical. The component of an incident beam having vertical polarization is passed, and any orthogonally polarised component is reflected and dumped (onto Radar Absorbing Material, constituting a 'power-dump').

(2) For measurement of cross-polar transmittances, the setting of wire-grid G2 is changed to reflect horizontal polarization and the VNA head-unit H2 is rotated 90° to receive horizontal polarization.

The measurement procedures for the VNA-CD method-2:

(1) Set up the co-polar measurement configuration as shown in Figure 6.5 without positioning the sample S and without the 45°wire grid in the position G. The spectrum thus acquired serves as the background or reference measurement. Locate and align the sample with ferrite in position S and make the co-polar measurement T_{CO}.

(2) Put the G1 wire grid in position G and re-set for a new background. Keep the 45°wire grid G1 in position G, with the sample with ferrite into position S and record the measured transmittance $T_{1,45^\circ}$.

Put the G2 wire grid in position G (rotate the 45°wire grid 180° at the same position) and re-set the new background. Keep the 45°wire grid G2 in position G, with the sample with ferrite into position S and record the measured transmittance $T_{2,45^\circ}$.

References

[1] B. Ferguson and X.-C. Zhang, "Materials for terahertz science and technology," *Nature materials*, vol. 1, no. 1, pp. 26–33, 2002.

[2] M. Tonouchi, "Cutting-edge terahertz technology," *Nat Photon*, vol. 1, no. 2, pp. 97–105, February 2007.

[3] J. El Haddad, B. Bousquet, L. Canioni, and P. Mounaix, "Review in terahertz spectral analysis," *TrAC Trends in Analytical Chemistry*, vol. 44, pp. 98–105, March 2013.

[4] P. H. Siegel, "Terahertz Technology in Biology and Medicine," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 10, pp. 2438–2447, October 2004.

[5] J. B. Baxter and G. W. Guglietta, "Terahertz Spectroscopy," *Analytical Chemistry*, vol. 83, no. 12, pp. 4342–4368, June 2011.

[6] C. Wang, J. Gong, Q. Xing, Y. Li, F. Liu, X. Zhao, L. Chai, C. Wang, and A. M. Zheltikov, "Application of Terahertz Time-Domain Spectroscopy in Intracellular Metabolite Detection," *Journal of Biophotonics*, vol. 3, no. 10–11, pp. 641–645, October 2010.

[7] E. F. Pliński and S. Plińska, "Sensing with Terahertz Radiation of Pharma-and Bio-materials," *Procedia Engineering*, vol. 47, pp. 929–932, 2012.

[8] Y.-S. Lee, "Introduction to the Special Issue on "Terahertz Spectroscopy of Carbon Nanomaterials," *Journal of Infrared, Millimeter and Terahertz Waves*, pp. 1–2, 2012.

[9] Y.-S. Lee, *Principles of terahertz science and technology*, vol. 170. Springer Science & Business Media, 2009.

[10] C. A. Schmuttenmaer, "Exploring Dynamics in the Far-Infrared with Terahertz Spectroscopy," *Chemical Reviews*, vol. 104, no. 4, pp. 1759–1780, April 2004.

[11] H. H. Mantsch and D. Naumann, "Terahertz spectroscopy: The renaissance of far infrared spectroscopy," *Journal of Molecular Structure*, vol. 964, no. 1–3, pp. 1–4, February 2010.

[12] W. Gao, X. Degang, and Y. Jianquan, “Review of explosive detection using terahertz spectroscopy technique,” in *Electronics and Optoelectronics (ICEOE), 2011 International Conference on*, vol. 4, pp. 4–22, 2011.

[13] P. U. Jepsen, D. G. Cooke, and M. Koch, “Terahertz spectroscopy and imaging - Modern techniques and applications,” *Laser & Photonics Reviews*, vol. 5, no. 1, pp. 124–166, January 2011.

[14] W. Feng, “Review of terahertz semiconductor sources,” *Journal of Semiconductors*, vol. 33, no. 3, p. 031001, March 2012.

[15] T. Kleine-Ostmann and T. Nagatsuma, “A Review on Terahertz Communications Research,” *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 32, no. 2, pp. 143–171, February 2011.

[16] J. Federici and L. Moeller, “Review of terahertz and subterahertz wireless communications,” *Journal of Applied Physics*, vol. 107, no. 11, p. 22, 2010.

[17] E. Pickwell and V. P. Wallace, “Biomedical applications of terahertz technology,” *Journal of Physics D: Applied Physics*, vol. 39, no. 17, pp. R301–R310, September 2006.

[18] J. A. Zeitler, P. F. Taday, D. A. Newnham, M. Pepper, K. C. Gordon, and T. Rades, “Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting - a review,” *Journal of Pharmacy and Pharmacology*, vol. 59, no. 2, pp. 209–223, February 2007.

[19] K. Kawase, Y. Ogawa, Y. Watanabe, and H. Inoue, “Non-destructive terahertz imaging of illicit drugs using spectral fingerprints,” *Optics express*, vol. 11, no. 20, pp. 2549–2554, 2003.

[20] J. F. Federici, “Review of Moisture and Liquid Detection and Mapping using Terahertz Imaging,” *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 33, no. 2, pp. 97–126, February 2012.

[21] J. Qin, Y. Ying, and L. Xie, “The detection of agricultural products and food using terahertz spectroscopy: A review,” *Applied Spectroscopy Reviews*, vol. 48, no. 6, pp. 439–457, 2013.

[22] A. A. Gowen, C. O’Sullivan, and C. P. O’Donnell, “Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring

and quality control," *Trends in Food Science & Technology*, vol. 25, no. 1, pp. 40–46, May 2012.

[23] A. Pang, M. J. Warren, and R. W. Pickersgill, "Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site," *Acta Crystallographica Section D*, vol. 67, no. 2, pp. 91–96, February 2011.

[24] A. Pang, S. Frank, I. Brown, M. J. Warren, and R. W. Pickersgill, "Structural Insights into Higher Order Assembly and Function of the Bacterial Microcompartment Protein PduA," *Journal of Biological Chemistry*, vol. 289, no. 32, pp. 22377–22384, August 2014.

[25] M. E. Salvucci, K. W. Osteryoung, S. J. Crafts-Brandner, and E. Vierling, "Exceptional sensitivity of Rubisco activase to thermal denaturation *in vitro* and *in vivo*," *Plant physiology*, vol. 127, no. 3, pp. 1053–1064, 2001.

[26] G. Mei, A. Di Venere, M. Buganza, P. Vecchini, N. Rosato, and A. Finazzi-Agro, "Role of quaternary structure in the stability of dimeric proteins: the case of ascorbate oxidase," *Biochemistry*, vol. 36, no. 36, pp. 10917–10922, 1997.

[27] D. Shortle and M. S. Ackerman, "Persistence of native-like topology in a denatured protein in 8 M urea," *Science*, vol. 293, no. 5529, pp. 487–489, 2001.

[28] J.-H. Cho and D. P. Raleigh, "Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins," *Journal of molecular biology*, vol. 353, no. 1, pp. 174–185, 2005.

[29] E. Herczenik and M. F. Gebbink, "Molecular and cellular aspects of protein misfolding and disease," *The FASEB Journal*, vol. 22, no. 7, pp. 2115–2133, 2008.

[30] T. J. Huang, D.-S. Yang, P. E. Fraser, and A. Chakrabartty, "Alternate Aggregation Pathways of the Alzheimer β-Amyloid Peptide: An *in vitro* Model of Preamyloid," *Journal of Biological Chemistry*, vol. 275, no. 46, pp. 36436–36440, 2000.

[31] X. Yan, J. Watson, P. S. Ho, and M. L. Deinzer, "Mass spectrometric approaches using electrospray ionization charge states and hydrogen-deuterium exchange for determining protein structures and their conformational changes," *Molecular & Cellular Proteomics*, vol. 3, no. 1, pp. 10–23, 2004.

[32] C. B. Anfinsen, E. Haber, M. Sela, and F. White, "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain," *Proceedings of the National Academy of Sciences*, vol. 47, no. 9, pp. 1309–1314, 1961.

[33] C. Anfinsen and H. Scheraga, "Experimental and theoretical aspects of protein folding," *Advances in protein chemistry*, vol. 29, pp. 205–300, 1975.

[34] S. K. Pal, J. Peon, B. Bagchi, and A. H. Zewail, "Biological water: femtosecond dynamics of macromolecular hydration," *The Journal of Physical Chemistry B*, vol. 106, no. 48, pp. 12376–12395, 2002.

[35] S. Bone and R. Pethig, "Dielectric studies of the binding of water to lysozyme," *Journal of molecular biology*, vol. 157, no. 3, pp. 571–575, 1982.

[36] M. Lang, X. Jordanides, X. Song, and G. Fleming, "Aqueous solvation dynamics studied by photon echo spectroscopy," *The Journal of chemical physics*, vol. 110, no. 12, pp. 5884–5892, 1999.

[37] T. Li, A. A. Hassanali, Y.-T. Kao, D. Zhong, and S. J. Singer, "Hydration dynamics and time scales of coupled water-protein fluctuations," *Journal of the American Chemical Society*, vol. 129, no. 11, pp. 3376–3382, 2007.

[38] F. Pizzitutti, M. Marchi, F. Sterpone, and P. J. Rossky, "How protein surfaces induce anomalous dynamics of hydration water," *The Journal of Physical Chemistry B*, vol. 111, no. 26, pp. 7584–7590, 2007.

[39] A. A. Hassanali, T. Li, D. Zhong, and S. J. Singer, "A molecular dynamics study of Lys-Trp-Lys: structure and dynamics in solution following photoexcitation," *The Journal of Physical Chemistry B*, vol. 110, no. 21, pp. 10497–10508, 2006.

[40] D. E. Khoshtariya, E. Hansen, R. Leecharoen, and G. C. Walker, "Probing protein hydration by the difference O · H (O · D) vibrational spectroscopy: Interfacial percolation network involving highly polarizable water-water hydrogen bonds," *Journal of molecular liquids*, vol. 105, no. 1, pp. 13–36, 2003.

[41] T. Yokomizo, J. Higo, and M. Nakasako, "Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme," *Chemical physics letters*, vol. 410, no. 1, pp. 31–35, 2005.

[42] B. Honig and F. E. Cohen, "Adding backbone to protein folding: why proteins are polypeptides," *Folding and Design*, vol. 1, no. 1, pp. R17–R20, 1996.

[43] P. H. Thibodeau, C. A. Brautigam, M. Machius, and P. J. Thomas, "Side chain and backbone contributions of Phe508 to CFTR folding," *Nature structural & molecular biology*, vol. 12, no. 1, pp. 10–16, 2005.

[44] L. Meinhold and J. C. Smith, "Protein dynamics from X-ray crystallography: Anisotropic, global motion in diffuse scattering patterns," *Proteins: Structure, Function, and Bioinformatics*, vol. 66, no. 4, pp. 941–953, December 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*, vol. 15, no. 5, pp. 1175–1181, 2006.

[46] 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*, vol. 82, no. 15, pp. 4995–4999, 1985.

[47] O. Sushko, R. Dubrovka, and R. S. Donnan, "Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected," *The Journal of Chemical Physics*, vol. 142, no. 5, p. 055101, February 2015.

[48] E. P. J. Parrott, Y. Sun, and E. Pickwell-MacPherson, "Terahertz spectroscopy: Its future role in medical diagnoses," *Journal of Molecular Structure*, vol. 1006, no. 1–3, pp. 66–76, December 2011.

[49] T. Ding, A. P. J. Middelberg, T. Huber, and R. J. Falconer, "Far-infrared spectroscopy analysis of linear and cyclic peptides, and lysozyme," *Vibrational Spectroscopy*, vol. 61, pp. 144–150, July 2012.

[50] E. Castro-Camus and M. B. Johnston, "Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy," *Chemical Physics Letters*, vol. 455, no. 4–6, pp. 289–292, April 2008.

[51] D. M. Leitner, M. Gruebele, and M. Havenith, "Solvation dynamics of biomolecules: modeling and terahertz experiments," *HFSP journal*, vol. 2, no. 6, pp. 314–323, 2008.

[52] T. H. Basey-Fisher, S. M. Hanham, H. Andresen, S. A. Maier, M. M. Stevens, N. M. Alford, and N. Klein, "Microwave Debye relaxation analysis of dissolved proteins: Towards free-solution biosensing," *Applied Physics Letters*, vol. 99, no. 23, p. 233703, 2011.

[53] S. Laurette, A. Treizebre, F. Affouard, and B. Bocquet, "Subterahertz characterization of ethanol hydration layers by microfluidic system," *Applied Physics Letters*, vol. 97, no. 11, p. 111904, 2010.

[54] A. G. Markelz, "Terahertz Dielectric Sensitivity to Biomolecular Structure and Function," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, no. 1, pp. 180–190, 2008.

[55] 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*, vol. 110, no. 47, pp. 24255–24259, November 2006.

[56] 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*, vol. 104, no. 52, pp. 20749–20752, 2007.

[57] M. Heyden, D. J. Tobias, and D. V. Matyushov, "Terahertz absorption of dilute aqueous solutions," *The Journal of Chemical Physics*, vol. 137, no. 23, p. 235103, 2012.

[58] A. Markelz, S. Whitmire, J. Hillebrecht, and R. Birge, "THz time domain spectroscopy of biomolecular conformational modes," *Physics in medicine and biology*, vol. 47, no. 21, p. 3797, 2002.

[59] S. E. Whitmire, D. Wolpert, A. G. Markelz, J. R. Hillebrecht, J. Galan, and R. R. Birge, "Protein flexibility and conformational state: A comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin," *Biophysical journal*, vol. 85, no. 2, pp. 1269–1277, 2003.

[60] A. Markelz, A. Roitberg, and E. Heilweil, "Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz," *Chemical Physics Letters*, vol. 320, no. 1, pp. 42–48, 2000.

[61] 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*, vol. 130, no. 8, pp. 2374–2375, February 2008.

[62] C. Zhang and S. M. Durbin, "Hydration-induced far-infrared absorption increase in myoglobin," *The Journal of Physical Chemistry B*, vol. 110, no. 46, pp. 23607–23613, 2006.

[63] J.-Y. Chen, J. R. Knab, S. Ye, Y. He, and A. G. Markelz, "Terahertz dielectric assay of solution phase protein binding," *Applied Physics Letters*, vol. 90, no. 24, p. 243901, 2007.

[64] J. Xu, K. W. Plaxco, and S. J. Allen, "Absorption spectra of liquid water and aqueous buffers between 0.3 and 3.72 THz," *The Journal of Chemical Physics*, vol. 124, no. 3, p. 036101, 2006.

[65] N. Q. Vinh, S. J. Allen, and K. W. Plaxco, "Dielectric Spectroscopy of Proteins as a Quantitative Experimental Test of Computational Models of Their Low-Frequency Harmonic Motions," *Journal of the American Chemical Society*, vol. 133, no. 23, pp. 8942–8947, June 2011.

[66] M. Heyden and M. Havenith, "Combining THz spectroscopy and MD simulations to study protein-hydration coupling," *Methods*, vol. 52, no. 1, pp. 74–83, September 2010.

[67] M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx, "Dissecting the THz spectrum of liquid water from first principles via correlations in time and space," *Proceedings of the National Academy of Sciences*, vol. 107, no. 27, pp. 12068–12073, July 2010.

[68] M.-P. Gaigeot, M. Martinez, and R. Vuilleumier, "Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides," *Molecular Physics*, vol. 105, no. 19–22, pp. 2857–2878, 2007.

[69] P. H. Berens, D. H. J. Mackay, G. M. White, and K. R. Wilson, "Thermodynamics and quantum corrections from molecular dynamics for liquid water," *The Journal of Chemical Physics*, vol. 79, no. 5, p. 2375, 1983.

[70] R. Liu, M. He, R. Su, Y. Yu, W. Qi, and Z. He, "Insulin amyloid fibrillation studied by terahertz spectroscopy and other biophysical methods," *Biochemical and Biophysical Research Communications*, vol. 391, no. 1, pp. 862–867, January 2010.

[71] C. Hua, C. Gui-Ying, L. Shu-Qin, and W. Li, "Reversible conformational changes of PsbO protein detected by terahertz Time-domain spectroscopy," *Chinese Physics Letters*, vol. 26, no. 8, p. 084204, 2009.

[72] C. Hua, Q. Yuan-Gang, P. Wei-Xian, K. Ting-Yun, L. Liang-Bi, and W. Li, "Investigation of Chlorophyll Protein 43 and 47 Denaturation by Terahertz Time-Domain Spectroscopy," *Chinese Physics Letters*, vol. 24, no. 7, p. 2131, 2007.

[73] Y. Qu, H. Chen, X. Qin, L. Li, L. Wang, and T. Kuang, "Thermal denaturation of CP43 studied by Fourier transform-infrared spectroscopy and terahertz time-domain spectroscopy," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1774, no. 12, pp. 1614–1618, December 2007.

[74] Y. Qu, H. Chen, X. Qin, L. Wang, L. Li, and T. Kuang, "The guanidine hydrochloride-induced denaturation of CP43 and CP47 studied by terahertz time-domain spectroscopy," *Science in China Series C: Life Sciences*, vol. 50, no. 3, pp. 350–355, 2007.

[75] H. Yoneyama, M. Yamashita, S. Kasai, K. Kawase, R. Ueno, H. Ito, and T. Ouchi, "Terahertz spectroscopy of native-conformation and thermally denatured bovine serum albumin (BSA)," *Physics in Medicine and Biology*, vol. 53, no. 13, pp. 3543–3549, July 2008.

[76] X. Li, X. Fu, J. Liu, Y. Du, and Z. Hong, "Investigation of thermal denaturation of solid bovine serum albumin by terahertz dielectric spectroscopy," *Journal of Molecular Structure*, vol. 1049, pp. 441–445, October 2013.

[77] 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*, vol. 101, no. 4, pp. 925–933, August 2011.

[78] C. U. Stehle, W. Abuillan, B. Gompf, and M. Dressel, "Far-infrared spectroscopy on free-standing protein films under defined temperature and hydration control," *The Journal of chemical physics*, vol. 136, no. 7, p. 075102, 2012.

[79] T. Suzuki, Y. Ogawa, T. Kondo, S. Kamba, and N. Kondo, "THz spectroscopy based high sensitivity measurement of protein using a metal mesh device," in *SPIE Defense, Security, and Sensing*, 2011, pp. 802706–802706.

[80] F. W. Teale, "Cleavage of the haem-protein link by acid methylethylketone.," *Biochimica et biophysica acta*, vol. 35, p. 543, October 1959.

[81] U. K. Laemmli, "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, August 1970.

[82] H. Edelhoch, "Spectroscopic determination of tryptophan and tyrosine in proteins," *Biochemistry*, vol. 6, no. 7, pp. 1948–1954, 1967.

[83] Z. Chi and S. A. Asher, "Ultraviolet Resonance Raman Examination of Horse Apomyoglobin Acid Unfolding Intermediates," *Biochemistry*, vol. 38, no. 26, pp. 8196–8203, June 1999.

[84] G. Böhm, R. Muhr, and R. Jaenicke, "Quantitative analysis of protein far UV circular dichroism spectra by neural networks," *Protein engineering*, vol. 5, no. 3, pp. 191–195, 1992.

[85] N. Karpowicz, H. Zhong, J. Xu, K.-I. Lin, J.-S. Hwang, and X.-C. Zhang, "Comparison between pulsed terahertz time-domain imaging and continuous wave terahertz imaging," *Semiconductor Science and Technology*, vol. 20, no. 7, pp. S293–S299, July 2005.

[86] Tanouchi, M. "New frontier of terahertz technology." *18th International Conference on Applied Electromagnetics and Communications.* 2005.

[87] J. M. Chamberlain, Institute of Electrical and Electronic Engineers, Eds., *THz 2002: 2002 IEEE Tenth International Conference on Terahertz Electronics*

proceedings: Selwyn College, University of Cambridge, Cambridge, UK, 9th and 10th September 2002. Piscataway, NJ: IEEE, 2002.

[88] D. J. Segelstein, "The complex refractive index of water," University of Missouri–Kansas City, 2011.

[89] K. Shala, B. Yang, R. Dudley, T. T. Chowdhury, and R. S. Donnan, "Quantification of glucose and cytosine hydrochloride by THz-TDS," in *2009 Loughborough Antennas & Propagation Conference*, 2009.

[90] D. K. Ghodgaonkar, V. V. Varadan, and V. K. Varadan, "Free-space measurement of complex permittivity and complex permeability of magnetic materials at microwave frequencies," *Instrumentation and Measurement, IEEE Transactions on*, vol. 39, no. 2, pp. 387–394, 1990.

[91] T. Ding, R. Li, J. A. Zeitler, T. L. Huber, L. F. Gladden, A. P. J. Middelberg, and R. J. Falconer, "Terahertz and far infrared Spectroscopy of alanine-rich peptides having variable ellipticity," *Opt. Express*, vol. 18, no. 26, pp. 27431–27444, December 2010.

[92] B. G. Hawkins, A. E. Smith, Y. A. Syed, and B. J. Kirby, "Continuous-flow particle separation by 3D insulative dielectrophoresis using coherently shaped, dc-biased, ac electric fields," *Analytical chemistry*, vol. 79, no. 19, pp. 7291–7300, 2007.

[93] J. W. Bye and R. J. Falconer, "Three Stages of Lysozyme Thermal Stabilization by High and Medium Charge Density Anions," *The Journal of Physical Chemistry B*, vol. 118, no. 16, pp. 4282–4286, April 2014.

[94] V. Sharma, F. Böhm, M. Seitz, G. Schwaab, and M. Havenith, "From solvated ions to ion-pairing: a THz study of lanthanum(iii) hydration," *Physical Chemistry Chemical Physics*, vol. 15, no. 21, p. 8383, 2013.

[95] B. Yang, R. J. Wylde, D. H. Martin, P. Goy, R. S. Donnan, and S. Caroopen, "Determination of the gyrotropic characteristics of hexaferrite ceramics from 75 to 600 GHz," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 12, pp. 3587–3597, 2010.

[96] L. Duvillaret, F. Garet, and J.-L. Coutaz, "Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy," *Applied Optics*, vol. 38, no. 2, p. 409, January 1999.

[97] J. Werner, E. Wernersson, V. Ekholm, N. Ottosson, G. Öhrwall, J. Heyda, I. Persson, J. Söderström, P. Jungwirth, and O. Björneholm, "Surface Behavior of Hydrated Guanidinium and Ammonium Ions: A Comparative Study by Photoelectron Spectroscopy and Molecular Dynamics," *The Journal of Physical Chemistry B*, vol. 118, no. 25, pp. 7119–7127, June 2014.

[98] P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes, and J. M. Cruickshank, "The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution," *Proceedings of the National Academy of Sciences*, vol. 100, no. 8, pp. 4557–4561, 2003.

[99] J. Hunger, S. Niedermayer, R. Buchner, and G. Hefter, "Are Nanoscale Ion Aggregates Present in Aqueous Solutions of Guanidinium Salts?," *The Journal of Physical Chemistry B*, vol. 114, no. 43, pp. 13617–13627, November 2010.

[100] C. Camilloni, A. Guerini Rocco, I. Eberini, E. Gianazza, R. A. Broglia, and G. Tiana, "Urea and Guanidinium Chloride Denature Protein L in Different Ways in Molecular Dynamics Simulations," *Biophysical Journal*, vol. 94, no. 12, pp. 4654–4661, June 2008.

[101] N. Samanta, D. D. Mahanta, and R. K. Mitra, "Collective hydration dynamics of guanidinium chloride solutions and its possible role in protein denaturation: a terahertz spectroscopic study," *Phys. Chem. Chem. Phys.*, vol. 16, no. 42, pp. 23308–23315, September 2014.

[102] M. S. Akhtar, A. Ahmad, and V. Bhakuni, "Guanidinium Chloride- and Urea-Induced Unfolding of the Dimeric Enzyme Glucose Oxidase," *Biochemistry*, vol. 41, no. 11, pp. 3819–3827, March 2002.

[103] O. Sushko, J. Qiu, R. Dubrovka, R. W. Pickersgill, and R. S. Donnan, "Sub-terahertz spectroscopy as a probe for protein stability in an ionic environment," in *Infrared, Millimeter, and Terahertz waves (IRMMW-THz), 40th International Conference on*, pp. 1–2, 2015.

[104] A. K. Upadhyay, A. Singh, K. J. Mukherjee, and A. K. Panda, "Refolding and purification of recombinant L-asparaginase from inclusion bodies of *E. coli* into active tetrameric protein," *Frontiers in Microbiology*, vol. 5, p. 486, September 2014.

[105] F. Hofmeister, "On the understanding of the effects of salts," *Arch. Exp. Pathol. Pharmakol.(Leipzig)*, vol. 24, pp. 247–260, 1888.

[106] W. Kunz, J. Henle, and B. W. Ninham, "Zur Lehre von der Wirkung der Salze' (about the science of the effect of salts): Franz Hofmeister's historical papers," *Current Opinion in Colloid & Interface Science*, vol. 9, no. 1–2, pp. 19–37, Aug. 2004.

[107] K. J. Tielrooij, N. Garcia-Araez, M. Bonn, and H. J. Bakker, "Cooperativity in Ion Hydration," *Science*, vol. 328, no. 5981, pp. 1006–1009, May 2010.

[108] R. Mancinelli, A. Botti, F. Bruni, M. A. Ricci, and A. K. Soper, "Perturbation of water structure due to monovalent ions in solution," *Physical Chemistry Chemical Physics*, vol. 9, no. 23, p. 2959, 2007.

[109] S. J. Irudayam and R. H. Henchman, "Long-range hydrogen-bond structure in aqueous solutions and the vapor-water interface," *The Journal of Chemical Physics*, vol. 137, no. 3, p. 034508, 2012.

[110] D. C. Rees and A. D. Robertson, "Some thermodynamic implications for the thermostability of proteins," *Protein Science*, vol. 10, no. 6, pp. 1187–1194, 2001.

[111] R. L. Baldwin, "In Search of the Energetic Role of Peptide Hydrogen Bonds," *Journal of Biological Chemistry*, vol. 278, no. 20, pp. 17581–17588, May 2003.

[112] R. Aurora, T. P. Creamer, R. Srinivasan, and G. D. Rose, "Local interactions in protein folding: lessons from the α-helix," *Journal of Biological Chemistry*, vol. 272, no. 3, pp. 1413–1416, 1997.

[113] W. Kauzmann, "Some Factors in the Interpretation of Protein Denaturation," *Advances in protein chemistry*, vol. 14, pp. 1–63, 1959.

[114] K. A. Dill, "Theory for the folding and stability of globular proteins," *Biochemistry*, vol. 24, no. 6, pp. 1501–1509, 1985.

[115] J. P. Collman, R. Boulatov, C. J. Sunderland, and L. Fu, "Functional Analogues of Cytochrome c Oxidase, Myoglobin, and Hemoglobin," *Chemical Reviews*, vol. 104, no. 2, pp. 561–588, February 2004.

[116] I. Sirangelo, E. Bismuto, S. Tavassi, and G. Irace, "Apomyoglobin folding intermediates characterized by the hydrophobic fluorescent probe 8-anilino-1-naphthalene sulfonate (ANS)," *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, vol. 1385, no. 1, pp. 69–77, 1998.

[117] I. Sirangelo, F. Dal Piaz, C. Malmo, M. Casillo, L. Birolo, P. Pucci, G. Marino, and G. Irace, "Hexafluoroisopropanol and Acid Destabilized Forms of Apomyoglobin Exhibit Structural Differences," *Biochemistry*, vol. 42, no. 2, pp. 312–319, January 2003.

[118] M. S. Kay and R. L. Baldwin, "Packing interactions in the apomyglobin folding intermediate," *Nature structural biology*, vol. 3, no. 5, pp. 439–445, 1996.

[119] F. Hughson, P. Wright, and R. Baldwin, "Structural characterization of a partly folded apomyoglobin intermediate," *Science*, vol. 249, no. 4976, pp. 1544–1548, 1990.

[120] D. Barrick and R. L. Baldwin, "The molten globule intermediate of apomyoglobin and the process of protein folding," *Protein Science*, vol. 2, no. 6, pp. 869–876, 1993.

[121] I. Sirangelo, C. Iannuzzi, C. Malmo, and G. Irace, "Tryptophanyl substitutions in apomyoglobin affect conformation and dynamic properties of AGH subdomain," *Biopolymers*, vol. 70, no. 4, pp. 649–654, 2003.

[122] J. Tirado-Rives and W. L. Jorgensen, "Molecular dynamics simulations of the unfolding of apomyoglobin in water," *Biochemistry*, vol. 32, no. 16, pp. 4175–4184, 1993.

[123] A. Onufriev, D. A. Case, and D. Bashford, "Structural details, pathways, and energetics of unfolding apomyoglobin," *Journal of molecular biology*, vol. 325, no. 3, pp. 555–567, 2003.

[124] T. Uzawa, S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, and T. Fujisawa, "Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and

compactness," *Proceedings of the National Academy of Sciences*, vol. 101, no. 5, pp. 1171–1176, 2004.

[125] K. Gast, H. Damaschun, R. Misselwitz, M. Müller-Frohne, D. Zirwer, and G. Damaschun, "Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate," *European Biophysics Journal*, vol. 23, no. 4, pp. 297–305, 1994.

[126] Y. V. Griko, P. L. Privalov, S. Y. Venyaminov, and V. P. Kutyshenko, "Thermodynamic study of the apomyoglobin structure," *Journal of Molecular Biology*, vol. 202, no. 1, pp. 127–138, 1988.

[127] S. C. Harrison and E. R. Blout, "Reversible conformational changes of myoglobin and apomyoglobin," *Journal of Biological Chemistry*, vol. 240, pp. 299–303, 1965.

[128] C. A. Kerfeld, S. Heinhorst, and G. C. Cannon, "Bacterial Microcompartments," *Annual Review of Microbiology*, vol. 64, no. 1, pp. 391–408, October 2010.

[129] J. M. Shively, C. E. Bradburne, H. C. Aldrich, T. A. Bobik, J. L. Mehlman, S. Jin, and S. H. Baker, "Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes - polyhedral bodies," *Canadian Journal of Botany*, vol. 76, no. 6, pp. 906–916, 1998.

[130] T. A. Bobik, G. D. Havemann, R. J. Busch, D. S. Williams, and H. C. Aldrich, "The Propanediol Utilization (pdu) Operon of *Salmonella enterica* Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1, 2-Propanediol Degradation," *Journal of bacteriology*, vol. 181, no. 19, pp. 5967–5975, 1999.

[131] D. D. Sriramulu, M. Liang, D. Hernandez-Romero, E. Raux-Deery, H. Lunsdorf, J. B. Parsons, M. J. Warren, and M. B. Prentice, "*Lactobacillus reuteri* DSM 20016 Produces Cobalamin-Dependent Diol Dehydratase in Metabolosomes and Metabolizes 1,2-Propanediol by Disproportionation," *Journal of Bacteriology*, vol. 190, no. 13, pp. 4559–4567, July 2008.

[132] E. Bock, D. Duvel, and K. R. Peters, “Characterization of a phage-like particle from cells of Nitrobacter. I. Host-particle correlation and particle isolation,” *Archives of microbiology*, vol. 97, no. 2, pp. 115–127, April 1974.

[133] S. R. Brinsmade, T. Paldon, and J. C. Escalante-Semerena, “Minimal Functions and Physiological Conditions Required for Growth of *Salmonella enterica* on Ethanolamine in the Absence of the Metabolosome,” *Journal of Bacteriology*, vol. 187, no. 23, pp. 8039–8046, December 2005.

[134] I. Stojiljkovic, A. J. Bäumler, and F. Heffron, “Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster.,” *Journal of bacteriology*, vol. 177, no. 5, pp. 1357–1366, 1995.

[135] D. Walter, M. Ailion, and J. Roth, “Genetic characterization of the pdu operon: use of 1, 2-propanediol in *Salmonella typhimurium*,” *Journal of bacteriology*, vol. 179, no. 4, pp. 1013–1022, 1997.

[136] H. Seedorf, W. F. Fricke, B. Veith, H. Brüggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer, and G. Gottschalk, “The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 2128–2133, 2008.

[137] J. Jorda, D. Lopez, N. M. Wheatley, and T. O. Yeates, “Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria,” *Protein Science*, vol. 22, no. 2, pp. 179–195, February 2013.

[138] C. A. Kerfeld, M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Phillips, M. Beeby, and T. O. Yeates, “Protein structures forming the shell of primitive bacterial organelles,” *Science*, vol. 309, no. 5736, pp. 936–938, 2005.

[139] D. Heldt, S. Frank, A. Seyedarabi, D. Ladikis, J. B. Parsons, M. J. Warren, and R. W. Pickersgill, “Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*,” *Biochemical Journal*, vol. 423, no. 2, pp. 199–207, October 2009.

[140] T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, and J. M. Shively, “Protein-based organelles in bacteria: carboxysomes and related

microcompartments," *Nature Reviews Microbiology*, vol. 6, no. 9, pp. 681–691, September 2008.

[141] J. B. Parsons, S. D. Dinesh, E. Deery, H. K. Leech, A. A. Brindley, D. Heldt, S. Frank, C. M. Smales, H. Lunsdorf, A. Rambach, M. H. Gass, A. Bleloch, K. J. McClean, A. W. Munro, S. E. J. Rigby, M. J. Warren, and M. B. Prentice, "Biochemical and Structural Insights into Bacterial Organelle Form and Biogenesis," *Journal of Biological Chemistry*, vol. 283, no. 21, pp. 14366–14375, May 2008.

[142] G. D. Havemann and T. A. Bobik, "Protein Content of Polyhedral Organelles Involved in Coenzyme B12-Dependent Degradation of 1,2-Propanediol in *Salmonella enterica* Serovar Typhimurium LT2," *Journal of Bacteriology*, vol. 185, no. 17, pp. 5086–5095, September 2003.

[143] J. B. Parsons, S. Frank, D. Bhella, M. Liang, M. B. Prentice, D. P. Mulvihill, and M. J. Warren, "Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement," *Molecular Cell*, vol. 38, no. 2, pp. 305–315, April 2010.

[144] S. Cheng, S. Sinha, C. Fan, Y. Liu, and T. A. Bobik, "Genetic Analysis of the Protein Shell of the Microcompartments Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation by Salmonella," *Journal of Bacteriology*, vol. 193, no. 6, pp. 1385–1392, March. 2011.

[145] C. S. Crowley, D. Cascio, M. R. Sawaya, J. S. Kopstein, T. A. Bobik, and T. O. Yeates, "Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell," *Journal of Biological Chemistry*, vol. 285, no. 48, pp. 37838–37846, November 2010.

[146] E. Y. Kim, C. M. Jakobson, and D. Tullman-Ercek, "Engineering transcriptional regulation to control Pdu microcompartment formation," *PloS one*, vol. 9, no. 11, p. e113814, 2014.

[147] A. D. Lawrence, S. Frank, S. Newnham, M. J. Lee, I. R. Brown, W.-F. Xue, M. L. Rowe, D. P. Mulvihill, M. B. Prentice, M. J. Howard, and M. J. Warren, "Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor," *ACS Synthetic Biology*, vol. 3, no. 7, pp. 454–465, July 2014.

[148] S. M. Read and D. H. Northcote, “Minimization of variation in the response to different proteins of the Coomassie blue G dye-binding assay for protein,” *Analytical Biochemistry*, vol. 116, no. 1, pp. 53–64, 1981.

[149] S. F. de S. Groth, R. Webster, and A. Datyner, “Two new staining procedures for quantitative estimation of proteins on electrophoretic strips,” *Biochimica et biophysica acta*, vol. 71, pp. 377–391, 1963.

[150] H. J. Liebe, G. A. Hufford, and T. Manabe, “A model for the complex permittivity of water at frequencies below 1 THz,” *International Journal of Infrared and Millimeter Waves*, vol. 12, no. 7, pp. 659–675, 1991.

[151] S. Zhang, J. Zhou, Y.-S. Park, J. Rho, R. Singh, S. Nam, A. K. Azad, H.-T. Chen, X. Yin, A. J. Taylor, and X. Zhang, “Photoinduced handedness switching in terahertz chiral metamolecules,” *Nature Communications*, vol. 3, pp. 942, July 2012.

[152] R. W. Woody, “Circular dichroism in Biochemical Spectroscopy”, vol. 246, *Academic Press*, pp. 34–71, 1995.

[153] L. Whitmore and B. A. Wallace, “Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases,” *Biopolymers*, vol. 89, no. 5, pp. 392–400, 2008.

[154] C. Giri, F. Goesmann, C. Meinert, A. C. Evans, and U. J. Meierhenrich, “Synthesis and Chirality of Amino Acids Under Interstellar Conditions,” *Biochirality*, vol. 333, P. Cintas, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 41–82, 2012.

[155] M. L. Neidig, A. T. Wecksler, G. Schenk, T. R. Holman, and E. I. Solomon, “Kinetic and Spectroscopic Studies of N694C Lipoxygenase: A Probe of the Substrate Activation Mechanism of a Nonheme Ferric Enzyme,” *Journal of the American Chemical Society*, vol. 129, no. 24, pp. 7531–7537, June 2007.

[156] A. F. Drake, “Polarisation modulation-the measurement of linear and circular dichroism,” *Journal of Physics E: Scientific Instruments*, vol. 19, no. 3, pp. 170, 1986.

[157] P. L. Polavarapu and C. Zhao, “Vibrational circular dichroism: a new spectroscopic tool for biomolecular structural determination,” *Fresenius' journal of analytical chemistry*, vol. 366, no. 6–7, pp. 727–734, 2000.

[158] N. Go, 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*, vol. 80, no. 12, pp. 3696–3700, 1983.

[159] P. Dauber-Osguthorpe, D. J. Osguthorpe, P. S. Stern, and J. Moult, “Low Frequency Motion in Proteins,” *J. Comput. Phys.*, vol. 151, no. 1, pp. 169–189, May 1999.

[160] 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*, vol. 80, no. 21, pp. 6571–6575, 1983.

[161] B. Yang, R. J. Wylde, D. H. Martin, P. Goy, R. S. Donnan, and S. Caroopen, “Determination of the gyrotropic characteristics of hexaferrite ceramics from 75 to 600 GHz,” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 12, pp. 3587–3597, 2010.

[162] B. Yang and R. S. Donnan, “Enhanced rapid and accurate sub-THz magneto-optical characterization of hexaferrite ceramics,” *Journal of Magnetism and Magnetic Materials*, vol. 323, no. 15, pp. 1992–1997, August 2011.

[163] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, “Clustal W and Clustal X version 2.0,” *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, November 2007.

[164] N. Pugliano and R. J. Saykally, "Measurement of quantum tunneling between chiral isomers of the cyclic water trimer," *Science-New York Then Washington*, pp.1937-1937, 1992.

[165] F. N. Keutsch and R. J. Saykally, "Water clusters: untangling the mysteries of the liquid, one molecule at a time," *Proceedings of the National Academy of Sciences*, 98(19), pp.10533-10540, 2001.

[166] I. Roy and M. N. Gupta, "Freeze-drying of proteins: Some emerging concerns," *Biotechnology and applied biochemistry*, 39(2), pp.165-177, 2004.

[167] T. Globus, M. Bykhovskaia, D. Woolard, and B. Gelmont, "Sub-millimetre wave absorption spectra of artificial RNA molecules," *Journal of Physics D: Applied Physics*, 36(11), pp.1314, 2003.

[168] R. Parthasarathy, T. Globus, T. Khromova, N. Swami, and D. Woolard, "Dielectric properties of biological molecules in the Terahertz gap," *Applied Physics Letters*, 87(11), pp.113901. 2005.

[169] T. Globus, D. Woolard, T. W. Crowe, T. Khromova, B. Gelmont, and J. Hesler, "Terahertz Fourier transform characterization of biological materials in a liquid phase," *Journal of Physics D: Applied Physics*, 39(15), pp.3405, 2006.

[170] R. K. Galvão, S. Hadjiloucas, A. Zafiropoulos, G. C. Walker, J. W. Bowen, and R. Dudley, "Optimization of apodization functions in terahertz transient spectrometry," *Optics letters*, 32(20), pp.3008-3010, 2007.

[171] E. Bründermann, B. Born, S. Funkner, M. Krüger, and M. Havenith, "Terahertz spectroscopic techniques for the study of proteins in aqueous solutions," *In SPIE OPTO: Integrated Optoelectronic Devices*, pp. 72150E-72150E, February 2009.

[172] J. Dielmann, V. C. Nibali, B. Born, E. Brundermann, and M. Havenith, "Kinetic terahertz absorption spectroscopy of protein solutions," *In 2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, September 2013.

[173] B. M. Fischer, M. Walther, and P. U. Jepsen, "Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy," *Physics in medicine and biology*,47(21), pp.3807, 2002.

[174] T. D. Dorney, R. G. Baraniuk, and D. M. Mittleman, "Material parameter estimation with terahertz time-domain spectroscopy," *JOSA A*, 18(7), pp.1562-1571, 2001.

[175] S. Hadjiloucas, G. C. Walker and J. W. Bowen. "A 1-Port De-embedding Technique for the Quasi-Optical Characterization of Integrated Components," *IEEE Sensors Journal*, 13(1), pp. 111-123, 2013.