MYOGLOBIN FAR-INFRARED ABSORPTION AND PROTEIN HYDRATION EFFECTS STUDIED BY TERAHERTZ TIME-DOMAIN SPECTROSCOPY

A Dissertation
Submitted to the Faculty
of
Purdue University
by
Chenfeng Zhang

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December 2006
Purdue University
West Lafayette, Indiana
ACKNOWLEDGMENTS

Look back at my six years’ study at the Department of Physics, Purdue University as a Ph.D. student, there are many people I want to acknowledge. Without their help and support, this thesis would never be possible.

First of all, I want to thank my advisors, Prof. Stephen M. Durbin and Prof. Andrew M. Weiner.

I remember the talk I had with Prof. Durbin once before I started working on my Ph.D. project with him. He told me that a Ph.D. is considered qualified only if he can do independent scientific research. For the past six years, I have been learning from this important lesson, with tremendous support and uncountable instructive discussions from Prof. Durbin. With great patience, he showed me the way to write and polish a scientific paper. I owed him much on this. His kindness will always remain in my memory. I feel fortunate to be his student.

It was more than four years ago when I sat in the nonlinear optics class taught by Prof. Weiner, and then another important class of his, ultrafast optics. I was greatly impressed by his lectures---it seemed that there was no question he didn’t have a clear and knowledgeable answer for. These two courses laid the theoretical foundation for my future research in his ultrafast optics and optical fiber communications laboratory. While working in the lab and being embarrassed by the questions raised in the group meetings, I always wished I have had done better in the classes. During the whole PhD project, the discussions with Prof. Weiner always turned out to be very illuminating and
fruitful. I am deeply impressed by his critical thinking and the discussions with him put this thesis on a more solid foundation.

I also want to thank Prof. John P. Finley, Prof. Earl W. Prohofsky, and Prof. Sergei Savikhin for sitting in my committee. The time you put in my defense and your patience in reading my thesis are highly appreciated.

Special thanks go to Dr. Daniel Leaird, Dr. Haifeng Wang, and Dr. Kristl Adams. In the lab, I benefited tremendously from the skills and knowledge of Dr. Leaird in optics experiments. I really appreciate his help. I have been sharing the same Ti:Sapphire laser source with Dr. Wang. When I was stuck with laser problems, he was always there, willing to help. I truly enjoyed the time we spent together in the lab. Dr. Adams just graduated from the group this May. When I needed anything in biology, she was always the first person I asked for help.

There are, of course, many other people in the group I got help from. The time I spent in the lab with them will always remain an unforgettable memory in my mind.

My parents, Xinmin Zhang and Xiuxia Wang, brought me up as a righteous person. They did their best to support me through my education, both physically and spiritually. If I would be successful in some way, they are part of it.

Most importantly, I want to thank my wife, Junni Liu. No matter whenever it is, wherever I am, and whatever I would become, I know she will always be around, supporting, encouraging, and loving me. We will be together, forever, for the time to come.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... vi
LIST OF FIGURES ........................................................................................................ vii
ABSTRACT .................................................................................................................. xii

CHAPTER 1. INTRODUCTION ............................................................................. 1

2.1. Generation of Terahertz Radiation ................................................................. 6
   2.1.1. Transient Photoconductivity Method ...................................................... 6
   2.1.2. Optical Rectification Method ................................................................. 11
   2.1.3. Other THz Generation Mechanisms .................................................... 13

2.2. Detection of Terahertz Radiation ................................................................. 14
   2.2.1. Photoconductive Detection .................................................................. 14
   2.2.2. Electro-optic Sampling ....................................................................... 16
   2.2.3. Comparisons between Photoconductive Detection and Electro-optic Sampling ...................................................................................................... 17

2.3. Terahertz Time-Domain Spectroscopy ....................................................... 19

2.4. Applications of Terahertz Time-domain Spectroscopy .............................. 22
   2.4.1. Material Properties Characterization ................................................. 23
   2.4.2. Biological Identification ..................................................................... 24
   2.4.3. THz Sensing and Imaging .................................................................. 24
   2.4.4. Broadband Communication ................................................................ 25

CHAPTER 2. TERAHERTZ TIME-DOMAIN SPECTROSCOPY ...................... 5

CHAPTER 3. FAR-INFRARED ABSORPTION PROPERTIES OF MYOGLOBIN IN POWDERS AND IN SOLUTIONS ................................................................. 27

3.1. Introduction to Myoglobin Structure and Function ....................................... 27

3.2. Sample Preparation ...................................................................................... 30
   3.2.1. Myoglobin Powder Samples ................................................................ 30
   3.2.2. Myoglobin Solution Samples ................................................................ 35

3.3. Experiments .................................................................................................. 37
   3.3.1. Sample Cell Considerations .............................................................. 37
   3.3.2. Measurement Details ....................................................................... 40
   3.3.3. FTIR Spectroscopy .......................................................................... 42

3.4. Data Analysis and Material Parameters Extraction Procedures ............... 43
   3.4.1. Refractive Index of HDPE .................................................................. 43
   3.4.2. Parameters Extraction Methods for Powder Samples ....................... 44
   3.4.3. Parameters Extraction Methods for Solution Samples ....................... 56
   3.4.4. Windowing Effect on the Data Analysis .............................................. 60
3.4.5. Experimental Error Estimation ............................................................ 69
3.5. Results and Discussions ........................................................................... 72
  3.5.1. Absorption Spectrum of Water Vapor ................................................. 73
  3.5.2. Absorption Spectra of Myoglobin Powders ................................. 75
  3.5.3. Absorption Spectra of Myoglobin Aqueous Solutions ....................... 82
  3.5.4. Comparing Myoglobin Powder Absorption Spectrum with Absorption
         Spectrum of Water Vapor ....................................................................... 86
  3.5.5. Comparison with FTIR Measurements ........................................... 88
3.6. Conclusions .............................................................................................. 90
CHAPTER 4. BIOLOGICAL WATER AND HYDRATION EFFECT .............. 91
  4.1. Biological Water in Protein ................................................................. 92
  4.2. Basic Physics of Liquid Water Absorption in the THz Region ............... 96
  4.3. Myoglobin Far-Infrared Absorption in a Non-interacting Model .......... 99
  4.4. Myoglobin Far-Infrared Absorption in an Interacting Model .............. 108
  4.5. Conclusions ........................................................................................ 120
CHAPTER 5. THESIS SUMMARY .......................................................... 121
BIBLIOGRAPHY ......................................................................................... 123
VITA ........................................................................................................ 137
LIST OF TABLES

Table 3-1. Hydration result for Mb powders. Each specimen was stored for 3-4 days in a sealed desiccator with one saturated salt solution with a specific equilibrium relative humidity. The water content can be expressed in three related notations. Some notes for the table: (i) The “as-received sample” means the Mb lyophilized powder was not treated by any of the solutions. (ii) The “desiccant” means the sample was placed in a desiccator with desiccants to produce a dry environment. .................................................... 33

Table 3-2. Preparation of Mb aqueous solutions, indicating the masses of Mb lyophilized powder and deionized water needed. ........................................... 36

Table 3-3. Some parameters of the sample. The sample thickness is determined using the algorithm above. For the two samples with highest water content, we used the thinner sample holder so that the absorption was not too strong. ..................................................................................................................... 53

Table 3-4. Some parameters of Mb solutions. We first measured a certain volume of sample (second column in the table) and then the sample is weighed to get the mass. The sample density and Mb molar concentration can therefore be calculated. ........................................................................................................... 59

Table 3-5. Parameters for fitting Eqn. (3.23) to the absorption curves in Figure 3.21. ............................................................................................................. 79

Table 3-6. Fitting parameters for the absorption curves in Figure 3.23. .......... 83

Table 4-1. Relative weight and time constant for the four dispersion terms in Eqn. (4.1) [115] .......................................................... 94

Table 4-2. Absorption coefficients of liquid water and dry Mb at selected frequencies. The absorption coefficients are described in different units, accounting for the density and molar concentration of liquid water and dry Mb. ............................................................................................................. 102

Table 4-3. Density, Mb molar concentration, and water molar concentration for Mb powders and solutions at selected water contents. The liquid water (100 wt%) and dry Mb (1.2 wt%) are also included. The molecular weight of Mb and water are given by $MW_{Mb} = 16952 \text{ g/mol}$ and $MW_{H_2O} = 18 \text{ g/mol}$ . ............ 103
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2-1.</td>
<td>Terahertz gap in the electromagnetic spectrum (From Ref. [28])</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2-2.</td>
<td>Generation of THz radiation from a PC antenna pumped with femtosecond optical</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>pulses. The photoconductor consists of two gold strips deposited on a LT-GaAs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>substrate. The PC antenna is viewed both from the side (side view) and the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>top (top view). (Figure after Ref. [39])</td>
<td></td>
</tr>
<tr>
<td>Figure 2-3.</td>
<td>Calculated photocurrent in the emitter and amplitude of the radiated THz</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>field versus time. The incident optical pulse is assumed to have a Gaussian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>temporal profile. (Figure taken from Ref. [41]. The circles are data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>extracted from Ref. [42].)</td>
<td></td>
</tr>
<tr>
<td>Figure 2-4.</td>
<td>Generation of THz pulses by optical rectification in a nonlinear material.</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>The induced THz field is proportional to the second-order time derivative of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the incident optical intensity. See explanations in the text.</td>
<td></td>
</tr>
<tr>
<td>Figure 2-5.</td>
<td>Detection of THz pulses by photoconductive antenna. The photocarriers</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>generated in the GaAs substrate by the gated optical pulse are driven by the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>incoming THz pulse and result in a transient current whose amplitude is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>proportional to the THz field.</td>
<td></td>
</tr>
<tr>
<td>Figure 2-6.</td>
<td>Principle of photoconductive sampling. The photoconductive antenna acts as</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>a sampling gate that detects the THz signal when the optical pulse is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>present. The detected signal is actually a convolution between THz pulse and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>optical pulse. By changing the delay between the optical pulse and THz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pulse, the entire THz waveform can be mapped out sequentially in time.</td>
<td></td>
</tr>
<tr>
<td>Figure 2-7.</td>
<td>Principle of electro-optic sampling (From Ref. [63]). The birefringence</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>induced by the THz pulse in the electro-optic crystal will be seen by the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>optical sampling pulse. The incoming circularly polarized beam will become</td>
<td></td>
</tr>
<tr>
<td></td>
<td>elliptically polarized after passing the crystal. The phase retardation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>between the two polarization components is proportional to the THz electric</td>
<td></td>
</tr>
<tr>
<td></td>
<td>field.</td>
<td></td>
</tr>
<tr>
<td>Figure 2-8.</td>
<td>Schematic of the experimental setup. BS: Beam splitter, PBS: Polarization</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>beam splitter, GDC: Grating-dispersion compensator, Tx: THz transmitter, Rx:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>THz receiver</td>
<td></td>
</tr>
<tr>
<td>Figure 3-1.</td>
<td>Structure of Mb (left) and heme group (right). The heme group is the pink</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>disk nested in the Mb molecule. There is a water molecule (denoted as 'W'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in the left figure) attached to the sixth liganding position of ferrous iron</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ion when there is no oxygen present. The six liganding position of heme</td>
<td></td>
</tr>
<tr>
<td></td>
<td>molecule can be seen in the right figure. Four of them are in the heme</td>
<td></td>
</tr>
<tr>
<td></td>
<td>plane and two others are perpendicular to the plane.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3-2. Measuring the water content of a hydrated Mb sample. The mass of the sample drops significantly upon heating due to the loss of water. When heated for a prolonged period of time, the sample mass will stabilize. When heat stops, the sample mass quickly rises due to absorption of water from the ambient atmosphere.................................................................34

Figure 3-3. Side view of the sample holder for Mb powder. The walls of the sample holder are two HDPE plates. There is a surface recess with a depth of either 1.00 mm or 0.50 mm in the center of one HDPE plate to hold the sample.................................................................37

Figure 3-4. Sample cell and adapter to hold the cell in place.........................39

Figure 3-5. THz time-domain waveform and its Fourier transform (inset). Solid line is the spectrum taken in dry nitrogen, and dotted line is the spectrum taken in ordinary laboratory air (~35% relative humidity). There are many oscillations on the time-domain waveform for humid air due to the vibrational and rotational motions of water molecules..............................................41

Figure 3-6. A photo of the BOMEM DA.3 rapid scan Michelson interferometer in Prof. Ramdas’s lab.................................................................42

Figure 3-7. Measurement scheme requires one measurement without sample, and another with sample (Medium 2) of length L. The sample holder (Medium 1 and 3) is made of HDPE.................................................................46

Figure 3-8. Procedure to determine the thickness of a HDPE plate. The total variation of the complex refractive index of the first order (top) and second order (bottom) is plotted as a function of guessed sample thickness and the real sample thickness is determined as the local minimum of the curve (3.075 mm). .............................................................................................................50

Figure 3-9. Procedure to determine the thickness of a HDPE plate. The local minimum of the total variation is at 3.055 mm. .................................................................51

Figure 3-10. Data collection and analysis for Mb powder at 19.0 wt% water content. (A) Comparison of time waveforms recorded without sample (solid line) and with Mb powder (dotted line) in the N₂-purged chamber; arrow denotes first Fabry-Perot reflection within the sample cell. (B) Fourier-transform frequency spectra of the two waveforms, plus their phase angle (inset). (C) Ratio of amplitudes versus frequency, and phase difference (inset). (D) Calculated absorption coefficient (mm⁻¹) before (dotted line) and after (solid line) correction for Fabry-Perot oscillations due to sample cell dimensions. Negative absorption coefficient at the lowest frequencies is most likely due to system uncertainties and data analysis artifacts (see text) .......55

Figure 3-11. Time-domain waveforms of liquid water sample with thickness of 100 (solid curve) and 200 μm (dashed curve).........................................................61

Figure 3-12. The effect of rectangular window on the time-domain waveforms of liquid water. To reduce the abruptness at the window edge due to truncation, some linear extrapolations are used to smooth the signal.........................62
Figure 3-13. Windowing effect on the time-domain waveforms of liquid water. Three types of window functions are used in the figure: Hamming window, Blackman-Harris window, and Flat Top window...

Figure 3-14. Three types of window functions used in Figure 3.13

Figure 3-15. Absorption spectra of liquid water obtained by applying different window functions on the time-domain waveform. The liquid water absorption spectrum from the publish results (black line) [125] is also included for a comparison.

Figure 3-16. Refractive index of liquid water obtained by applying different window functions on the time-domain waveform. The liquid water refractive index data from publish results (black line) [125] are also included for a comparison.

Figure 3-17. Fourier transform of different window functions used.

Figure 3-18. Absorption spectrum of liquid water with standard deviations shown as dashed lines below and above the average. The black line is the published results for comparison [125].

Figure 3-19. Refractive index of liquid water with standard deviations shown as dashed lines below and above the average. The black line is the published results for comparison [125].

Figure 3-20. Absorption spectrum of water vapor taken in room air (blue line) along with the simulated results (red line) [127] for comparison. Excellent agreement is obtained except for the relative amplitude of each peak.

Figure 3-21. The absorption spectra of Mb powders from 3.6 to 42 wt% water contents. The quadratic curve on each line is the fit based on data points from 0.1 to 1.2 THz where the signal-to-noise ratio is good. Error bars at 0.1, 0.6, and 1.2 THz are standard deviations derived from multiple measurements. Weak oscillations at low frequencies are likely residual Fabry-Perot oscillations from the sample cell, and structures at high frequencies are noises due to weak signals and are within the error bars.

Figure 3-22. The absorption spectra of Mb samples from 3.6 to 42 wt% water contents normalized to sample density, along with the quadratic fit. Error bars at 0.1, 0.6, and 1.2 THz are the corresponding error bars in Figure 3.21 divided by the sample density. These absorption spectra can be characterized as essentially smooth, continuous, and without sharp identifiable features.

Figure 3-23. The absorption spectra of Mb solutions at different water concentrations. The straight line is a linear fit for each sample based on the frequency range with good signal-to-noise ratio (0.3-0.9 THz). The error bars shown at 0.3, 0.6, and 0.9 THz are the standard deviations derived from multiple measurements.

Figure 3-24. The absorption spectra of Mb solutions normalized to sample density at different water concentrations. The straight line is a linear fit for each sample based on the frequency range with good signal-to-noise ratio (0.3-0.9 THz). The error bars shown at 0.3, 0.6, and 0.9 THz are the
corresponding error bars in Figure 3.23 divided by the sample density, based on the principle presented in Section 3.5.2.................

Figure 3-25. Comparison of absorption spectra from water vapor and a typical Mb sample (19.0 wt% water content). Water vapor has sharp, assignable lines that can be used to identify its presence, while biological water in protein has a smooth profile without identifiable characteristics..............

Figure 3-26. Comparison of absorption spectra obtained from THz and FTIR spectrometers, for an Mb powder with 32.0 wt% water content. FTIR displays Fabry-Perot oscillations and increased uncertainty at low frequency due to weak source intensity, and requires normalization by a factor of 1.18 to coincide with the THz spectrum......................................................

Figure 4-1. Schematic of spherical Mb molecule and the surrounding water layers. The diameter of the Mb sphere is approximated to be 3.0 nm and the outside water layer extends to about 0.35 nm.........................................................

Figure 4-2. Schematic of the local tetrahedral structure in liquid water in which one water molecule forms four hydrogen bonds with neighboring molecules (Picture from Ref. [152])...............................................................................

Figure 4-3. Normalized absorption spectra of Mb powders (top) and solutions (bottom) copied from Section 3.5. The absorption spectrum for Mb powder is fit by a quadratic curve, while the absorption spectrum for Mb solution is better fit by a straight line. The values on the fitting curve will be used as the basis for the discussions in the following sections, instead of using the original data................................................................................................

Figure 4-4. Normalized areal absorption coefficient of Mb-water mixtures as a function of water content at selected frequencies (0.35, 0.5, and 0.8 THz). The absorption coefficients used here for each sample are taken from the fitting curves presented in Figure 4.3, in order to average over any uncertainties in the measurements. The measured data points are denoted by open circles with error bar on each data point. The solid lines connecting the high and low water contents data points are used to guide the eyes. The gap left between 42 and 70 wt% samples indicates that we don't have data points for the intermediate water contents. The dotted straight lines are the results calculated based on an ideal model in which the Mb molecules and water molecules are assumed to be two non-interacting species. It can be seen that increased absorption is observed for many of the samples measured at all the selected frequencies, as compared with the ideal model. ........................................................................................................

Figure 4-5. Molar concentration of free water and biological water as a function of water content. It can be seen that there is no free water for water contents up to around 40 wt%. For Mb solutions, the free water molar concentration increases linearly with water content and the biological water molar concentration decreases linearly because of the decreasing amount of available Mb. .........................................................................................................
Figure 4-6. Molar absorptivity of Mb as a function of water content at frequencies of 0.35, 0.5, and 0.8 THz. The closed circles with error bars connected by solid line are the actual data points. The dashed line denotes the absorptivity of dry Mb as a base level. An enhanced molar absorption is seen throughout, with a dramatic increase at the dilute limit. ................................................. 113

Figure 4-7. Comparison between our results on solvated Mb absorption and the published results on BSA absorption [158]. Error bars on BSA results are copied from Ref. [158] and the error bars for our results are only shown on the 0.5 THz curve for a representative. The increase in BSA molar absorptivity above 95 wt% water is consistent with our results. However, the large experimental uncertainties on the data points could not exclude the possibility of no change in BSA molar absorptivity. ........................................ 116

Figure 4-8. Model-dependence investigation of the results. The blue line is the Mb molar absorptivity when taking all the protein-bound water as non-absorbing and the red line is the results when taking all the protein-bound water as normal free water. It can be seen that any models between these two extremes will yield results between the blue and the red lines, just as what we got for the black curve. The two trends on the Mb molar absorptivity are still clearly identifiable. ................................................................. 119

Absorption measurements were made of the heme protein myoglobin mixed with water from 1.2 to 98 wt% (weight percentage) in the frequency range 0.1-2.0 THz, using THz time-domain spectroscopy. It was found that the absorption is dominated by the water content, but even the driest specimens with hydration level below 4 wt% have a nearly continuous spectrum without identifiable sharp features. Inhomogeneous broadening plus the intrinsically high spectral density of vibrational modes in the region below 2.0 THz apparently combine to obscure the lowest frequency vibrational modes expected for protein molecules of this size. A continuous absorption spectrum for hydrated protein samples suggests that the absorption mechanisms are similar to those in liquid water, and hinders the spectroscopic identification of biomolecules in this frequency range.

The interaction of proteins with an aqueous environment leads to a thin region of "biological water" whose molecules have properties that differ from bulk water, in particular reduced absorption of far-infrared radiation caused by protein-induced perturbation of the water dipole moment. Based on the myoglobin far-infrared absorption measurements, the effect of biological water on myoglobin is carefully studied. Measurements show that absorption per protein molecule is increased by the presence of biological water. Analysis shows greater THz absorption when compared to a non-interacting protein-water model. Including the suppressed absorption of biological water leads to a substantial hydration-dependent
increase in absorption per protein molecule over a wide range of concentration and frequencies, meaning that water increases the protein's polarizability.
CHAPTER 1. INTRODUCTION

This work deals with an important protein molecule, myoglobin (Mb). In particular, the far-infrared absorption spectra of Mb powders and aqueous solutions are systematically measured and the role of water molecules associated with Mb is carefully studied.

Mb is the primary oxygen-carrying protein of muscle tissues [1] that gives muscle its red color. It is among the first protein molecules whose structure was well-studied [2]. Mb is even called “the hydrogen atom of biology” for its relative simplicity and biological importance [3] and often serves as a prototype for biologically complex systems. There have been extensive studies on Mb in the past 50 years or so. Among these is the study of interactions between Mb and Mb-associated water molecules, which has long been an attractive research topic [4-7].

Various methods have been utilized to clarify protein hydration processes and to investigate protein-water interactions, such as dielectric spectroscopy [8-11], NMR spectroscopy [12-16], and X-ray scattering and neutron scattering techniques [17-22]. With the development of ultrafast optics, a new tool named terahertz time-domain spectroscopy (THz-TDS) was recently introduced into this area [23]. THz-TDS, as a new spectroscopic technique working in the far-infrared region, compares favorably to the traditional Fourier transform infrared (FTIR) spectroscopy due to its single-cycle pulse generation and coherent, time-gated detection which reduce the thermal background effectively [24]. Furthermore, THz spectroscopy should be particularly appropriate in studying the hydration
effect on biomolecules because that THz radiation is very sensitive to the sample water content due to the large dipole moment for water molecules [25].

In this thesis, we will present a complete THz-TDS study on the far-infrared absorption spectra of Mb-water system as a function of hydration level. The samples measured cover the hydration levels from 1.2 to 98 wt% (water weight percentage), including Mb powders and solutions. The frequencies studied cover the range from 0.1 to 2.0 THz. The interaction between Mb molecules and the associated water molecules, so-called biological water, was carefully analyzed based on the far-infrared absorption properties of the Mb-water system.

The results of this PhD work are mainly in two aspects:

1. Extract the intrinsic far-infrared absorption spectrum of myoglobin in hydrated and dehydrated powders and myoglobin solutions

The far-infrared absorption properties of Mb powders and solutions are studied at a series of carefully controlled water contents. It is found that the absorption spectra of the samples under study are smooth throughout the available frequency range, without any identifiable sharp lines which can be assigned to the intrinsic characteristics of the protein.

2. Isolate the role of biological water in myoglobin powders and solutions

The effects of protein-associated water molecules on the Mb far-infrared absorption spectrum are studied. With the measurements of the Mb absorption spectrum at gradually increasing water content, we can learn how the biological water in protein species like Mb will influence its far-infrared absorption properties and the role of hydration water molecules in large biomolecules.
The subsequent chapters in this thesis are organized as follows:

Chapter 2 will focus on the basic principles of THz-TDS. The generation and detection mechanisms of THz radiation will be introduced, especially those used in this thesis work. The experimental instrumentation for THz-TDS will be discussed in detail. Some of the popular applications of THz spectroscopy will be described at the end.

Chapter 3 will concentrate on the main experimental results of this PhD work. The method to prepare Mb powder and solution samples is described. The issues related with experiments and measurements are detailed. In THz experiments, one important concern in the data analysis is the material parameters extraction. Multiple reflections on the interface between sample and sample cell were commonly observed, interfering with the primary transmitted beam and complicating parameter extraction [26]. A reliable algorithm to extract the material parameters is crucial for the success of the experiment. The algorithms proposed by L. Duvillaret [26] and T. D. Dorney [27] are introduced in Chapter 3 and applied in the experimental data analysis. Finally, the experimental results are presented and discussed. The absorption spectra of Mb samples are measured at a series of carefully controlled hydration levels ranging from 1.2 to 98 wt%. It is shown that the absorption is dominated by water content, for Mb powders as well as solutions. A nearly continuous absorption spectrum with increasing frequency was observed that proved to be devoid of any obvious features characteristic of Mb.

Chapter 4 will dig deeper into the scientific insights of this PhD work. The Mb absorption properties are interpreted in terms of absorptivity per Mb molecule. Two models are constructed to understand the hydration effects of biological water in Mb absorption, in which the Mb and water molecules are first taken as non-interacting species and then their interactions are considered. Some
interesting and important scientific implications are presented, in particular the plasticizing effect of hydrated water on the protein structure. For a meaningful discussion of the hydration effects on the Mb molecule, a deep understanding of THz absorption physics of liquid water is necessary. So the basic physics of liquid water absorption in the THz region is also introduced in Chapter 4.

The whole work will be summarized in Chapter 5.
CHAPTER 2. TERAHERTZ TIME-DOMAIN SPECTROSCOPY

In the electromagnetic spectrum, the intermediate infrared band has wavelengths as long as 10 μm (30 THz), and the microwave band has frequencies as high as 30 GHz (10 mm). For lack of effective sources and sensitive detectors, the region between these two (0.03-30 THz), which is called terahertz (THz) or far-infrared band (or sub-millimeter wave), remained a gap for a long time (Figure 2.1) [28]. The pioneering work done by D. H. Auston and his coworkers [29-34] led to the advent of THz time-domain spectroscopy (THz-TDS) in the 1980s [35-37] which bridged the THz gap.

Figure 2-1. Terahertz gap in the electromagnetic spectrum (From Ref. [28])

The advent of THz-TDS owed to the development of ultrafast lasers, semiconductor technologies, and nonlinear optics. In general, the generation and detection of THz radiation relies on either transient photoconductivity in semiconductor materials or optical rectification in electro-optic crystals, both with an ultrafast laser as a pumping source.
In this Chapter, we will focus on the generation and detection mechanisms of THz radiation utilizing transient photoconductivity, since this is the method used in our spectrometer. Other generation and detection methods such as optical rectification and difference frequency photomixing will also be briefly touched upon. The last part of this chapter will be on the details of the THz spectrometer used in this work, and common applications of THz-TDS.

2.1. Generation of Terahertz Radiation

The two major mechanisms used to generate THz radiation are transient photoconductivity in semiconductors with short majority carrier lifetime, and optical rectification in electro-optic crystals with large second order nonlinearity. While attention will be focused on the transient photoconductivity method, other mechanisms including the optical rectification will also be discussed briefly.

2.1.1. Transient Photoconductivity Method

The transient photoconductivity method for the generation of pulsed THz radiation is based on the photoconductive (PC) antenna [30, 33, 34], which is also commonly known as an “Auston switch”, named after the pioneer in this area [38]. The schematic diagram of a PC antenna is shown in Figure 2.2, viewed from the side and the top. The physical properties of the THz emitter, so-called THz transmitter (Tx) in our spectrometer, can be understood based on this PC antenna.
The main structure of a PC antenna is a photoconductor, which consists of two metal strips deposited on a semiconductor substrate. The extensively used semiconductor substrate nowadays is the low-temperature grown GaAs (LT-GaAs) for its short carrier life time and relatively high carrier mobility. The metal strips are usually made of gold for its good conductivity. There is a small gap between the two metal strips. This gap is the active region where the femtosecond optical pulses are focused.

Physically, the PC THz emitter behaves like a Hertzian dipole antenna [39, 40]. When the PC gap is pumped by high-intensity femtosecond optical pulses with energy greater than the semiconductor bandgap, for example, 1.43 eV for GaAs at room temperature, photocarriers will be generated within the semiconductor.
The acceleration of photocarriers under an external DC bias field and the following recombination will result in a pulsed photocurrent in the PC antenna. This transient current through the switch will first rise very rapidly upon generation of photocarriers, and then decay with a time constant given by the carrier life time of the semiconductor. The photocurrent signal produces a time-varying dipole moment \( p(t) \) whose radiation field, \( E(r, \theta, t) \) at a distance \( r \), and angle \( \theta \) relative to the dipole axis, can be estimated by the classical field of an elementary Hertzian dipole:

\[
E(r, \theta, t) = \frac{1}{4\pi\varepsilon_0 n^2} \left[ \frac{p}{r^3} + \frac{n}{cr^2} \frac{\partial p}{\partial t} + \frac{n^2}{c^2 r} \frac{\partial^2 p}{\partial t^2} \right] \sin \theta,
\]  

(2.1)

where \( c \) is the speed of light in vacuum, \( n = \sqrt{\varepsilon_0/\varepsilon} \) is the refractive index of the semiconductor substrate with \( \varepsilon \) and \( \varepsilon_0 \) being the permittivity of semiconductor and vacuum, respectively. The dipole moment \( p(t) \) is related to the photocurrent, \( i(t) \), by the expression

\[
i(t) = \frac{1}{l} \frac{\partial p}{\partial t},
\]  

(2.2)

where \( l \) is the effective length of the dipole antenna, which is assumed to be small relative to the shortest radiated wavelength.

There are three terms in the expression for the radiated electric field Eqn. (2.1), which correspond to the quasistatic, near and far field components, varying respectively as \( r^{-3} \), \( r^{-2} \) and \( r^{-1} \). The far field term \( (1/4\pi\varepsilon_0 c^2 r)(\partial^2 p/\partial t^2) \) is the THz field we are interested in, which will dominate as we go further from the source. The radiated field is proportional to the second time derivative of \( p(t) \) and hence has a time variation equal to the first time derivative of the transient photocurrent.
Since the current modulation occurs in the subpicosecond regime, the radiated field is also in the subpicosecond regime, i.e., THz pulse. The distinction between the far and near fields is given by the simple relation $r \gg \tau_p c / n$, where $\tau_p$ is the pulse duration [30]. If $\tau_p$ is extremely short, as in the case of femtosecond pulses, the distance for the far field to occur is also short.

The typical temporal behavior of the photocurrent and the associated radiated THz far field when the PC antenna is pumped by an optical pulse can be simulated numerically. A figure taken from Ref. [41] is shown (Figure 2.3) to illustrate this behavior. The incident optical pulse is assumed to have a Gaussian temporal profile. It can be seen that the emitter photocurrent has a fast rising edge and a slowly decaying tail. The radiated THz field is proportional to the time derivative of the emitter photocurrent. Hence a short unipolar current pulse will radiate a bipolar THz far field.

It is illustrated in Figure 2.2 left (side view) that the generated THz pulse is emitted from the substrate side, instead of from the gold strips side. This is based on the antenna theory that a dipole antenna on the surface of dielectric material emits roughly $(\varepsilon/\varepsilon_0)^{3/2}/2$ times (about 20 times for typical semiconductor substrate) more power to the dielectric material than directly to the air [39]. To couple the THz radiation into free space, a hemispherical lens is fabricated on the back side of the LT-GaAs substrate. The lens is usually made of high-resistivity silicon for good index matching and a better transmission.
Figure 2-3. Calculated photocurrent in the emitter and amplitude of the radiated THz field versus time. The incident optical pulse is assumed to have a Gaussian temporal profile. (Figure taken from Ref. [41]. The circles are data extracted from Ref. [42].)
2.1.2. Optical Rectification Method

THz pulse emission due to optical rectification is caused by second-order optical nonlinearity. It was first reported by M. Bass et al in 1962 [43] and was considered as a form of Cherenkov radiation [29, 31, 44]. The difference is that the radiation in this case is produced by electromagnetic beams (photons) rather than by moving charged particles.

Optical rectification was also studied as a difference-frequency generation process [45, 46], in which the frequency difference is close to zero (quasi-DC field). As shown schematically in Figure 2.4 below, typically, visible or near-infrared femtosecond pulses are focused on an electro-optic material with second-order nonlinearity $\chi^{(2)}$. A femtosecond optical pulse contains within it various spectral components and any two frequencies can contribute to the difference-frequency generation by beating with each other. A weighted sum of all these contributions will result in a broad radiation extending from DC to far- and mid-infrared.

Figure 2-4. Generation of THz pulses by optical rectification in a nonlinear material. The induced THz field is proportional to the second-order time derivative of the incident optical intensity. See explanations in the text.
Assume the incident femtosecond pulse $E(t)$ is a plane wave with complex frequency spectrum $E(\omega)$, the following equation gives the second-order nonlinear polarization in the frequency domain within an electro-optic material

$$P_{OR}^{(2)}(\omega) = \varepsilon_0 \chi^{(2)} \int E(\omega_1)E^*(\omega_1 - \omega)d\omega_1.$$  \hspace{1cm} (2.3)

The Fourier transform of Eqn. (2.3) gives the polarization field in the time domain

$$P_{OR}^{(2)}(t) = \varepsilon_0 \chi^{(2)} E(t)E^*(t) \approx \chi^{(2)} I(t).$$  \hspace{1cm} (2.4)

This equation shows that the pulse width of the generated radiation depends on that of the pump beam.

From Eqn. (2.1), we know that the far-field radiated electric field is proportional to the second-order time derivative of the induced dipole moment $p(t)$. Since the polarization field $P_{OR}^{(2)}(t)$ is the total dipole moment per unit volume, we will have the same relationship between the far-field radiated electric field and the optical rectification polarization field, given by

$$E_{\text{THz}}(t) \propto \frac{\partial^2 P_{OR}^{(2)}(t)}{\partial t^2}.$$  \hspace{1cm} (2.5)

Given the crystal structure and sufficient information about the incident pulses, Eqn. (2.5) can be used to calculate the far-field waveform of the radiation. However, in practice, many factors such as the crystal orientation, its thickness, its absorption and dispersion, diffraction effects, and saturation effects will affect the radiation efficiency, the temporal waveform shape, and the frequency distribution of the emitted THz radiation [47].
2.1.3. Other THz Generation Mechanisms

In addition to transient photoconductivity and optical rectification, the generation of THz radiation can also be achieved by other techniques, such as continuous wave (CW) THz radiation generation by photomixing, the free electron laser (FEL), and the THz quantum cascade lasers (QCL).

Both transient photoconductivity and optical rectification discussed in Sections 2.1.1 and 2.2.2 respectively rely on ultrafast optical pulses as a pumping source to generate THz pulses. Instead of using ultrafast pulsed lasers, it is possible to use two CW lasers to generate coherent CW THz radiation, via photomixing [48, 49]. The key component in this technique is a photomixer, which is a compact, all-solid-state source that uses a pair of single-frequency tunable lasers to generate a THz difference frequency by photoconductive mixing in LT-GaAs [48, 50]. The output frequency is tuned over several THz by fixing one laser and detuning the other by a few nanometers in wavelength.

An FEL is a large scale device requiring a high energy electron beam to operate. It is capable of generating tunable, coherent and high power radiation over a large part of the electromagnetic spectrum, ranging from millimeter waves up to potentially X-rays. For example, the FEL at University of California, Santa Barbara covers the frequency range from 120 GHz to 4.8 THz and there has been a lot of work done at this facility [51-53].

QCLs are based on semiconductor quantum lasers, which are man-made quantum mechanical systems in which the energy levels can be designed and engineered to be at any chosen values [54]. The mid-infrared QCL (operating at 4.2 μm wavelength) was first developed at Bell Laboratories in 1994 [55]. The development of THz QCLs was much more difficult. The first QCL operating below the reststrahlen band at 4.4 THz was developed by R. Kohler et al in 2001.
[56], which was based on a chirped superlattice structure that had been successfully developed at mid-infrared frequencies. The QCLs offer new possibilities for chemical and biological sensing and imaging applications.

2.2. Detection of Terahertz Radiation

It was shown that the THz generation principles presented in Section 2.1 could also be used for THz detection. Two major methods for THz detection, photoconductive detection and electro-optic sampling, will be discussed in this section, along with the comparisons between these two methods.

2.2.1. Photoconductive Detection

The same mechanism used in photoconductive THz generation can be used for THz detection as well [30, 33, 57]. As shown in the figure below (Figure 2.5), the photoconductive detection setup is very similar to the generation setup shown in Figure 2.1. The only change here is that the external DC bias is replaced by an amperemeter connected with a lock-in amplifier to measure the transient current occurring in the PC dipole antenna induced by the incoming THz pulses.

In the pump-probe scheme shown in Figure 2.5, the antenna is gated by a probe optical pulse that generates a bunch of photocarriers in the photoconductor. The electric field of the incoming THz pulse is polarized across the antenna and serves as the voltage bias. The carrier lifetime $\tau$ of the PC substrate is typically much shorter than the THz pulse so that the antenna will act as a sampling gate which samples the THz field within a time $\tau$. The principle of photoconductive sampling is illustrated in Figure 2.6.
When the optical pulse is present, the generated photocarriers are driven by the THz field and form a photocurrent through the gap between the antenna’s leads. The photocurrent measured at the detector, proportional to the electric field of the focused THz radiation, is amplified with a low-noise current amplifier and is fed to a lock-in amplifier. The THz waveform is obtained by measuring the average photocurrent versus time delay between the THz pulses and the gating optical pulses. A Fourier analysis of the temporal profile of the received THz pulse reveals the amplitude and phase spectrum.

Figure 2-5. Detection of THz pulses by photoconductive antenna. The photocarriers generated in the GaAs substrate by the gated optical pulse are driven by the incoming THz pulse and result in a transient current whose amplitude is proportional to the THz field.
Figure 2-6. Principle of photoconductive sampling. The photoconductive antenna acts as a sampling gate that detects the THz signal when the optical pulse is present. The detected signal is actually a convolution between THz pulse and optical pulse. By changing the delay between the optical pulse and THz pulse, the entire THz waveform can be mapped out sequentially in time.

2.2.2. Electro-optic Sampling

Developed originally for the local field characterization of ultrafast electrical transient [58, 59], the electro-optic sampling has developed into a powerful method for detecting THz pulses in free space [60-62]. A typical experimental setup for the free-space detection of THz pulses using electro-optic sampling is shown in Figure 2.7 [63].

The principle of electro-optic sampling is based on Pockels effect (first order non-linearity) in an electro-optic crystal. The electric field of a THz pulse will induce a small birefringence in the crystal through a nonlinearity of the first order. Initially, the ultrafast probe beam is linearly polarized. After passing through the crystal, the probe beam will gain a small elliptical polarization. In the first order approximation, this ellipticity is proportional to the instantaneous THz electric field.
applied to the crystal. Since the THz pulse is much longer than the laser pulse, the THz field can be approximately treated as a DC bias field. Thus, by varying the delay between THz and optical probe pulses, the whole time profile of the pulse can be mapped.

The nonlinear crystals normally used for electro-optic sampling are LiTaO$_3$ [60], ZnTe [62], and poled polymer [61]. ZnTe was shown to yield the best performance and thus is most commonly used.

![Figure 2-7. Principle of electro-optic sampling (From Ref. [63]). The birefringence induced by the THz pulse in the electro-optic crystal will be seen by the optical sampling pulse. The incoming circularly polarized beam will become elliptically polarized after passing the crystal. The phase retardation between the two polarization components is proportional to the THz electric field.](image)

### 2.2.3. Comparisons between Photoconductive Detection and Electro-optic Sampling

Before the development of electro-optic sampling, the PC antenna was the primary technique for THz pulse detection. Due to its intrinsic gated detection nature, the PC antenna has superior detection sensitivity and a high signal-to-noise ratio. Furthermore, the detection bandwidth of a PC antenna with a short
dipole length can exceed 5 THz [60]. However, the measured waveform is not a simple cross-correlation of the incoming THz and optical gating pulses, but a convolution of the laser pulse envelope, the response function of the antenna, and the THz pulse itself. Thus, the detection bandwidth of a PC antenna is limited by the carrier lifetimes of the photoconductive materials and the antenna geometry. Even if the temporal resolution of the PC antenna is reduced below 100 fs, the measured signal will still not represent the true THz waveform.

As an alternative method for THz pulses characterization, free-space electro-optic sampling is an instantaneous technique and can provide an exact cross-correlation measurement of the incident THz and optical pulses. Assuming the spectrum of the THz pulses lies below the first phonon resonance of the sensor crystal (for example, 8.06 THz for GaAs, 5.31 THz for ZnTe [62], and 6 THz for LiTaO₃ [32]), electro-optic sampling offers a flat frequency response over an ultrawide bandwidth, and overcomes the limitations of photoconductive antennas that result from their resonant dipole structure and the effects of finite photocarrier lifetime. Furthermore, since electro-optic sampling is purely a free-space optical technique, it does not require electrode contact or wiring on the sensor crystal. However, to take full advantage of free-space electro-optic sampling, two major problems must be addressed: (1) the velocity-matching condition between the THz and optical pulses; and (2) the multireflections of the THz and optical pulses within the sensor crystal [60].

For a detailed comparison between the performances of PC antenna and electro-optic sampling, readers are referred to the work by Y. Cai et al [64] and S. G. Park et al [65].
2.3. Terahertz Time-Domain Spectroscopy

The advent of optoelectronic THz emitters and receivers has enabled the successful development of THz-TDS [66, 67]. THz-TDS works in two major configurations: transmission and reflection. The transmission configuration is the most commonly used and is more reliable, but it can only work well when the samples are transparent or at least semi-transparent in the THz frequency range. In the cases when: (i) the samples are strongly absorbing or reflective, (ii) the sample thickness is involved which complicates the material parameters extraction, the reflection configuration is more favorable [68-72]. However, the usage of reflection configuration is limited because it is very sensitive to the sample positioning and apparatus stability. THz-TDS can also be used in emission spectrometers [73, 74] and time-resolved optical-pump THz-probe experiments [75-80]. In this section, we will mainly concentrate on the transmission configuration.

A schematic of the experimental setup used for this thesis research is shown in Figure 2.8. The main components of the setup are a home-made Ti:Sapphire femtosecond laser and a commercial fiber-pigtailed THz spectroscopic system purchased from Picometrix, Inc. [81].
Figure 2-8. Schematic of the experimental setup. BS: Beam splitter, PBS: Polarization beam splitter, GDC: Grating-dispersion compensator, Tx: THz transmitter, Rx: THz receiver.
The Ti:Sapphire femtosecond laser, pumped by an argon ion laser purchased from Spectra-Physics, emits laser pulses with ~100 fs pulse width and 800 nm central wavelength, as required by the THz system. A beam splitter is integrated in the THz system to separate the laser power into two portions: one is used to pump the THz transmitter (Tx) and another (passing a delay stage) to probe the THz receiver (Rx). PC antenna fabricated on LT-GaAs substrate is used for both the transmitter and receiver. The detected current is converted to voltage by a lock-in amplifier and then sent to a control box for analysis and is displayed on a computer.

Because fibers are used to deliver the femtosecond pulses to the transmitter and receiver in the THz spectrometer, there will be positive group velocity dispersion (GVD) imparted onto the pulses as they travel along the fiber. A grating-dispersion compensator (GDC) is applied before the laser pulses are coupled into the fiber, which imparts a compensating negative GVD onto the pulse. There are two micrometers on the GDC box, which can be tuned to obtain an optimal compensation effect. The optimal dispersion compensation will show up as an optimal bandwidth in the detected THz spectrum.

In the actual experiments, back-reflections from the fiber may perturb the stability of the Ti:Sapphire laser. An optical isolator is used to prevent these back-reflections. For real time diagnostics of the laser performance, small fractions of the laser power are split off to a spectrometer and an autocorrelator to monitor the spectrum and pulse width of the laser, respectively. Since the THz spectrometer requires that the pump laser is horizontally polarized and there is a maximum power level for the GDC input, a polarization beam splitter, together with a quarter wave plate, is used to tune the polarization state and input power.

THz radiation is emitted from the transmitter and coupled into free space by an integrated hyper-hemispherical silicon lens [82]. Appropriate lenses (high-density
polyethylene lenses or silicon lenses) are used to manipulate the THz radiation in free space. The studied sample is positioned at the focal point of the lenses. Nitrogen gas boil-off from a liquid nitrogen tank is used to purge the THz spectroscopy system to eliminate absorption from ambient water vapor.

2.4. Applications of Terahertz Time-domain Spectroscopy

THz-TDS offers tremendous advantages relative to other far-infrared spectroscopic techniques in terms of sensitivity, dynamic range, and direct measurement of both real and imaginary parts of the dielectric function [66]. Compared with its conventional counterpart, FTIR spectroscopy [24], some unique features of THz-TDS make it a very suitable tool for various applications:

1. A wide bandwidth (0.03-30 THz) enables spectroscopic investigations of molecular rotational and vibrational motions and studies of carrier dynamics in semiconductors, superconductors, and dielectrics.

2. Femtosecond time-gated detection of the far-infrared electric field, with time resolution <100 fs, enables spectral information over the entire bandwidth to be determined since the spectral bandwidth is inversely proportional to the time resolution.

3. The electric field is measured in THz-TDS, as opposed to intensity, which gives rise to both amplitude and phase information in the frequency domain. This enables direct and simultaneous determination of the refractive index and the absorption coefficient.

4. The detection scheme is a laser-gated coherent technique, and as such is blind to most background emission such as blackbody radiation.
Because of all these advantages, this technique has been used successfully in many research fields. In the remainder of this chapter, we will briefly review some of the important applications.

2.4.1. Material Properties Characterization

Due to the high sensitivity and the ability to obtain real and imaginary parts of the complex dielectric function simultaneously, THz-TDS has become a very powerful tool in characterizing the material properties in the far-infrared range, particularly of lightweight molecules [83, 84] and semiconductors such as GaAs and silicon wafers [85]. Water molecules in the vapor phase were among the first to be studied by THz-TDS [83]. THz-TDS was also used to study the coherent transient effects in \( \text{N}_2\text{O} \) and methyl chloride vapors subsequent to excitation by THz pulses [86].

High-temperature superconductor characterization is another important application of THz-TDS. Superconducting thin films have been analyzed to determine various material parameters [87]. THz-TDS has recently been used to study \( \text{MgB}_2 \), the material newly discovered to be superconducting at the surprisingly high transition temperature of 39 K and is not currently well understood [88].

Experiments with optical-pump THz-probe system can reveal additional information about materials. In these experiments, the materials are excited using ultrafast optical pulses, and THz pulses are used to probe the dynamic far-infrared optical properties of the excited materials [89].
2.4.2. Biological Identification

The collective vibrational and rotational motions of biological molecules like proteins and DNA are predicted to occur in the THz range (1 THz = 4.1 meV). Many biological compounds, especially smaller molecules, show very strong, highly specific frequency-dependent absorption and dispersion in this range. By studying the far-infrared absorption properties of these biological samples, it is feasible to identify them depending on their characteristic features [90-96]. THz-TDS has also shown ability to infer information on the conformational states of biomolecules [97, 98]. The complex refractive indices of DNA and other biomolecules have been determined and show absorption consistent with a large density of low-frequency infrared-active modes [99, 100].

In biomedical research, the identification of polynucleotides with unknown base sequences usually requires gene chips composed of fluorescently labeled polynucleotides with known base sequences. Fluorescent labeling can affect diagnostic accuracy and increase the cost and preparation time of gene chips. THz-TDS offers an approach for label-free genetic diagnostics. It was shown that THz-TDS is capable of differentiating single- and double-stranded DNA owing to the associated changes in refractive index [101, 102]. It may be possible to design a marker-free THz biochip for gene sensing based on this method [103, 104].

2.4.3. THz Sensing and Imaging

Different materials show different absorption and dispersion properties in the THz range. When a THz radiation propagates through a material, the time-domain waveforms will carry characteristic information of the material under study. This makes highly sensitive imaging of material composition possible.
Pulsed THz imaging was first demonstrated in 1995 [25], and since then has been used in a wide range of applications such as chemical composition detection [105], industrial quality inspection [106], and biomedical diagnostics [107]. The attraction of THz imaging is largely due to the availability of phase-sensitive spectroscopic images, which holds the potential for material identification [89].

Based on the characteristic features of small molecules in gas phase, THz-TDS can also be used in gas sensing [108]. In some cases where high temperature is involved, THz-TDS is actually the only applicable way to detect the gas composition [109, 110] due to its immunity to thermal background.

2.4.4. Broadband Communication

In the communication community, there is always a demand for a broader frequency band for signal communication and data transfer. While the current communication system is pushing up to the GHz bandwidth range, there may be a need for communications in THz frequencies someday.

There are indeed some benefits offered by wireless communications operating in the THz region, particularly for high bandwidth, short path, and line of sight wireless links [111]. However, the feasibility of THz communications is driven by commercial interests rather than technological issues.

For THz wireless communication in free space, one hindrance is that the atmosphere strongly attenuates THz energy. It is possible to find some atmospheric windows in which the attenuation is weak, for example, at around 400 GHz. One advantage for wireless communication in this range is that smaller
transmitter power is required for communicating the same distance compared with microwaves [111]. Because of the highly directional nature of THz propagation and the inability of THz radiation to pass through buildings, it may find applications in very high data rate transfers over short distances in a multipoint to point/multipoint basis.

It is also possible that THz frequencies may be used in wave guiding communications providing advances in compact sources, effective wave guiding components, and sensitive detectors. For real world communication, many other devices such as amplifiers, modulators, multiplexers, and isolators would need to be developed. All these components are still under development.
CHAPTER 3. FAR-INFRARED ABSORPTION PROPERTIES OF MYOGLOBIN IN POWDERS AND IN SOLUTIONS

This chapter will concentrate on the main experimental results of this thesis work. It starts with a brief introduction to the structure and function of Mb, followed by a detailed description of the Mb powder and solution samples preparation procedure. Then some important issues in the experiments and with data analysis are documented, including the material parameters extraction algorithms. Finally, the experimental results on the far-infrared absorption spectra of Mb powders and solutions at carefully controlled hydration levels ranging from 1.2 to 98 wt% (weight percentage) are presented and discussed, followed by a brief conclusion.

3.1. Introduction to Myoglobin Structure and Function

Mb is an iron-containing single-chain protein that gives muscle its red color, due to the presence of an optically active heme molecule. It is a relatively simple protein molecule whose 3D structure was among the first to be mapped by X-ray diffraction [2]. The Mb polypeptide chain consists of 153 amino acids and is folded to form a cradle (4.4×4.4×2.5 nm) that nestles the heme prosthetic group [112]. Figure 3.1 (left) is a schematic of its structure. The pink disk shown in the figure is the heme group. As shown in Figure 3.1 (right), heme is a porphyrin molecule containing four pyrrole rings linked together by methenyl bridges in a plane with a ferrous iron ion (Fe$^{2+}$) in the center. Four of the six liganding positions of the iron ion are connected with nitrogen atoms in the pyrroles. The
fifth and sixth ligands lie above and below the heme plane. The fifth liganding position is attached to the imidazole side chain of amino acid residue histidine F8. When Mb binds O₂ to become oxymyoglobin, the O₂ molecule adds to the heme iron ion as the sixth ligand. On the oxygen-binding side of the heme lies another histidine residue, His E7. While its imidazole function lies too far away to interact with the Fe atom, it is close enough to contact the O₂. Therefore, the O₂-binding site is a sterically hindered region. Biologically important properties are believed to stem from this hindrance.

Although it is well-known that the main function of Mb is storage of O₂, the mechanism for binding is not fully understood. It is believed that the collective vibrational motions of the molecule, which occur in the far-infrared frequency range and can lead to conformational transformation of the molecule, play an important role in the functional mechanisms of biological macromolecules [113, 114]. Due to its relative simplicity and biological importance, Mb serves as a prototype for biologically complex systems and is called “the hydrogen atom of biology” [3].

Despite the physiological importance of Mb, there are few studies on its far-infrared absorption properties [98]. This may be partly due to the lack of effective far-infrared radiation sources. The invention of THz-TDS provided a powerful technique for the far-infrared absorption studies of Mb [23]. In the sections to follow in this chapter, the experimental results of absorption spectra for Mb powders and solutions at hydration levels ranging from 1.2 to 98 wt% are presented and the Mb absorption properties discussed.
Figure 3-1. Structure of Mb (left)\(^1\) and heme group (right)\(^2\). The heme group is the pink disk nested in the Mb molecule. There is a water molecule (denoted as 'W' in the left figure) attached to the sixth liganding position of ferrous iron ion when there is no oxygen present. The six liganding position of heme molecule can be seen in the right figure. Four of them are in the heme plane and two others are perpendicular to the plane.

\(^1\) Picture is taken from the internet: 
http://library.tedankara.k12.tr/chemistry/vol5/vol5.htm

\(^2\) Picture is taken from the internet: 
http://www.kiriya-chem.co.jp/q&a/image/heme-o2-eq.gif
3.2. Sample Preparation

All the samples are prepared out of horse heart Mb lyophilized powder purchased from Sigma (Lot No. 122K7057) and used without further purification. Absorption spectra for both Mb powders and solutions at various hydration levels are measured. The preparation methods for these two kinds of samples are different. For the solutions, the mass of deionized water and Mb lyophilized powder are calculated based on the pre-determined hydration level. For the powders, the samples are prepared first and the hydration levels are determined afterwards. The methods for preparing powder and aqueous solution samples are described in detail below.

3.2.1. Myoglobin Powder Samples

The Mb powders at various hydration levels were obtained by allowing as-received lyophilized powders to equilibrate with different saturated salt solutions in a desiccator over several days. Different saturated salt solutions have different equilibrium relative humidities (ERH), as indicated in Table 3.1 [21]. The sample hydration level is determined by the ERH of the solution and the preparation time.

Sample water content can be expressed in different notations. Throughout this thesis, the weight percentage of water notation (wt%) is used, which is defined in the following simple equation:

\[
\text{wt}\% = \frac{m_{H_2O}}{m_{H_2O} + m_{Mb}} \times 100\%.
\]

where \(m_{H_2O}\) and \(m_{Mb}\) are the masses of water and Mb in the mixture,
respectively. The masses can be measured with an accuracy of ±1% relative error so that the relative uncertainty in the water content will be around ±2%. Two other common notations, the number of water molecules per Mb molecule and grams of water per gram of Mb, are also included in Table 3.1 for comparison.

One may notice in Table 3.1 that the final water content doesn't always scale with the ERH of the saturated solution. That is to say, higher ERH doesn't always yield higher water content. This is because the sample preparation time is not identical. To explain this clearly, another fact is worth noting. It was observed that laboratory air typically has relative humidity around 40%, and if equilibrated with this environment, the as-received Mb sample will retain its original water content at around 10 wt%. Any relative humidity much higher than 40% (e.g. 80% for KCl solution) will result in a hydrated Mb sample with water content higher than 10 wt% and lower relative humidity (e.g. 18% for NaOH solution) will yield a dehydrated Mb sample with water content below 10 wt%. If the ERH of the saturated solution is higher than 40%, the longer the preparation time, the higher the Mb sample water content. If the ERH of the saturated solution is lower than 40%, the longer the preparation time, the lower the Mb sample water content. This will explain the discrepancy in Table 3.1 concerning the solution ERH and Mb sample water content. Fortunately, what we care about is the final water content of the sample, not the relationship between ERH and water content. As long as the water content measurements are accurate, we can properly interpret the results.

To measure the water content of the powders, two samples were prepared side by side for each hydration level. After preparation, one sample was placed in a sand-bath (a beaker with a layer of sand at the bottom) and heated to just above 100 °C for one hour. All the water in the sample was assumed to be vaporized in this process. The sample was weighed before and after heating. The weight loss
is the initial mass of water in the specimen. The measurement procedure was performed in room air with relative humidity around 40%. Data were recorded as quickly as possible (within 30 sec) to prevent the sample from rehydration by the ambient humidity. One example of these measurements is demonstrated in Figure 3.2. The sample mass drops significantly during the first 5-10 minutes of heating and then remains at a stable value for the rest of the time, indicating all the water is vaporized. When heating stops, the sample mass increases rapidly until it reaches the value close to that before heating, indicating that all the water is recovered by absorbing water vapor from air. In Figure 3.2 as well as Table 3.1, the uncertainties in the mass measurements are ±1%.

With this method, Mb powders with water content ranging from 1.2 to 42 wt% were prepared (Table 3.1). We failed to obtain samples with water content greater than 42 wt%. At that concentration, the Mb powder had converted into a dense liquid. Earlier measurements of dielectric response [115] deduced a maximum level of 0.6 grams of water per gram of Mb, corresponding to 37.5 wt% water and in reasonable agreement with our 42 wt% result.
Table 3-1. Hydration result for Mb powders. Each specimen was stored for 3-4 days in a sealed desiccator with one saturated salt solution with a specific equilibrium relative humidity. The water content can be expressed in three related notations. Some notes for the table: (i) The “as-received sample” means the Mb lyophilized powder was not treated by any of the solutions. (ii) The “desiccant” means the sample was placed in a desiccator with desiccants to produce a dry environment.

<table>
<thead>
<tr>
<th>Solution/Material used for the sample preparation</th>
<th>Equilibrium Relative Humidity (ERH%)</th>
<th>Total sample mass (mg)</th>
<th>Water content in different notations</th>
<th>No. of H₂O molecules per Mb molecule</th>
<th>H₂O gram of H₂O per gram of Mb</th>
<th>H₂O (wt%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNO₃</td>
<td>86</td>
<td>13.9</td>
<td>695</td>
<td>0.74</td>
<td>42.4</td>
<td></td>
</tr>
<tr>
<td>Pure liquid water</td>
<td>80</td>
<td>12.9</td>
<td>488</td>
<td>0.52</td>
<td>34.1</td>
<td></td>
</tr>
<tr>
<td>KCl</td>
<td>80</td>
<td>11.8</td>
<td>447</td>
<td>0.48</td>
<td>32.2</td>
<td></td>
</tr>
<tr>
<td>Mg(NO₃)₂</td>
<td>70</td>
<td>11.7</td>
<td>325</td>
<td>0.34</td>
<td>25.6</td>
<td></td>
</tr>
<tr>
<td>NaCl</td>
<td>73</td>
<td>10.5</td>
<td>222</td>
<td>0.24</td>
<td>19.0</td>
<td></td>
</tr>
<tr>
<td>NaBr</td>
<td>55</td>
<td>10.0</td>
<td>179</td>
<td>0.19</td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>As-received sample</td>
<td>N/A</td>
<td>9.8</td>
<td>107</td>
<td>0.11</td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>CH₃COOK</td>
<td>25</td>
<td>9.1</td>
<td>78</td>
<td>0.08</td>
<td>7.7</td>
<td></td>
</tr>
<tr>
<td>MgCl₂</td>
<td>31</td>
<td>9.4</td>
<td>64</td>
<td>0.07</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>LiCl</td>
<td>19</td>
<td>10.6</td>
<td>47</td>
<td>0.05</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>NaOH</td>
<td>18</td>
<td>8.4</td>
<td>35</td>
<td>0.04</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>Desiccant</td>
<td>18</td>
<td>8.4</td>
<td>11</td>
<td>0.01</td>
<td>1.2</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3-2. Measuring the water content of a hydrated Mb sample. The mass of the sample drops significantly upon heating due to the loss of water. When heated for a prolonged period of time, the sample mass will stabilize. When heat stops, the sample mass quickly rises due to absorption of water from the ambient atmosphere.
3.2.2. Myoglobin Solution Samples

For the Mb solutions, a different scheme was applied. Suppose we want to produce solutions with water contents of 70, 75, 80, 85, 90, 95, and 98 wt%. The following formula can be used to calculate the amount of deionized water $m_{\text{H}_2\text{O}}$ needed to add to a given amount of lyophilized Mb powder $m_{\text{Mb}}$ to produce solutions with specific water content $x$:

$$m_{\text{H}_2\text{O}} = \frac{x - x_0}{1 - x} m_{\text{Mb}},$$  \hspace{1cm} (3.2)

where $x_0 = 10$ wt% is the water content of the as-received lyophilized Mb powder. Based on this formula, the mass of lyophilized Mb powder and deionized water can be calculated (Table 3.2). Due to uncertainties in the mass measurements, the actual water content of the solution may not be exactly what it is intended to be. The relative error is around $\pm 2\%$.

One thing worth noting in Table 3.2 is that for water concentrations down to about 80 wt%, the Mb powders appeared to fully dissolve, while a suspension of Mb particles was visible for the samples at 75 and 70 wt% water concentrations, indicating incomplete mixing. Though it is indicated in the Material Safety Data Sheet (MSDS) of Mb that the solubility in liquid water is only 20 mg/ml [116], researchers typically obtain concentrations of 170 mg/ml (about 85 wt% water content) at pH 6.9 [115], which is roughly the same as our dissolving limit.
Table 3-2. Preparation of Mb aqueous solutions, indicating the masses of Mb lyophilized powder and deionized water needed.

<table>
<thead>
<tr>
<th>Water content of Mb solution $x$ (wt%)</th>
<th>$\frac{x-x_0}{1-x}$</th>
<th>Mass of Mb lyophilized powder $m_{\text{Mb}}$ (mg)</th>
<th>Mass of deionized water $m_{\text{H}_2\text{O}}$ (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>98</td>
<td>43.87</td>
<td>21.7</td>
<td>951.9</td>
</tr>
<tr>
<td>95</td>
<td>16.95</td>
<td>32.9</td>
<td>557.5</td>
</tr>
<tr>
<td>90</td>
<td>7.97</td>
<td>78.0</td>
<td>621.9</td>
</tr>
<tr>
<td>85</td>
<td>4.98</td>
<td>88.3</td>
<td>439.9</td>
</tr>
<tr>
<td>80</td>
<td>3.49</td>
<td>116.4</td>
<td>405.8</td>
</tr>
<tr>
<td>75</td>
<td>2.59</td>
<td>130.3</td>
<td>337.4</td>
</tr>
<tr>
<td>70</td>
<td>1.99</td>
<td>138.0</td>
<td>274.8</td>
</tr>
</tbody>
</table>
3.3. Experiments

In this section, experimental details will be described, including the sample cell design, details about THz time-domain measurements, and FTIR spectroscopy.

3.3.1. Sample Cell Considerations

The sample cell for Mb powders is made of two 3.18 mm (1/8 inches) thick pieces of high-density polyethylene (HDPE) clamped together, with a surface recess in the center of one piece to hold the Mb powder. The diameter of each HDPE plate is 38.1 mm (1.5 inches) and the cylindrical recess has a diameter of 12.7 mm (0.5 inches) and a depth of either 1.00 mm or 0.50 mm. The cell with the thinner recess is for higher water content samples with strong absorption (e.g. Mb powder with 42 wt% water content). A side view of the sample holder is shown in Figure 3.3.

![Figure 3-3. Side view of the sample holder for Mb powder. The walls of the sample holder are two HDPE plates. There is a surface recess with a depth of either 1.00 mm or 0.50 mm in the center of one HDPE plate to hold the sample.](image)
In the THz region, HDPE is one of the most widely used materials for lenses and window applications due to its transparency, machinability, stability and availability [117]. Depending on the amount of impurities incorporated during the manufacturing processes, the absorption coefficient of HDPE may vary from sample to sample. Our measurements confirmed an absorption coefficient less than 0.04 mm$^{-1}$ over the frequency range of interest (0.2-2.0 THz) and this was consistent with other published data [118-120]. For practical applications, HDPE is often assumed to be transparent in the THz region. The refractive index of HDPE is determined as $n_{\text{HDPE}} = 1.533 \pm 0.007$. There will be more discussion on the refractive index of HDPE in the next section.

While preparing the samples, care was taken to ensure that the recess was uniformly filled with Mb powders. Because radiation at 1.0 THz frequency corresponds to wavelength of 300 µm, which is much larger than typical Mb particle sizes, elastic scattering effects from Mb particles should be negligible. Vacuum grease was applied between the two HDPE pieces as a sealant to maintain the hydration level of the enclosed sample.

For solution samples with high water content (above 70 wt%), the THz absorption is much stronger, so that very short but accurately known path lengths are required. We used commercially available cuvettes (purchased from Nova Biotech, www.novabiotech.com) made of optical glass with sample compartment thicknesses of 100 µm (part # G584) and 200 µm (part # G585), as shown in Figure 3.4 below (left). The width of the sample compartment is 8.0 mm and the thickness of the glass is 1.2 mm. By comparing the absorption spectra from both path lengths, it was possible to distinguish the absorption coefficients of the sample from interference effects due to the sample holder, as discussed in Section 3.4. Also shown in Figure 3.4 (right) is a spring loaded adapter to hold the glass cuvette (also purchased from Nova Biotech, part # AC-680).
Simple measurements on the THz absorption properties of the optical glass were performed. The optical glass absorbs THz radiation more strongly than HDPE but the absorption coefficient is still less than 0.1 mm$^{-1}$ for the whole frequency range of interest. Considering the thickness of the optical glass of 1.2 mm, the intensity attenuation due to the sample holder is around 20%, which is calculated as $1 - \exp(-2 \cdot 1.2 \cdot 0.1) \approx 0.2$. There is a factor of 2 in the exponential term because there are two pieces of glass, one for the bottom piece and one for the cover piece. The refractive index of the material is shown to be around $2.0 \pm 0.1$ and is nearly frequency-independent. Fortunately, the optical constants of this glass material don't come into play in the data analysis for Mb solutions (Section 3.4.3).

Figure 3-4. Sample cell and adapter to hold the cell in place$^3$.

---

$^3$ Pictures are taken from the vendor's webpage: http://www.novabiotech.com/cells.htm
3.3.2. Measurement Details

The experimental setup of THz-TDS used in this work has been described in Section 2.3 (Figure 2.8). In this section, some of the experimental parameters will be presented.

HDPE and silicon lenses with focal length of 7.62 cm (3.0 inches) are used to manipulate THz radiation in free space. THz radiation is collimated and then focused to a spot size of less than 5.0 mm at the sample location by the lenses. Since the surface recess for holding the powder samples has a diameter of 12.7 mm (8.0 mm of compartment width for the glass cuvette holding the solution samples), as discussed in Section 3.3.1, no additional aperture is applied to narrow the beam down.

The signal measured by the detector is the time-domain waveform. It is then processed by fast Fourier transform (FFT) to obtain the frequency spectrum. The entire scanning time window for the time-domain waveforms is about 70 ps, resulting in spectral resolution of around 0.5 cm\(^{-1}\) (15 GHz). The scan rate of the delay stage is 5 ps/s, so it takes about 14 seconds to run a single scan. To increase the signal-to-noise ratio (SNR), 16 scans were recorded and averaged by the computer to obtain one measurement, corresponding to a data collection time of about 4 minutes. The resulting SNR is above 3000:1. To further account for system stability issues, multiple measurements were taken for one sample and the data were averaged to obtain the final spectrum. All measurements were taken at room temperature.

In Figure 3.5, a signal waveform taken in room air with relative humidity around 35% is included, along with its FFT frequency domain spectrum shown as an inset. Also shown in the figure is the scan taken in dry nitrogen. Compared with this smooth reference spectrum (with some wiggles due to experimental noises),
the absorption lines due to water vapor in room air are clearly identifiable. To test the reproducibility of the THz system, this spectrum was taken routinely to ensure the same features were obtained every time.

Figure 3-5. THz time-domain waveform and its Fourier transform (inset). Solid line is the spectrum taken in dry nitrogen, and dotted line is the spectrum taken in ordinary laboratory air (~35% relative humidity). There are many oscillations on the time-domain waveform for humid air due to the vibrational and rotational motions of water molecules.
To quantitatively validate the THz-TDS measurements, selected samples were measured with conventional FTIR spectroscopy. This work was performed by Enver Tarhan in Prof. Ramdas's lab in the Department of Physics, Purdue University. The spectrometer is a BOMEM DA.3 [121] rapid scan Michelson interferometer (Figure 3.6). A mercury source along with a set of mylar beamsplitters of different thicknesses were employed for the THz range, and a composite silicon bolometer operating at 1.8 K was used as detector [122]. The same specimens and holders used in the THz-TDS measurements were used without modification for the FTIR measurements.

Figure 3-6. A photo of the BOMEM DA.3 rapid scan Michelson interferometer in Prof. Ramdas's lab.
3.4. Data Analysis and Material Parameters Extraction Procedures

Data analysis procedures for determining the complex refractive index and absorption coefficient of Mb samples from THz-TDS data are described in this section. The methods to calculate the parameters for powder and solution samples are different and they are also treated differently. Since the time-domain measurement cannot go infinitely long and there must be some kind of time window applied on the THz time profile to cut it off before applying the FFT analysis, the windowing effect on the data analysis is also investigated in some detail. At the end of this section, the experimental error estimation procedure is described.

3.4.1. Refractive Index of HDPE

Since HDPE was used to make the powder sample holder in the current experiments, a better knowledge of its refractive index is essential. As confirmed by our measurements and consistent with previous published results (Section 3.3.1), HDPE is nearly transparent in the frequency range of interest (0.2-2 THz); the imaginary part of its refractive index is essentially zero. We further assume that there is negligible pulse distortion due to group velocity dispersion of the THz pulse in HDPE and this is also confirmed by the measurements, so that the only effect on the pulse while it transmits through an HDPE plate is a time delay that depends on the refractive index of the material and its thickness. So the refractive index of HDPE can be written as:

$$n_{\text{HDPE}} = n_{\text{Air}} + \frac{c\Delta t}{l},$$  \hspace{1cm} (3.3)

where $n_{\text{Air}}$ is the refractive index of air, $c$ is the speed of light in vacuum, $\Delta t$ is
the measured time delay caused by inserting an HDPE plate, and \( l \) is the thickness of the HDPE plate. For a thickness of 3.059 mm (determined by the algorithm presented in the next subsection), we obtained a refractive index of \( n_{\text{HDPE}} = 1.533 \pm 0.007 \), which is consistent with the widely accepted values between 1.52-1.54 [117]. Further analysis confirms that the refractive index of HDPE is essentially frequency-independent, with a variation of less than 0.01 over the usable frequency range (0.2-2.0 THz).

### 3.4.2. Parameters Extraction Methods for Powder Samples

One important problem in THz-TDS is how to extract the material parameters reliably from the time-domain waveforms [26]. A commonly used method is to record the waveform twice, in the presence and absence of the sample to be characterized. As shown in Figure 3.7, medium 1 and 3 are the HDPE sample cell plates, and medium 2 is the Mb powder. The air around the sample cell is taken as medium 0. For the reference measurement, instead of leaving an air gap between two pieces of HDPE with thickness the same as the Mb powder, we clamped two identical HDPE plates together so that the Fabry-Perot factor from an air gap need not be considered (see Eqn. (3.5)). The recorded waveform shows the temporal profile of the THz electric field. The Fourier transform gives its spectral components at angular frequency \( \omega \). In the following equations, \( E_{\text{Sample}}(\omega) \) and \( E_{\text{Ref}}(\omega) \) are the spectral components for the sample and reference measurements, respectively. These can be represented as:

\[
E_{\text{Sample}}(\omega) = \eta(\omega)E_0(\omega)T_{12}(\omega)P_2(\omega,L)FP(\omega)T_{23}(\omega) ,
\]

\[
E_{\text{Ref}}(\omega) = \eta(\omega)E_0(\omega)T_{13}(\omega)P_{\text{Air}}(\omega,L) .
\]
where \( FP(\omega) \) is the factor accounting for Fabry-Perot effect (multiple reflections) in the sample:

\[
FP(\omega) = \sum_{k=0}^{\infty} \left[ R_{23}(\omega) P_{21}^2(\omega) R_{21}(\omega) \right]^k.
\]  

(3.6)

In these equations, \( R_{ab}(\omega) = \frac{\bar{n}_a - \bar{n}_b}{\bar{n}_a + \bar{n}_b} \) is the reflection coefficient at the \( a-b \) interface and \( T_{ab}(\omega) = \frac{2\bar{n}_a}{\bar{n}_a + \bar{n}_b} \) is the transmission coefficient from medium \( a \) to medium \( b \). \( P_a(\omega, L) = \exp(-i\bar{n}_a \omega L / c) \) is the propagation coefficient in medium \( a \) over a distance \( L \). \( \bar{n}_a = n_a - i\kappa_a \) is the complex refractive index of medium \( a \).

The expression in Eqn. (3.6) means each echo \( k \) after the primary peak endures one backward reflection \( R_{23}(\omega) \), one forward reflection \( R_{21}(\omega) \), and two propagations \( P_{21}^2(\omega) \) through the sample, so that the total effect is the multiplication of all the terms summed over all the echoes. However, only the first, which occurs at a delay of less than 10 ps after the main peak, needs to be considered, because higher order reflections have amplitudes below the noise level and are not detectable.

\( \eta(\omega) \) in Eqns. (3.4) and (3.5) is given by the following expression which accounts for the transmission, propagation, and multiple reflection terms in medium 1 & 3:

\[
\eta(\omega) = T_{01}(\omega) P_1(\omega) FP_1(\omega) \cdot P_3(\omega) FP_3(\omega) T_{30}(\omega),
\]

(3.7)

It should be noticed that the Fabry-Perot factors in \( \eta(\omega) \) for Eqns. (3.4) and (3.5) have different behaviors because the surrounding materials are different. However, the echoes created in medium 1 and 3 occur on a time scale much
larger than the signal of interest. In fact, the first echo corresponding to multiple
time at about 17 ps after the main peak was applied and the THz signal
outside this time window was set to zero in order to filter out the echoes due to
the HDPE sample cell wall. So we assume $\eta(\omega)$ is the same for Eqns. (3.4) and
(3.5) and therefore can be divided out in the subsequent analysis.

Figure 3-7. Measurement scheme requires one measurement without sample,
and another with sample (Medium 2) of length L. The sample holder (Medium 1
and 3) is made of HDPE.
The complex transmission coefficient of the sample is obtained by dividing $E_{\text{Sample}}(\omega)$ by $E_{\text{Ref}}(\omega)$ (let medium 1 and 3 be HDPE and medium 2 be Mb so that $T_{13}(\omega)$ in Eqn. (3.5) is unity):

$$T(\omega) = \frac{4n_{\text{HDPE}}\tilde{n}_{\text{Mb}}}{(n_{\text{HDPE}} + \tilde{n}_{\text{Mb}})^2} \exp(-i\omega\tilde{n}_{\text{Mb}} - n_{\text{Air}})L/c)FP(\omega),$$

with

$$FP(\omega) = \frac{1}{1 - \left(\frac{\tilde{n}_{\text{Mb}} - n_{\text{HDPE}}}{\tilde{n}_{\text{Mb}} + n_{\text{HDPE}}}\right) \exp(-2i\tilde{n}_{\text{Mb}}\omega L/c)},$$

which is the summation in Eqn. (3.6). What we must solve here is an inverse electromagnetic problem, i.e. extracting the complex refractive index of sample $\tilde{n}_{\text{Mb}} = n_{\text{Mb}} - i\kappa_{\text{Mb}}$ from the measured transmission coefficient. Fabry-Perot effects due to surface reflections can produce artificial oscillations on the absorption spectrum, and tend to complicate the interpretation of experimental results [99].

For this analysis, we employed the algorithm presented by L. Duvillaret et al [26]. An error function with $n_{\text{Mb}}$ and $\kappa_{\text{Mb}}$ as variables characterizing the difference between theoretical and experimental transmission coefficient is defined, and an iterative procedure performed to obtain refined material parameters until the error is lower than a pre-assigned value related to the experimental uncertainty. In this algorithm, the Fabry-Perot factor is treated as a perturbation and taken into account in each loop. Once the complex refractive index $\tilde{n}_{\text{Mb}} = n_{\text{Mb}} - i\kappa_{\text{Mb}}$ of the Mb sample is obtained, the absorption coefficient $\alpha(\omega)$ is related to the imaginary part of $\tilde{n}_{\text{Mb}}$ by:
This extraction procedure for the frequency-dependent complex refractive index function assumes a measured sample thickness $L$. The precision of sample thickness measurement proved to be critical for the accuracy of extracted parameters [123]. A modification to the algorithm allowed us to determine the thickness and the complex refractive index of the sample simultaneously [27]. In this method, the complex refractive index of the sample was determined applying the above algorithm at each trial thickness over a range around the measured sample thickness. Once the material parameters $\tilde{n}_{\text{Mb}} = n_{\text{Mb}} - i\kappa_{\text{Mb}}$ were obtained for a particular thickness $L$, a total variation metric was defined as follows to measure the smoothness of the complex refractive index function with frequency at this thickness:

$$D(\omega_m) = |n_{\text{Mb}}(\omega_m) - n_{\text{Mb}}(\omega_{m-1})| + |\kappa_{\text{Mb}}(\omega_m) - \kappa_{\text{Mb}}(\omega_{m-1})|.$$ (3.11)

$$TV = \sum_m D(\omega_m),$$ (3.12)

where the index $m$ denotes frequency. The total variation $TV$ is a function of thickness and its value measures the ruggedness of the computed complex refractive index curve. The fewer oscillations the computed complex refractive index curve has, the smaller the total variation metric will be. The sample thickness $L_0$ was identified by the deepest local minimum of $TV$ [27] and the sample parameters can then be calculated at $L_0$.

In the real implementation of the above algorithm, the sample thickness determination is performed in two steps. In the first step, a larger thickness range is guessed and the material parameters are computed at a coarse thickness step size. Once a deepest local minimum is reached for $TV$ at a certain thickness $L_0$, a
smaller thickness range is utilized around \( L_0 \) and the material parameters are computed at a finer step size.

Figure 3.8 and 3.9 show the procedure of determining the thickness of a HDPE plate using the algorithm above. The nominal thickness of the HDPE plate is 3.175 mm (1/8 inches), so in the first step, the thickness is guessed in the range of \( L = 3.175 \pm 0.5 \) mm with a step size of 0.1 mm. The material parameters are determined at each thickness and the total variation for the complex refractive index as defined in Eqn. (3.12) is determined. The total variation as a function of guessed thickness is shown in Figure 3.8. Also shown in Figure 3.8 is the so-called total variation of the second order defined as [27]

\[
TV2 = \sum_m |D(\omega_m) - D(\omega_{m+1})|.
\]

This is an extra metric to determine the local minimum for the total variation. From Figure 3.8, the local minimum is determined at \( L_0 = 3.075 \) mm.

In the second step shown in Figure 3.9, the sample thickness is guessed around \( L_0 = 3.075 \) mm with a finer step size of 0.01 mm. The local minimum is at 3.055 mm. In this case for HDPE plate, we take one more step with a step size one order of magnitude smaller (0.001 mm) and the sample thickness is finally determined at 3.059 mm, which is also the sample thickness used in Section 3.4.1 to determine the refractive index of HDPE.
Figure 3-8. Procedure to determine the thickness of a HDPE plate. The total variation of the complex refractive index of the first order (top) and second order (bottom) is plotted as a function of guessed sample thickness and the real sample thickness is determined as the local minimum of the curve (3.075 mm).
Figure 3-9. Procedure to determine the thickness of a HDPE plate. The local minimum of the total variation is at 3.055 mm.
All the sample thicknesses of Mb powders are determined applying the same procedure presented above and are listed in Table 3.3. Using the thickness information, together with the total mass of the sample and the diameter of the sample holder (12.7 mm), the density of the sample can be calculated and are also listed in Table 3.3. The density of the sample will be used later to calculate the absorption coefficient of the sample per cross sectional area. Another related parameter listed in the table is the nominal molar concentration of Mb, $M_{M_b}$, defined as:

$$M_{M_b} = \frac{(1-x)m}{V \cdot MW_{M_b}},$$

(3.14)

where $x$ is the sample water content and $m$ is the total sample mass. $V = \pi r^2 l$ is the sample volume with $r$ and $l$ being the sample radius and thickness, respectively. $MW_{M_b} = 16952$ (unit in g/mol or mg/m-mol) is the molecular weight of Mb.

In Table 3.3, the uncertainty in sample density is determined by the uncertainties in sample mass $m$, sample thickness $l$, and sample holder radius $r$. Based on the error analysis theory, the relative error of the sample density is given by

$$\frac{\Delta \rho}{\rho} = \sqrt{\left(\frac{\Delta m}{m}\right)^2 + \left(\frac{\Delta l}{l}\right)^2 + 4\left(\frac{\Delta r}{r}\right)^2}.$$

(3.15)

All the three quantities on the right hand side can be determined accurate to $\pm 1\%$, so the sample density has a relative error about $\pm 2\%$. Similarly, the Mb molar concentration listed in Table 3.3 will have a relative error about $\pm 3\%$ if we consider the uncertainty in the sample water content is $\pm 2\%$ (Section 3.2.1).
Table 3-3. Some parameters of the sample. The sample thickness is determined using the algorithm above. For the two samples with highest water content, we used the thinner sample holder so that the absorption was not too strong.

<table>
<thead>
<tr>
<th>Sample water content $x$ (wt%)</th>
<th>Total sample mass $m$ (mg)</th>
<th>Sample thickness $l$ (mm)</th>
<th>Sample density $\rho = \frac{m}{\pi l^2}$ (mg/mm$^3$)</th>
<th>Mb molar concentration $M_{\text{Mb}} = \frac{(1-x)m}{V \cdot MW_{\text{Mb}}}$ (milli-mol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>42.4</td>
<td>13.9</td>
<td>0.499</td>
<td>0.220</td>
<td>7.47</td>
</tr>
<tr>
<td>34.1</td>
<td>12.9</td>
<td>0.512</td>
<td>0.200</td>
<td>7.73</td>
</tr>
<tr>
<td>32.2</td>
<td>11.8</td>
<td>1.133</td>
<td>0.082</td>
<td>3.29</td>
</tr>
<tr>
<td>25.6</td>
<td>11.7</td>
<td>1.002</td>
<td>0.092</td>
<td>4.05</td>
</tr>
<tr>
<td>19.0</td>
<td>10.5</td>
<td>1.097</td>
<td>0.076</td>
<td>3.61</td>
</tr>
<tr>
<td>16.0</td>
<td>10.0</td>
<td>1.180</td>
<td>0.067</td>
<td>3.32</td>
</tr>
<tr>
<td>10.2</td>
<td>9.8</td>
<td>1.171</td>
<td>0.066</td>
<td>3.50</td>
</tr>
<tr>
<td>7.7</td>
<td>9.1</td>
<td>1.054</td>
<td>0.068</td>
<td>3.71</td>
</tr>
<tr>
<td>6.4</td>
<td>9.4</td>
<td>0.984</td>
<td>0.075</td>
<td>4.16</td>
</tr>
<tr>
<td>4.7</td>
<td>10.6</td>
<td>1.000</td>
<td>0.084</td>
<td>4.70</td>
</tr>
<tr>
<td>3.6</td>
<td>8.4</td>
<td>1.147</td>
<td>0.058</td>
<td>3.29</td>
</tr>
<tr>
<td>1.2</td>
<td>8.4</td>
<td>1.090</td>
<td>0.061</td>
<td>3.55</td>
</tr>
</tbody>
</table>
Figure 3.10 gives an example of the absorption spectrum determination procedure for a selected sample with water content of 19.0 wt%. The echo due to reflections from the front and back interfaces of the sample with the sample cell can be seen at about 8.0 ps after the main peak in the time domain waveform (see arrow in Figure 3.10 (A)). This echo is responsible for the Fabry-Perot oscillations seen below 1.0 THz in the transmission amplitude spectrum (Figure 3.10 (C)). The phase information is shown as an inset in both the FFT frequency domain spectrum (Figure 3.10 (B)) and the complex transmission spectrum (Figure 3.10 (C)). The complex refractive index of the sample can be extracted from the complex transmission spectrum, and the absorption spectrum of the sample can be calculated thereafter. In Figure 3.10 (D), the comparison between the absorption spectrum with and without Fabry-Perot effect illustrates slightly less oscillations at low frequencies when the Fabry-Perot factor is included as a perturbation in the extraction algorithm. While the results published by other researcher seem to suffer from the same problem [97, 99, 100], there is still work needed to be done on this side to improve the oscillation effect.

The plots in Figure 3.10 illustrate the two main sources of uncertainty in the interpretation of THz-TDS data. The first is the Fabry-Perot effect just mentioned, where the thickness oscillations are the source of peaks in the low frequency region of the absorption spectrum. Any peaks in THz-TDS absorption spectrum should be demonstrated to be independent of the specimen configuration. The second problem is noisy peaks seen at higher frequencies, caused by the baseline noise from the spectrometer. That is, the output signal consists of the real signal plus an additive noise level. If the specimen is strongly absorbing, the high frequency amplitudes are more strongly affected, and the output signal above a given frequency may be entirely due to background noise. The peaks seen above 1.5 THz in Figure 3.10 (C) & (D) are largely due to this background effect. Thinner, less absorbing specimens are required to get better data at these frequencies.
Figure 3-10. Data collection and analysis for Mb powder at 19.0 wt% water content. (A) Comparison of time waveforms recorded without sample (solid line) and with Mb powder (dotted line) in the N₂-purged chamber; arrow denotes first Fabry-Perot reflection within the sample cell. (B) Fourier-transform frequency spectra of the two waveforms, plus their phase angle (inset). (C) Ratio of amplitudes versus frequency, and phase difference (inset). (D) Calculated absorption coefficient (mm⁻¹) before (dotted line) and after (solid line) correction for Fabry-Perot oscillations due to sample cell dimensions. Negative absorption coefficient at the lowest frequencies is most likely due to system uncertainties and data analysis artifacts (see text).
Negative absorption coefficients seen at the lowest frequencies in Figure 3.10 (D), and also in the samples with lower water content presented in Section 3.5 below, raise questions about the system uncertainties and data analysis artifacts. While this phenomenon is not uncommon in the THz-TDS studies of biomolecules [94-100], additional study is required to track it down.

3.4.3. Parameters Extraction Methods for Solution Samples

The measurement and data analysis scheme applied for Mb in aqueous solution is analogous to the method used by C. A. Schmuttenmaer et al for the measurements of liquid water and other polar liquids [124]. In their experiments, the liquid was sealed in a polyethylene bag and pressed between two parallel windows of either HDPE or high-resistivity silicon. The material path length could be varied by keeping one window fixed and moving the other through a translation stage. The measurements were taken for several path lengths, at intervals of 50 or 100 μm, in the range of 50 to 1000 μm. At each path length \( d \), the time dependent electric field of the transmitted THz pulse was obtained and Fourier transformed to obtain the power \( P(\omega, d) \) and phase \( \phi(\omega, d) \) of each frequency component \( \omega \). The absorption coefficient \( \alpha(\omega) \) at each frequency was determined by the slope of a linear regression fit of \( \ln P(\omega, d) \) to path length \( d \) :

\[
\ln P(\omega, d) = \ln P_0(\omega) - \alpha(\omega)d .
\]  

(3.16)

Similarly, the frequency-dependent refractive index \( n(\omega) \) was calculated from the path length dependence of the phases \( \phi(\omega, d) \):

\[
\phi(\omega, d) = \phi_0(\omega) + \frac{2\pi\omega(n(\omega) - 1)}{c}d .
\]  

(3.17)
For the current measurements we had only two rectangular cuvettes of different thicknesses (100 and 200 μm) readily available to use as holders for Mb solutions. Measurements were taken separately with both sample holders and reproducible time-domain THz waveforms were obtained within experimental uncertainties. The temporal profile of the THz electric field was Fourier transformed to yield the spectral components at angular frequency \( \omega \). Before applying an FFT, some manipulation of the time-domain waveform are necessary. First, a DC offset due to the data acquisition circuit on the waveform is subtracted. Then a cutoff time of about 10 ps after the main THz peak is applied and the waveform is truncated to remove satellite peaks resulting from reflection on the sample cell walls (1.2 mm of optical glass) that occurs about 14 ps after the main peak and carry no significant information about the sample. Finally, the time-domain waveform is zero-padded to 4096 points (including measured data of 3060 points) to increase the effective frequency resolution. After performing an FFT on the time-domain waveform to obtain the frequency-dependent amplitude \( E(\omega) \) and phase \( \phi(\omega) \) of the electric field, Eqns. (3.16) and (3.17) are converted and the absorption coefficient \( \alpha(\omega) \) and refractive index \( n(\omega) \) of each sample are written as:

\[
\alpha(\omega) = \frac{\ln P_1(\omega) - \ln P_2(\omega)}{d},
\]

\[
n(\omega) = \frac{(\phi_1(\omega) - \phi_2(\omega))c}{d\omega} + 1.
\]

In these formulas, \( P_1 \) and \( P_2 \), \( \phi_1 \) and \( \phi_2 \) are the transmission intensity (\( P \propto E^2 \)) and phase shift at 100 and 200 μm path lengths, respectively; \( d \) equals 100 μm and it is the path length difference between the two cuvettes.

In using Eqns. (3.18) and (3.19), it was assumed that the reflection and transmission coefficients due to the sample cell walls are identical and therefore
can be divided out. The Fabry-Perot reflections occur 14 ps after the primary transmitted peak and can be removed by applying a time window 10 ps after the primary peak. The first Fabry-Perot echo due to the sample itself is entangled with the main peak and can not be separated by windowing. We can roughly estimate its amplitude by evaluating the first-order Fabry-Perot factor using Eqn. (3.6). Taking \( k = 1 \), the amplitude of the first Fabry-Perot echo can be approximated as

\[
|FP^{(1)}(\omega)| = \left| R_{23}(\omega) P_2^2(\omega) R_{21}(\omega) \right| = \left( \frac{n_{\text{Water}} - n_{\text{Glass}}}{n_{\text{Water}} + n_{\text{Glass}}} \right)^2 \exp(-\alpha_{\text{Water}} L),
\]

which is a function of frequency. Eqn. (3.10) is used to evaluate the propagation term \( P_2^2(\omega) \) in Eqn. (3.20).

Let us estimate the value of Eqn. (3.20) at 0.5 THz as an example. At 0.5 THz, the refractive indices of liquid water \( n_{\text{Water}} \) and optical glass \( n_{\text{Glass}} \) are around 2.3 [125] and 2.0 (Section 3.3.1), respectively. The absorption coefficient of liquid water \( \alpha_{\text{Water}} \) is 16 mm\(^{-1}\) [125] and the water layer thickness \( L \) is 0.1 mm. Based on these numbers, Eqn. (3.20) can be evaluated to be around 0.1%. At higher frequencies, the absorption coefficient of liquid water increases and its refractive index decreases so that this number will be even smaller and the Fabry-Perot effect due to the sample itself is negligible. Therefore the only differences between the two measurements are stronger absorption and larger phase shift for the longer path length, allowing the use of Eqns (3.18) and (3.19) to extract the material parameters.

For the data analysis in the subsequent sections, the densities of Mb solutions are determined by measuring the mass and volume of each sample. The results are listed in Table 3.4. Also listed is the molar concentration of Mb as calculated
by Eqn. (3.14), for future use. Compared with the Mb molar concentration for lower water content samples (higher Mb weight content) in Table 3.3, it is noticed that the Mb molar concentration is higher for Mb solutions. This is largely due to the much higher density of aqueous solutions because the powder can hardly be packed as dense as the liquid.

In Table 3.4, the uncertainty in sample density is determined by the uncertainties in sample mass $m$ and volume $V$. Since both of these quantities can be measured with an accuracy of $\pm 1\%$, the relative error in sample density is around $\pm 1\%$. The Mb molar concentration will have an error around $\pm 2\%$ if considering the water content uncertainty of $\pm 2\%$ (Section 3.2.2).

### Table 3-4. Some parameters of Mb solutions. We first measured a certain volume of sample (second column in the table) and then the sample is weighed to get the mass. The sample density and Mb molar concentration can therefore be calculated.

<table>
<thead>
<tr>
<th>Sample water content $x$ (wt%)</th>
<th>Sample volume $V$ (mm$^3$)</th>
<th>Measured sample mass $m$ (mg)</th>
<th>Sample density $\rho = \frac{m}{V}$ (mg/mm$^3$)</th>
<th>Mb molar concentration $M_{\text{Mb}} = \frac{(1-x)m}{V \cdot MW_{\text{Mb}}}$ (milli-mol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>98</td>
<td>50.0</td>
<td>49.8</td>
<td>1.00</td>
<td>1.18</td>
</tr>
<tr>
<td>95</td>
<td>50.0</td>
<td>50.4</td>
<td>1.01</td>
<td>2.97</td>
</tr>
<tr>
<td>90</td>
<td>50.0</td>
<td>51.5</td>
<td>1.03</td>
<td>6.08</td>
</tr>
<tr>
<td>85</td>
<td>50.0</td>
<td>51.5</td>
<td>1.03</td>
<td>9.11</td>
</tr>
<tr>
<td>80</td>
<td>20.0</td>
<td>20.7</td>
<td>1.04</td>
<td>12.2</td>
</tr>
<tr>
<td>75</td>
<td>20.0</td>
<td>20.2</td>
<td>1.01</td>
<td>14.9</td>
</tr>
<tr>
<td>70</td>
<td>20.0</td>
<td>20.4</td>
<td>1.02</td>
<td>18.1</td>
</tr>
</tbody>
</table>
3.4.4. Windowing Effect on the Data Analysis

As indicated in Section 3.4.2 and 3.4.3, in order to remove the multiple oscillations following the main peak, an appropriate window is applied on the time-domain waveform. It is then important to understand the effect of different window functions on the data analysis. We will take the measurement on liquid water as an example in the following analysis because liquid water is highly absorbing and its absorption spectrum and refractive index data are readily available.

Shown in Figure 3.11 are the time-domain waveforms for liquid water at 100 (solid curve) and 200 µm (dashed curve) sample thicknesses. Oscillations can be seen on the curves following the primary transmitted THz pulse. The big dip around 183 ps is due to a reflection in the THz receiver structure.

The simplest window that could be applied on the waveform is a rectangular window, which cuts off the signal at about 8 ps before the main peak and 10 ps after the main peak. The truncation time is so chosen because the THz signal is close to zero at 8 ps before and 12 ps after the main peak. Figure 3.12 shows the effect of the truncation. To smooth the abruptness at the edge of the window due to truncation, some extrapolations based on the linear extension of the data points near the window edge are used.
Figure 3-11. Time-domain waveforms of liquid water sample with thickness of 100 (solid curve) and 200 μm (dashed curve).
Figure 3-12. The effect of rectangular window on the time-domain waveforms of liquid water. To reduce the abruptness at the window edge due to truncation, some linear extrapolations are used to smooth the signal.
For the purpose of data analysis, it is preferable to use more sophisticated window functions [126]. Shown in Figure 3.13 are the effects of three commonly used window functions (the built-in window functions in MATLAB are applied) on the time-domain THz pulses, the Hamming window, the Blackman-Harris window, and the Flat Top window. While all three windows keep the peak portion of the THz pulses unchanged, they do alter the tails of the pulses due to different tapering behaviors.

Figure 3-13. Windowing effect on the time-domain waveforms of liquid water. Three types of window functions are used in the figure: Hamming window, Blackman-Harris window, and Flat Top window.

For reference, Figure 3.14 illustrates the shapes of the three window functions (1024 data points) used in Figure 3.13. It is noticed that the Hamming window doesn't fall to zero at the window edge, while the Flat Top window has a negative tail. We are not going to discuss the detailed behaviors of these window functions since this is beyond the scope of this thesis work.
So finally, what are the effects of the window function on the absorption spectrum of liquid water, as well as its refractive index data? Shown in Figure 3.15 is the absorption spectrum of liquid water obtained by applying different window functions on the time-domain waveform. The results from published data (the black line) [125] are also included for comparison. Good agreement is obtained between the measurements in this work and the published values.

The absorption spectrum (blue line) oscillates dramatically if we just use the measured raw time-domain waveform to extract the material parameters because of the various reflections present on the THz time profile due to sample holder and receiver structure and also the truncation effect when applying a rectangular window. For all the situations where a certain window function is used, the absorption spectra all follow the same trend but the oscillations are significantly reduced. For the two window functions—rectangular window and Hamming window—which are said to have high resolution (the ability to differentiate neighboring signals) but low dynamic range (the ratio between the
maximal and minimal detectable signal), there are some oscillations in the higher frequency range. These oscillations are most possibly due to the side lobes of the Fourier transform of the window function. The application of a window function in the time-domain is equivalent to convolving the actual signal frequency response with the Fourier transform of the window function in the frequency-domain. For example, the Fourier transform of a rectangular window is a sinc function. If the time width of the window is 20 ps, which is roughly the time window in our situation, the separation between primary zeros will be 100 GHz. This will add extra oscillations on the spectrum as well as influence the low frequency portion of the spectrum. For the two other window functions---Blackman-Harris window and Flat Top window, which are said to have higher dynamic range, there are much smaller oscillations on the absorption spectrum in the high frequency range compared with rectangular window and Hamming window. This advantage is understandable in that we need larger dynamic range more than better resolution since the absorption spectrum is smooth any way and there are not really any peaks to be resolved. Larger dynamic range is preferable in detecting the minimal signal from the background noise level.

In the low frequency end, all four windows applied show discrepancies from the published results on water. It is believed that this results from the Fourier side lobes of the window function and is not the subject of study in this thesis.

Figure 3.16 shows the refractive index data obtained by applying different window functions on the waveform, along with the published values for comparison (the black line) [125]. The agreement is excellent except at low frequency where the Fourier side lobes of the window function play an important role.
Figure 3-15. Absorption spectra of liquid water obtained by applying different window functions on the time-domain waveform. The liquid water absorption spectrum from the publish results (black line) [125] is also included for a comparison.
Figure 3-16. Refractive index of liquid water obtained by applying different window functions on the time-domain waveform. The liquid water refractive index data from published results (black line) [125] are also included for a comparison.
For the subsequent analysis, we applied a Blackman-Harris window to the THz time profile because it gives better results for the liquid water absorption spectrum and refractive index data over the frequency range of interest. However, it should be noted that the results and conclusions in this thesis don’t depend on the choice of window function.

For reference, Figure 3.17 illustrates the Fourier transform of the four window functions used. It can be seen that the rectangular window and Hamming window have sharp primary peaks and oscillatory tails, while the Blackman-Harris window and the Flat Top window are broader in the central peaks but the tails are perfectly flat.

Figure 3-17. Fourier transform of different window functions used.
3.4.5. Experimental Error Estimation

There are many factors in the experimental measurements as well as in the data analysis that could cause uncertainties in the final absorption spectrum, such as instability of the input laser power into the THz transmitter, additive background noise due to the detection electronics, windowing effects and residual Fabry-Perot oscillations. It is important to understand how such factors are taken into account in determining the error bars on the various absorption points.

As described in Section 3.4.2, in the actual experiments, for each sample with given water content, two sets of time-domain waveform measurements were taken. One set is for the reference (empty sample cell for Mb powders and 100 μm thick sample for Mb solutions) and the other is for the sample (actual sample for Mb powders and 200 μm thick sample for Mb solutions). For each set, the waveform was measured 3-4 times. Combining each pair of reference/sample measurements will yield one absorption spectrum and one refractive index curve. So in total we could have 9-16 curves. The final result was taken as the average of these curves at each frequency and the standard deviation can also be obtained from these multiple curves. This standard deviation is the error estimation for the absorption coefficient and refractive index at each frequency and will be used for the subsequent analysis. It is significant in that at least 68% of all the data points at each frequency are within one standard deviation from the average. However, it is observed that in the extreme case, the data points may be off from the average by ±10% while the relative error (standard deviation divided by the average) is typically within ±5%.

Shown in Figures 3.18 and 3.19 are the absorption spectrum and refractive index of liquid water, respectively, along with the standard deviations derived from the multiple data sets plotted below and above the average as dashed lines. The results from the published data (black line) [125] are also included for
comparison. A Blackman-Harris window is applied in the data analysis to obtain these results. It is obvious that the experimental error is much smaller at the intermediate frequencies around 0.5 THz as opposed to lower and higher frequencies.

The agreement of our results with the published data is excellent in the frequency range from 0.3 to 1.2 THz. Due to increasingly strong absorption from the sample above 1.2 THz, the transmitted THz signal falls below the instrumental noise level, resulting in poor signal-to-noise ratio. The signal below 0.3 THz suffers from the windowing effect as discussed in the previous subsection, and will not be the main concern of this thesis work. For the refractive index, the agreement is excellent from 0.1 to 1.5 THz. These measurements establish the general validity of the measurements and parameter extraction procedures employed on the Mb-water mixtures.

Figure 3-18. Absorption spectrum of liquid water with standard deviations shown as dashed lines below and above the average. The black line is the published results for comparison [125].
Figure 3-19. Refractive index of liquid water with standard deviations shown as dashed lines below and above the average. The black line is the published results for comparison [125].
3.5. Results and Discussions

Applying the algorithms described in Section 3.4, we obtain the absorption coefficient and refractive index as a function of frequency for the samples under study based on the measured time-domain waveforms. Since the refractive indices are essentially unchanged within experimental uncertainties for Mb powders and Mb solutions and they don't carry as much interesting information as the absorption coefficient, we will only concentrate on the absorption spectra in the following discussions.

Aside from the comparison with published liquid water results in the last section to validate the measurement and data analysis procedures, we separately validate the entire process by recording the THz absorption spectrum for water vapor and comparing it with simulated results from NIST by T. M. Korter et al [127]. Excellent agreement is obtained. Then the absorption spectra of Mb powder and aqueous solutions at a series of well-controlled hydration levels are presented and discussed. The hydration levels range from 1.2 to 98 wt% and the frequencies studied cover the regime from 0.1 to 2.0 THz.

To further validate the THz-TDS measurements on Mb powder samples, absorption spectra were also measured using FTIR spectroscopy. Comparison between these two measurements yields good agreement.
As indicated in Section 3.3.2, to test the reproducibility of THz system, we routinely took scans in room air with relative humidity around 35%, taking the scans in dry nitrogen as a reference, to ensure the same features on the absorption spectrum of water vapor were obtained every time.

Shown in Figure 3.20 is one of these absorption spectra, along with the simulated absorption spectrum for water vapor at room temperature [127]. The agreement of the line centers is very good. For frequencies beyond 2.0 THz, the peaks are suspicious and will not be taken as real because the THz signal falls below the system noise level. It should be noted that the vertical scale is expressed in arbitrary units for absorption coefficient, because we don't have absolute values for the simulated absorption spectrum of water vapor [127] to compare with. There are also significant discrepancies in the relative amplitudes of the peaks. This may partly due to the differences in room air condition between the current measurements and the simulations. While the simulations commonly assume 100% relative humidity at room temperature, the current measurements were performed in room air with relative humidity around 35%. The partial overlap between neighboring absorption lines may also influence the relative amplitude of each peak which is not considered here. All these factors are not the main concerns of this thesis.

Because the THz absorption spectra for materials in the gas phase usually show sharp characteristic peaks, THz-TDS can be used to detect specific molecular species in gaseous environments [108], especially when high temperature is a challenge, because of the inherent immunity of THz detection to thermal background [109, 110].
Figure 3-20. Absorption spectrum of water vapor taken in room air (blue line) along with the simulated results (red line) [127] for comparison. Excellent agreement is obtained except for the relative amplitude of each peak.
3.5.2. Absorption Spectra of Myoglobin Powders

The absorption spectra of Mb powders at four selected water contents are shown in Figure 3.21 below. The samples include one with the highest water content (42 wt%) in powders, one which is essentially dry (3.6 wt%) and two with intermediate water contents (19 and 32 wt%). The error bars shown on each curve at several representative frequencies (0.1, 0.6, and 1.2 THz) are standard deviations as derived from Section 3.4.5. The relative error on each data point is typically less than $\pm 5\%$. The standard deviations are larger at higher frequencies, as the amount of detectable transmitted radiation falls below the baseline noise level. The weak oscillations at low frequencies are likely residual Fabry-Perot oscillations from the sample cell. Also shown in Figure 3.21 on each absorption spectrum is a quadratic curve, which will be discussed below.

It is observed from Figure 3.21 that there is a big difference between the absorption spectrum of 42 wt% water content sample and that of 32 wt% water content sample. This is due to the difference in sample packing density (See Table 3.3). Because when preparing the high water content sample (e.g. 42 wt%) using saturated salt solution with high ERH, the sample will shrink more and result in a thinner layer, thus the sample volume is only about half of that for low water content sample, resulting in a doubled sample density. To account for this difference, the absorption coefficient $\alpha$ (unit in mm$^{-1}$) is divided by the sample density $\rho$ (unit in mg/mm$^3$) to obtain the normalized areal absorption coefficient $\alpha_{\text{Area}}$ (unit in (mg/mm$^2$)$^{-1}$):

\[
\alpha_{\text{Area}} = \frac{\alpha}{\rho}.
\]  

This expression denotes the absorption ability of the sample per unit cross sectional area, which is independent of sample packing density. In this equation,
the uncertainty in $\alpha_{\text{Area}}$ can be derived from the uncertainties in the total absorption coefficient $\alpha$ and the sample density $\rho$, 

$$\frac{\Delta \alpha_{\text{Area}}}{\alpha_{\text{Area}}} = \sqrt{\left(\frac{\Delta \alpha}{\alpha}\right)^2 + \left(\frac{\Delta \rho}{\rho}\right)^2}.$$  \hspace{1cm} (3.22)  

Considering the relative error of $\alpha$ and $\rho$ being $\pm 5\%$ and $\pm 2\%$, respectively, for Mb powder samples, the uncertainty in $\alpha_{\text{Area}}$ is dominated by the uncertainty in $\alpha$, so that the standard deviation of $\alpha_{\text{Area}}$ can be approximated by the standard deviation of $\alpha$ divided by the sample density.

The absorption spectra of the four samples in Figure 3.21 are converted using Eqn. (3.21) and replotted in Figure 3.22, along with the error bars derived based on the analysis above. For all samples shown in Figures 3.21 and 3.22, the far-infrared absorption spectra seen by THz-TDS show a fairly smooth increase with frequency. The spectra lack any identifiable absorption structures that can be assigned to the intrinsic vibrational or rotational modes of Mb molecule. For the samples with higher hydration levels, e.g. 19, 32, and 42 wt\%, it is clear that the water content is responsible for a significant fraction of the THz absorption.
Figure 3-21. The absorption spectra of Mb powders from 3.6 to 42 wt% water contents. The quadratic curve on each line is the fit based on data points from 0.1 to 1.2 THz where the signal-to-noise ratio is good. Error bars at 0.1, 0.6, and 1.2 THz are standard deviations derived from multiple measurements. Weak oscillations at low frequencies are likely residual Fabry-Perot oscillations from the sample cell, and structures at high frequencies are noises due to weak signals and are within the error bars.
Figure 3-22. The absorption spectra of Mb samples from 3.6 to 42 wt% water contents normalized to sample density, along with the quadratic fit. Error bars at 0.1, 0.6, and 1.2 THz are the corresponding error bars in Figure 3.21 divided by the sample density. These absorption spectra can be characterized as essentially smooth, continuous, and without sharp identifiable features.
Each absorption spectrum in Figures 3.21 and 3.22 can be fit by a simple quadratic curve given by:

\[ \alpha = A_0 f^2 + A_1 f + A_2, \]  

(3.23)

where \( f = \omega / 2\pi \) is the frequency, and \( A_0, A_1, \) and \( A_2 \) are the fitting parameters. The fit is based on the data points from 0.1 to 1.2 THz where the spectrum has enough signal-to-noise ratio. All the curves show good qualitative agreement and similar results were found for many other specimens within this range of hydration levels. The fitting parameters for the curves in Figures 3.21 are listed in Table 3.5. The fitting parameters for the curves in Figure 3.22 are just the parameters in Table 3.5 divided by the individual sample density. It is observed from Table 3.5 that the y-intercept \( A_2 \) has a small non-zero value, which is due to the experimental uncertainties at very low frequencies.

Table 3-5. Parameters for fitting Eqn. (3.23) to the absorption curves in Figure 3.21.

<table>
<thead>
<tr>
<th>Sample water content (wt%)</th>
<th>( A_0 )</th>
<th>( A_1 )</th>
<th>( A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>0.9945</td>
<td>4.9305</td>
<td>-0.2323</td>
</tr>
<tr>
<td>32</td>
<td>0.5078</td>
<td>0.9233</td>
<td>0.0039</td>
</tr>
<tr>
<td>19</td>
<td>0.4076</td>
<td>0.4495</td>
<td>-0.0296</td>
</tr>
<tr>
<td>3.6</td>
<td>0.3188</td>
<td>0.0017</td>
<td>-0.0054</td>
</tr>
</tbody>
</table>
One might expect identifiable modes from large molecules like Mb at THz frequencies, but these data show smooth spectra. Recall that the THz absorption spectrum for water vapor consists of sharp lines, while liquid water has a continuous spectrum. A water molecule in the liquid phase experiences a wide range of instantaneous interaction potentials due to the highly variable arrangement of neighboring molecules, leading to an inhomogeneous broadening of the absorption spectrum. Hydrated Mb may have a highly variable arrangement of neighboring water molecules as well, leading to inhomogeneous broadening of the absorption spectrum, but this cannot account for the dehydrated Mb sample at 3.6 wt% water content.

Analysis of Mb absorption must be separated into two parts, the response of an ideal, single molecule of the protein, and the effect of an ensemble of many molecules with a wide range of individual conformations and configurations of neighboring molecules. This idea will be explored further in the next chapter for investigating the roles of biological water in Mb far-infrared absorption. For a single Mb molecule, calculations have been made for a particular set of effective potentials that find approximately 400 normal modes in the spectral region below 3 THz [128]. Since this leads to an average of two modes within 15 GHz of our THz spectrometer resolution, most modes would not be individually resolvable. These calculations do predict that the lowest-lying modes below 0.5 THz will be separated by approximately 30 GHz, which should be resolvable by this spectrometer. The absence of such peaks in the measurements suggests that inhomogeneous broadening may be responsible. While Mb has a well-defined average structure, any real specimen will exhibit significant conformational disorder whereby the orientation of individual amino acid side chains and the number and location of biological water molecules vary from molecule to molecule. In addition, each protein molecule in the powder specimen can be in a slightly different local environment of neighboring protein molecules and biological water molecules, altering the local potential in a manner analogous to
that in liquid water. The observed continuous absorption spectrum for even dehydrated Mb must arise from a combination of high mode density, conformational disorder, and variability in the local protein molecule environments.

Furthermore, in a recent study by S. E. Whitmire et al on the far-infrared spectral measurements of bacteriorhodopsin (BR) thin films using THz-TDS [97], they calculated the far-infrared absorbance of BR using CHARMM based on the density of normal modes (403 modes below 3 THz for BR). Although the calculated spectrum they obtained based on the normal mode analysis show some absorption lines, their experimental results resemble ours in that the far-infrared absorption spectrum of this different protein species shows a broadly continuous nature.

With this discussion in mind, it seems that any hope of resolving the far-infrared absorption characteristics of Mb may rely on cryogenic experiments with crystalline samples, since high state density, conformational disorder, and configurational variability in local environment are believed to be the main causes obscuring the resonance peaks of Mb powders at room temperature. The use of crystalline samples would reduce the local environment issues. Samples recorded under cryogenic condition would suppress the populations in vibrational excited states and reduce the number of possible conformations populated. These all combined may potentially result in a better resolved absorption spectrum.
3.5.3. Absorption Spectra of Myoglobin Aqueous Solutions

The absorption spectra for selected Mb aqueous solutions at 75, 85, and 98 wt% water contents are shown in Figure 3.23, along with the measured liquid water spectrum. All the absorption spectra are obviously dominated by water and show no specific identifiable features, as just discussed in the last subsection. The standard deviations are derived from multiple measurements and shown as error bars at selective frequencies (0.1, 0.6, and 0.9 THz). The relative error is typically within ±5% for all the data points.

The absorption for 98 wt% water content sample shows a slight increase compared with that for pure liquid water but still within experimental uncertainty. This result will be discussed in more detail in the next chapter. Mb solutions at 85 and 75 wt% water contents demonstrate reduced absorption compared with pure water, a result of the substitution of polar water molecules by the relatively non-polar Mb molecules. The detailed effect of these substituting water molecules will be discussed in the next chapter.

It can be seen that in the main frequency range (0.3-0.9 THz) shown in Figure 3.23, the absorption coefficient changes almost linearly with frequency. Instead of using quadratic fitting, as we used for Mb powders, the absorption spectrum of Mb solution is better fit by a straight line given by:

$$\alpha = A_0 \cdot f + A_1,$$

(3.24)

with \(f\) being the frequency and \(A_0\) and \(A_1\) being the fitting parameters. The fit is based on the data points from 0.3 to 0.9 THz, in which the measured absorption spectrum of liquid water agrees well with the published results (Section 3.4.5). The fitting parameters \(A_0\) and \(A_1\) are listed in Table 3.6. The y-intercept \(A_1\) is not zero because we only fit the data points within the range of 0.3-0.9 THz.
As discussed in the last subsection as well as noted in our paper [23], the relatively smooth and broad nature of the absorption spectrum of hydrated Mb is likely due to an inhomogeneous broadening effect which results from the highly variable arrangement of the water molecules in the close vicinity of Mb molecule.

Similarly, we can normalize the absorption coefficient of Mb solutions using Eqn. (3.21) and obtain the absorption ability per cross sectional area of sample. The spectra are plotted in Figure 3.24, along with the error bars derived using the same principle as Mb powders in Section 3.5.2. Because the density of the Mb solution doesn't change much with water content (Table 3.4), the absorption spectra in Figure 3.24 have very similar behavior with the curves shown in Figure 3.23.

<table>
<thead>
<tr>
<th>Sample water content (wt%)</th>
<th>$A_0$</th>
<th>$A_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>14.14</td>
<td>8.98</td>
</tr>
<tr>
<td>98</td>
<td>15.02</td>
<td>8.96</td>
</tr>
<tr>
<td>85</td>
<td>14.90</td>
<td>7.23</td>
</tr>
<tr>
<td>75</td>
<td>3.24</td>
<td>6.86</td>
</tr>
</tbody>
</table>
Figure 3-23. The absorption spectra of Mb solutions at different water concentrations. The straight line is a linear fit for each sample based on the frequency range with good signal-to-noise ratio (0.3-0.9 THz). The error bars shown at 0.3, 0.6, and 0.9 THz are the standard deviations derived from multiple measurements.
Figure 3-24. The absorption spectra of Mb solutions normalized to sample density at different water concentrations. The straight line is a linear fit for each sample based on the frequency range with good signal-to-noise ratio (0.3-0.9 THz). The error bars shown at 0.3, 0.6, and 0.9 THz are the corresponding error bars in Figure 3.23 divided by the sample density, based on the principle presented in Section 3.5.2.
3.5.4. Comparing Myoglobin Powder Absorption Spectrum with Absorption Spectrum of Water Vapor

As can be seen in Figures 3.21 and 3.22 and discussed in Section 3.5.2, the absorption spectrum of Mb powder is lacking any characteristic peaks that can be used as an identification of the sample under study. This point is made more clearly in Figure 3.25, which shows the comparison between the absorption spectrum of water vapor and a typical Mb powder sample (19.0 wt% water content).

In the far-infrared absorption spectrum of water vapor, the absorption lines due to vibrational and rotational motions of water molecules are obvious and can serve as a fingerprint, for example, to identify the presence of water vapor in a propane-air flame [109]. Similar features are not seen with Mb powder. While a number of authors have investigated the potential of far-infrared identification of specific biomolecules [90, 91, 93-96, 100, 129], in which the size of the molecules are smaller than the protein we studied here, our results rule out this possibility for powdered Mb. Similar results can be expected for other materials of this type.
Figure 3-25. Comparison of absorption spectra from water vapor and a typical Mb sample (19.0 wt% water content). Water vapor has sharp, assignable lines that can be used to identify its presence, while biological water in protein has a smooth profile without identifiable characteristics.
3.5.5. Comparison with FTIR Measurements

To help put the THz absorption data on an absolute scale, selected samples were measured using a conventional FTIR spectrometer equipped with a composite silicon bolometer operating at 1.8 K [122] to extend the measurement range down to 1.0 THz. The same specimens and sample holders used in the THz-TDS measurements were used without modification for the FTIR measurements. For all samples, the absorption spectra obtained from two techniques are in consistent within 20%. Results for powdered Mb with 32 wt% water content are shown in Figure 3.12. The resolutions of the THz measurements and FTIR measurements are 0.5 cm\(^{-1}\) (15 GHz) and 1 cm\(^{-1}\) (30 GHz), respectively. The systematic errors are potentially somewhat larger for FTIR in the very low frequency range, but where comparable the FTIR spectra always show the same continuous spectra seen with THz-TDS. The good agreement with FTIR confirms the quantitative validity of THz-TDS, and that the continuous spectra are not artifacts of the THz spectrometer.
Figure 3-26. Comparison of absorption spectra obtained from THz and FTIR spectrometers, for an Mb powder with 32.0 wt% water content. FTIR displays Fabry-Perot oscillations and increased uncertainty at low frequency due to weak source intensity, and requires normalization by a factor of 1.18 to coincide with the THz spectrum.
3.6. Conclusions

The absorption spectra of Mb powders and solutions at a series of carefully controlled hydration levels were presented in this chapter. All the spectra are shown to be smooth in the frequency range studied, without identifiable features intrinsic to the sample.

It was also shown that the water content in Mb dominates its absorption ability, making the absorption curve analogous to that of liquid water. However, the driest Mb powder sample with essentially no water also shows the same smooth nature, raising questions about the possible mechanisms. It is believed that high state density, conformational disorder, and configurational variability in the local environment of Mb molecule are the main causes obscuring the resonance peaks of Mb powder sample, if there are any at room temperature.

Furthermore, a water molecule in the liquid phase experiences a wide range of instantaneous interaction potentials due to the highly variable arrangement of neighboring molecules, leading to an inhomogeneous broadening of the absorption spectrum. Hydrated Mb may have a highly variable arrangement of neighboring water molecules as well, leading to inhomogeneous broadening of the absorption spectrum.

The significance of water molecules in the absorption ability of Mb molecules will be studied more carefully in the next chapter, first based on an ideal model in which the two constituents are taken to be non-interacting, and then a more sophisticated model is constructed to include the interactions between water and Mb molecules.
CHAPTER 4. BIOLOGICAL WATER AND HYDRATION EFFECT

In the previous chapter, the absorption spectra of powdered Mb samples as well as Mb aqueous solutions at several hydration levels were presented and the features of the spectra discussed. We found that the Mb absorption spectrum is essentially smooth within the whole THz frequency range and water contents in the sample contribute significantly to the Mb far-infrared absorption properties.

Water molecules in the vicinity of a biomolecule, so-called “biological water” in contrast to bulk water, play several crucial roles. To maintain the biological activity and functionality of biomaterials like proteins and DNA, the existence of water is indispensable [4-6]. On the other hand, proteins themselves influence both the spatial and dynamical arrangement of the neighboring water layers [130-132].

In this Chapter, we will first take a look at the general roles of biological water in large biomolecules. The basic physics behind the absorption of liquid water in the THz region will also be investigated. We then discuss the effect of biological water on the far-infrared absorption properties of Mb molecule. The discussions will be based on the systematic measurement of the far-infrared absorption spectra of Mb powders and solutions presented in the last chapter. An ideal non-interacting model is first proposed to demonstrate the enhancement of Mb molecule absorptivity in Mb-water mixture. Then a more sophisticated interacting model is constructed to explain the Mb molecule absorptivity increase upon hydration.
4.1. Biological Water in Protein

Proteins require water to function, so the existence of water in protein and the role of water in protein dynamics has long been a focus of interest [4-7]. The water molecules in the immediate vicinity of a biological macromolecule, such as protein and DNA, play a crucial role in determining their structure, dynamics and functionality [133-136]. Conversely, it has been demonstrated both theoretically [137-139] and experimentally [140, 141] that the biological macromolecule also has a significant influence on the spatial arrangement and dynamical behavior of the surrounding water layers. The region of water in the immediate vicinity of a protein surface, extending through one to two layers of water molecules, is composed of “biological water,” so-called because of perturbations to the spatial arrangement, dynamics, and other physical attributes compared to bulk water [142, 143]. Shown in Figure 4.1 is a schematic of an Mb molecule and its surrounding water layers. The Mb molecule is approximated to be a sphere with diameter of around 3.0 nm and the outside biological water extends to about 0.35 nm [115].

Depending on the momentary state of existence, the biological water can be loosely classified into two constantly exchanging categories: internal water and peripheral water. Internal water refers to the few water molecules that remain strongly bound to the protein structure for a long time via multiple hydrogen bonds. Peripheral water is the water molecules in the immediate vicinity of the protein surface which experience much faster rotational and translational motions than the internal water [6, 142], but not as much as bulk water.

The existence of internal and peripheral water is essential to the protein molecule in many aspects. Bound with the protein backbone through a hydrogen bond network, the internal water is an integral part of the protein structure [144-146]. These water molecules can be identified crystallographically. It is also widely
recognized that a minimum threshold of water is required to fully activate the dynamics and functionality of most proteins [5, 138]. For example, enzymatic activity typically requires a water content of about 20% by weight [6, 7]. Some studies have indicated that the presence of water enhances the flexibility of protein molecules [144, 145]. Dielectric measurements of lysozyme [144, 145] and cytochrome c [145] powders showed an increased polarizability of the proteins due to water, which was described as a “plasticizer” that enables a larger dipole response to external electromagnetic radiation.

The interactions between biological macromolecules and water have been studied with various structural and dynamical probes such as X-ray and neutron scattering [19, 21, 22], NMR spectroscopy [12-14, 16], dielectric spectroscopy [8-11], and FTIR spectroscopy [147, 148]. For Mb in particular, there has been detailed dielectric characterization of protein hydration obtained by E. H. Grant et al [115] on the basis of experimental studies of the dynamic behavior of water near biomolecules. It was established that the frequency dependent dielectric constant of the Mb-water mixture system can be modeled as a sum of four dispersion terms as follows:

$$
\varepsilon(\omega) = \varepsilon_\infty + \frac{\Delta_1}{1 + i\omega \tau_1} + \frac{\Delta_2}{1 + i\omega \tau_2} + \frac{\Delta_3}{1 + i\omega \tau_3} + \frac{\Delta_4}{1 + i\omega \tau_4},
$$

where \( \varepsilon_\infty \) denotes the infinite frequency dielectric constant of bulk water, \( \Delta_i \) is the relative weight of a given relaxation type, and \( \tau_i \) is the respective time constant [142]. The values for \((\Delta_i, \tau_i)\) are listed in Table 4.1 for information [115].

The first dispersion term \((\Delta_1, \tau_1)\) represents the orientational relaxation of Mb molecule. Analysis of this dispersion gives an effective dipole moment of about
150 debye (1 debye=3.34×10^{-30}C·m) for the Mb molecule and a hydration shell of from one to two water molecules thickness (0.6 g water/g protein) [149]. The second and third terms, (Δ_2, τ_2) and (Δ_3, τ_3), correspond to the relaxation of biological water associated with the protein. With Δ_2 and Δ_3 having nearly equal values, they are interpreted as each being associated with orientational relaxation of one half of the total protein hydration [150]. Finally, the fourth term (Δ_4, τ_4) accounts for the orientational relaxation of bulk water.

Table 4-1. Relative weight and time constant for the four dispersion terms in Eqn. (4.1) [115].

<table>
<thead>
<tr>
<th>Dispersion relative weight</th>
<th>Time constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ_1</td>
<td>18.4</td>
</tr>
<tr>
<td>Δ_2</td>
<td>5.3</td>
</tr>
<tr>
<td>Δ_3</td>
<td>5.4</td>
</tr>
<tr>
<td>Δ_4</td>
<td>57.6</td>
</tr>
</tbody>
</table>

In this thesis, we report on the use of THz-TDS, an extension of FTIR spectroscopy to millimeter-scale wavelengths, to study the protein hydration effects in Mb. A relatively recent technique, THz-TDS exploits advances in pulsed THz sources and detectors that allow for absorption measurements up to several THz to be done with significantly improved sensitivity compared to earlier methods [24]. Since water’s large permanent dipole moment causes it to dominate absorption of THz radiation in most biological systems, this technique is especially well-suited to observing protein-induced polarization changes in biological water [25].
Figure 4-1. Schematic of spherical Mb molecule and the surrounding water layers. The diameter of the Mb sphere is approximated to be 3.0 nm and the outside water layer extends to about 0.35 nm.
4.2. Basic Physics of Liquid Water Absorption in the THz Region

Before studying the effect of biological water on the far-infrared absorption properties of Mb, it is illuminating to investigate the basic physics of the absorption by liquid water molecules in the THz region.

Water is the vital substance which supports and mediates biochemical reactions. It is considered one of the most important substances in this planet for the living because it possesses many unusual equilibrium and dynamical properties compared with the majority of liquids [151]. It is widely accepted that the anomalous properties of liquid water to a great extent are caused by the presence of hydrogen bonds. A water molecule can form up to four hydrogen bonds to other water molecules thereby creating the possibility for a local tetrahedral structure around the molecule. Figure 4.2 below is a schematic of this tetrahedral hydrogen bond structure [152].

![Figure 4-2. Schematic of the local tetrahedral structure in liquid water in which one water molecule forms four hydrogen bonds with neighboring molecules (Picture from Ref. [152]).](image-url)
One simple example of this hydrogen bond network influencing the property of water molecule is the effective dipole moment of liquid water. It is well-established that the permanent dipole moment of an isolated water molecule in the vapor phase is $1.84 \pm 0.02$ debye [151]. While in bulk water, there are no isolated water molecules. Each water molecule is surrounded by four others with which hydrogen bonds are formed. On applying an electric field, the neighboring dipoles have to reorient dependently, because the molecular rotations must be cooperative to preserve all hydrogen bonds. Due to these molecular association and correlation effects, the effective dipole moment of bulk water increases by a significant amount to around 2.5 debye [6]. This effective dipole moment is actually a measurement of the response of bulk liquid water molecules to the applied electric field.

The energy required to rupture one hydrogen bond in water is about 0.11 eV and the average lifetime of a hydrogen bond is on the picosecond time scale (THz frequency range) [72]. So it is the interplay between the formation and rupture of hydrogen bonds that causes the intermolecular dynamics on a picosecond time scale in liquid water which gives rise to resonance and relaxation processes in the THz and microwave spectral ranges. In the spectral region below 50 cm$^{-1}$ (1.5 THz), molecular dynamics simulation shows the calculated absorption spectrum is dominated by reorientation of permanent and induced dipole moments in the hydrogen bonded network of water molecules [153].

Theoretically, the following formula is widely used in molecular dynamics simulations to calculate the infrared and far-infrared absorption coefficient $\alpha(\omega)$ of liquid water with volume $V$ at temperature $T$ [153, 154]:

$$\alpha(\omega) = \frac{4\pi \omega \tanh(\beta \omega/2)}{3 \hbar(\omega)c V} \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle M(t) \cdot M(0) \rangle.$$

(4.2)
where \( n(\omega) \) is the refractive index of liquid water at angular frequency \( \omega \), \( \beta = 1/kT \) with \( k \) being the Boltzmann constant, \( h = h/2\pi \) is the Planck constant, \( C_T(t) = \langle M(t) \cdot M(0) \rangle \) is called the total dipole correlation function with \( M \) being the total dipole moment of the sample, which is the sum of the microscopic dipoles occurring in liquid water.

Molecular dynamics simulations [153] showed that the total dipole moment correlation function \( C_T(t) \) contains 70.6\% contribution from the permanent dipole moment of water molecule (1.84 debye) and 2.9\% contribution from the dipole induced dipole moment. The remaining 26.5\% comes mainly from an interference term correlating the permanent dipole moment of one molecule with the induced moments of its neighbors. It is also interesting to note that over 2/3 of the total dipole correlation function, and consequently the absorption spectrum, arises from collective contributions involving more than one water molecule [68]. The detailed balance of the permanent and induced dipole moments and specially their interference give a good quantitative agreement with the experimentally observed spectra.
4.3. Myoglobin Far-Infrared Absorption in a Non-interacting Model

The absorption spectra for Mb powders and aqueous solutions have been discussed in Chapter 3. In this section, we are trying to understand the absorption properties of Mb-water mixture in terms of an ideal non-interacting model. For convenience, Figure 3.22 in Section 3.5.2 and Figure 3.23 in Section 3.5.3 are copied in Figure 4.3. When inspecting Figure 4.3, it should be noticed that the frequency range plotted is different for Mb powders and solutions. Interestingly, it is also noticed from Figure 4.3 that the normalized absorption coefficients for liquid water, 98 wt% Mb solution, and 42 wt% Mb powder are roughly the same within error bars; this may well be a coincidence.

In Figure 4.3, each absorption spectrum for Mb powder is fit by a quadratic curve, based on the data points from 0.1 to 1.2 THz. Each absorption spectrum for Mb solution is better fit by a straight line, based on the data points from 0.3 to 0.9 THz. In the following analysis, absorption coefficients are taken from the fitting curves instead of the original spectra, in order to average over experimental uncertainties and remnant oscillations in the main frequency region.

Now, let's ask the question, what is the effect of water content on the absorption properties of Mb-water system? To answer this question, we should first plot the absorption coefficient of Mb-water mixture as a function of water content, at a fixed frequency.

Using the fitting curves from Figure 4.3 to describe the Mb-water system absorption measurements, we replot the results for fixed frequencies of 0.35, 0.5, and 0.8 THz in Figure 4.4. The error bar on each data point is the same error bar presented on Figure 4.3, and thus is the standard deviation of the total absorption coefficient derived from multiple measurements divided by the sample density (see Section 3.5.2). Also included in Figure 4.4 is a dotted straight line. It
corresponds to an ideal model, in which the two composing materials in Mb-water system, Mb molecules and liquid water molecules, are assumed to be non-interacting. That is, the absorption coefficient $\alpha$ (unit in mm$^{-1}$) of the mixture is given by the linear combination of the two constituents,

$$\alpha = \sigma_{\text{pure H}_2\text{O}} M_{\text{H}_2\text{O}} \sigma_{\text{Dry Mb}} M_{\text{Mb}}, \quad (4.3)$$

where $\sigma_{\text{pure H}_2\text{O}}$ and $\sigma_{\text{Dry Mb}}$ are the molar absorptivity of pure liquid water and dry Mb, respectively, defined as:

$$\sigma_{\text{pure H}_2\text{O}} = \frac{\alpha_{\text{pure H}_2\text{O}}}{M_{\text{pure H}_2\text{O}}}, \quad (4.4)$$

$$\sigma_{\text{Dry Mb}} = \frac{\alpha_{\text{Dry Mb}}}{M_{\text{Dry Mb}}}, \quad (4.5)$$

in which $\alpha_{\text{pure H}_2\text{O}}$ and $\alpha_{\text{Dry Mb}}$ are the absorption coefficients for pure liquid water and dry Mb and are considered to be constants at a given frequency. $M_{\text{pure H}_2\text{O}}$ and $M_{\text{Dry Mb}}$ are the nominal molar concentrations of pure liquid water and dry Mb and are of course constants.

In Eqn. (4.3), $M_{\text{H}_2\text{O}}$ and $M_{\text{Mb}}$ are the fractional molar concentrations (number of moles per unit volume) of water and Mb molecules, respectively. Recall the definition of $M_{\text{Mb}}$ in Eqn. (3.14)

$$M_{\text{Mb}} = \frac{(1-x)m}{V \cdot MW_{\text{Mb}}} = \frac{(1-x)\rho}{MW_{\text{Mb}}}, \quad (4.6)$$

similarly, $M_{\text{H}_2\text{O}}$ is defined as:
\[ M_{H_2O} = \frac{xm}{V \cdot MW_{H_2O}} = \frac{x \rho}{MW_{H_2O}}, \]  

(4.7) 

where \( x \) is the sample water content in weight percentage, \( m \) is the total mass of the sample, \( V \) is the sample volume, \( \rho \) is the sample density, \( MW_{Mb} = 16952 \text{ g/mol} \) and \( MW_{H_2O} = 18 \text{ g/mol} \) are the molecular weight for Mb and water, respectively.

Substituting Eqns. (4.6) and (4.7) into Eqn. (4.3), we obtain

\[ \alpha = \sigma_{Pure_{H_2O}} \frac{x \rho}{MW_{H_2O}} + \sigma_{Dry_{Mb}} \frac{(1-x) \rho}{MW_{Mb}}. \]  

(4.8) 

Notice \( \alpha/\rho \) gives the normalized areal absorption coefficient \( \alpha_{\text{Area}} \) (unit in \((\text{mg/mm}^2)^{-1}\)) defined in Eqn. (3.21), Eqn. (4.8) can be rewritten as:

\[ \alpha_{\text{Area}} = \frac{\alpha}{\rho} = \sigma_{Pure_{H_2O}} \frac{x}{MW_{H_2O}} + \sigma_{Dry_{Mb}} \frac{(1-x)}{MW_{Mb}}. \]  

(4.9) 

This is a linear function of the water content \( x \) and gives the dotted straight line in Figure 4.4.

To evaluate Eqn. (4.9), the molar absorptivity of pure liquid water \( \sigma_{Pure_{H_2O}} \) and dry Mb \( \sigma_{Dry_{Mb}} \) are required, which equivalently means we need the absorption coefficients for liquid water and dry Mb at various frequencies. This information is listed in Table 4.2 for selected frequencies at 0.35, 0.5, and 0.8 THz. Aside from the ordinary absorption coefficient \( \alpha \), the absorption coefficients are also expressed in normalized areal absorption coefficient \( \alpha_{\text{Area}} = \alpha/\rho \) and molar
absorptivity $\sigma = \alpha / M$ with $\rho$ and $M$ being the sample density and nominal molar concentration, respectively. The relation between $\rho$ and $M$ is given by Eqns. (4.6) and (4.7) while $x = 0$ is for dry Mb and $x = 1$ for pure liquid water. The densities for Mb powders at 1.2 (dry Mb), 3.6, 19, 32, 42 wt% water content from Table 3.3 and the densities for Mb solutions from Table 3.4 are copied in Table 4.3 for convenience. Also included in Table 4.3 are the fractional molar concentrations of water and Mb in the Mb-water mixture samples.

In Table 4.2, the liquid water absorption coefficients are taken from the published values [125]. At frequencies where the data for liquid water are not available from Ref. [125], cubic spline interpolation is used to obtain the absorption coefficients. For Mb, we used values obtained from our THz-TDS measurement of Mb powder containing about 1 wt% water, achieved by lengthy drying in a desiccator. The relative error in the absorption coefficients of dry Mb in Table 4.2 is around $\pm 5\%$ (Section 3.4.5). It is noticed from Table 4.2 that the molar absorptivity of dry Mb is about two orders of magnitude larger than liquid water. This is because of the much larger molecular size of Mb than water.

<table>
<thead>
<tr>
<th>Frequency (THz)</th>
<th>Liquid water [125]</th>
<th>Dry Mb (this work)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$ (mm$^{-1}$)</td>
<td>$\alpha_{\text{Area}}$ (mg/mm$^2$)$^{-1}$</td>
</tr>
<tr>
<td>0.35</td>
<td>14.0</td>
<td>14.0</td>
</tr>
<tr>
<td>0.5</td>
<td>16.0</td>
<td>16.0</td>
</tr>
<tr>
<td>0.8</td>
<td>20.4</td>
<td>20.4</td>
</tr>
</tbody>
</table>

Table 4-2. Absorption coefficients of liquid water and dry Mb at selected frequencies. The absorption coefficients are described in different units, accounting for the density and molar concentration of liquid water and dry Mb.
Table 4-3. Density, Mb molar concentration, and water molar concentration for Mb powders and solutions at selected water contents. The liquid water (100 wt%) and dry Mb (1.2 wt%) are also included. The molecular weight of Mb and water are given by $MW_{Mb} = 16952 \text{ g/mol}$ and $MW_{H_2O} = 18 \text{ g/mol}$.

$$M_{Mb} = \frac{(1-x)\rho}{MW_{Mb}} \quad \text{(m-mol/mm}^3\text{)}$$

$$M_{H_2O} = \frac{x\rho}{MW_{H_2O}} \quad \text{(m-mol/mm}^3\text{)}$$

<table>
<thead>
<tr>
<th>Sample water content $x$ (wt%)</th>
<th>Sample density $\rho$ (mg/mm$^3$)</th>
<th>Mb molar concentration $M_{Mb}$</th>
<th>Water molar concentration $M_{H_2O}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 (Liquid water)</td>
<td>1.00</td>
<td>N/A</td>
<td>$5.56 \times 10^{-2}$</td>
</tr>
<tr>
<td>98</td>
<td>1.00</td>
<td>$1.18 \times 10^{-6}$</td>
<td>$5.44 \times 10^{-2}$</td>
</tr>
<tr>
<td>95</td>
<td>1.01</td>
<td>$2.97 \times 10^{-6}$</td>
<td>$5.33 \times 10^{-2}$</td>
</tr>
<tr>
<td>90</td>
<td>1.03</td>
<td>$6.08 \times 10^{-6}$</td>
<td>$5.15 \times 10^{-2}$</td>
</tr>
<tr>
<td>85</td>
<td>1.03</td>
<td>$9.11 \times 10^{-6}$</td>
<td>$4.86 \times 10^{-2}$</td>
</tr>
<tr>
<td>80</td>
<td>1.04</td>
<td>$12.2 \times 10^{-6}$</td>
<td>$4.62 \times 10^{-2}$</td>
</tr>
<tr>
<td>75</td>
<td>1.01</td>
<td>$14.9 \times 10^{-6}$</td>
<td>$4.21 \times 10^{-2}$</td>
</tr>
<tr>
<td>70</td>
<td>1.02</td>
<td>$18.1 \times 10^{-6}$</td>
<td>$3.97 \times 10^{-2}$</td>
</tr>
<tr>
<td>42</td>
<td>0.220</td>
<td>$7.47 \times 10^{-6}$</td>
<td>$5.13 \times 10^{-3}$</td>
</tr>
<tr>
<td>32</td>
<td>0.082</td>
<td>$3.29 \times 10^{-6}$</td>
<td>$1.46 \times 10^{-3}$</td>
</tr>
<tr>
<td>19</td>
<td>0.076</td>
<td>$3.61 \times 10^{-6}$</td>
<td>$8.02 \times 10^{-4}$</td>
</tr>
<tr>
<td>3.6</td>
<td>0.058</td>
<td>$3.29 \times 10^{-6}$</td>
<td>$1.16 \times 10^{-4}$</td>
</tr>
<tr>
<td>1.2 (Dry Mb)</td>
<td>0.061</td>
<td>$3.55 \times 10^{-6}$</td>
<td>$4.07 \times 10^{-5}$</td>
</tr>
</tbody>
</table>
Figure 4-3. Normalized absorption spectra of Mb powders (top) and solutions (bottom) copied from Section 3.5. The absorption spectrum for Mb powder is fit by a quadratic curve, while the absorption spectrum for Mb solution is better fit by a straight line. The values on the fitting curve will be used as the basis for the discussions in the following sections, instead of using the original data.
Figure 4-4. Normalized areal absorption coefficient of Mb-water mixtures as a function of water content at selected frequencies (0.35, 0.5, and 0.8 THz). The absorption coefficients used here for each sample are taken from the fitting curves presented in Figure 4.3, in order to average over any uncertainties in the measurements. The measured data points are denoted by open circles with error bar on each data point. The solid lines connecting the high and low water contents data points are used to guide the eyes. The gap left between 42 and 70 wt% samples indicates that we don't have data points for the intermediate water contents. The dotted straight lines are the results calculated based on an ideal model in which the Mb molecules and water molecules are assumed to be two non-interacting species. It can be seen that increased absorption is observed for many of the samples measured at all the selected frequencies, as compared with the ideal model.
In Figure 4.4, the data points between 42 wt% and 70 wt% are absent. This is due to an apparent miscibility gap in the solubility of Mb in water. As we described in the sample preparation section in the last chapter, the Mb powder converts into a dense liquid for water content beyond 42 wt% so that we failed to obtain samples with water content greater than that. For Mb solutions with water content beyond 80 wt%, the samples appear to be uniform. For the specimens at 75 wt% and 70 wt% water, there was some visual evidence of undissolved Mb. However, the sizes of the possible suspending Mb particles are much smaller than the typical wavelength of the THz radiation (0.3 mm) so that the scattering loss is not a major effect here. Furthermore, the results of the subsequent analysis are not materially affected even if we leave out these two samples.

Deviations of the normalized areal absorption coefficient (Figure 4.4) from the non-interacting model (dotted straight line) increase systematically with increasing frequency, showing enhanced THz absorption. This is contrary to our expectation, because substantial evidence exists that biological water has a reduced effective dipole moment and hence would absorb less than bulk water.

Dielectric measurements by S. Bone et al [155] at 9.95 GHz for bovine serum albumin (BSA), cytochrome c, and lysozyme powders for water contents up to 19 wt% indicated that the dielectric absorption at this frequency due to water occupying the primary monolayer sorption sites is negligible because the water molecules are too rotationally hindered. It was concluded that only the water molecules bound to these proteins in the secondary hydration layers are able to contribute significantly to the dielectric absorption at this frequency, but these water molecules are much more tightly bound to the proteins than were originally expected so that the absorption is still not comparable to free bulk water. Combined these results with their later results on the dielectric properties of hydrated BSA for frequencies up to 100 kHz [156], they were able to deduce the effective dipole moment of protein-bound water. Calculations suggest that the
effective dipole moment for water molecule falls from 2.5 debye in the bulk water state to about 0.8 debye for the region of biological water around a protein, which indicates that the water molecules in the primary sorption sites of the protein are rotationally hindered to a significant extent [6, 157]. The 0.8 debye effective dipole moment of protein-bound water represents a reduced response of water molecules to applied electric field compared with free bulk water, which will show up as an absorption decrease in our case. Adding this known effect to the non-interacting model should lead to suppression, not enhancement of the absorption for the whole system.

Based on the analysis above, it is clear that the non-interacting model doesn't fit our measurements and a more sophisticated model is required to explain the discrepancy. This is the subject of the next section.
4.4. Myoglobin Far-Infrared Absorption in an Interacting Model

The deviations in Figure 4.4 between the measured absorption coefficient and the expected absorption coefficient from the ideal non-interacting model indicate that there must be interactions between water molecules and Mb molecules. We need a model which takes these interactions into account explicitly.

To investigate the possible Mb-water interactions, an important question to ask will be how the absorption per Mb molecule changes due to the interactions. To answer this question, it is important to understand the properties of water molecules at different hydration levels. For the driest Mb powders with water contents below 4 wt%, the only water molecules would be those few (less than 40 water molecules per Mb molecule) entrained in the protein interior (structural water) and would contribute negligibly to THz absorption. As water content increases, more water molecules would be at protein surfaces and would function as biological water which starts to absorb THz radiation but still contributes significantly less than its bulk counterpart (Section 4.3). The bulk phase water molecules would not appear until the whole surface of the protein is covered and the hydration reaches its saturation level (around 0.6 g water/g Mb). For Mb dissolved in liquid water, only the fraction corresponding to one or two water layers (less than 0.5 nm) surrounding each protein molecule would be biological water, with the remainder expected to simply be bulk water. We will construct a model describing these species of water and use it to extract the effective absorption per protein molecule from the data.

The interacting model makes the following specific assumptions for the amount of biological water as well as its absorption:

(i) The saturation hydration level is 0.6 g water/g Mb [115], which is about 37.5 wt% of water content in our scale. That is, each gram of Mb can
only accommodate up to 0.6 grams of biological water; any excessive amount of water will be treated as bulk water.

(ii) The very dry specimen with 3.6 wt% water will be treated as Mb alone. That is, the first 3.6 wt% of water is non-absorbing since all the water molecules are basically confined within the protein structure.

(iii) For the powder specimens with intermediate water contents of 19 and 32 wt%, all of the water is biological water based on the first assumption. From our knowledge about the dielectric response of protein-bound water [6, 155-157], we further assume water absorbs at 10 and 20% of the bulk water value for the 19 and 32 wt% specimens.

(iv) For the samples with water contents above the saturation hydration level of 37.5 wt%, including the most hydrated powder (42 wt%) and all the solutions, the portion of biological water can be calculated based on the available Mb and the rest will be treated as bulk water. We further assume the biological water absorbs at 30% of the bulk water value, corresponding to a 0.8 debye effective dipole moment (Section 4.3). At this hydration level, there will be one to two layers of water surrounding the Mb molecule, making the protein surface totally covered by water molecules (Figure 4.1).

(v) All the water beyond the saturation hydration level absorbs as bulk water, corresponding to a 2.5 debye dipole moment.

While this model may seem quite arbitrary in quantities, especially the third and fourth assumptions for the absorption of biological water, our qualitative conclusions would not depend on the choice of model. We will include an investigation on the model-dependence of the results in the last part of this section.
Based on the assumptions made above, the total absorption coefficient of Mb-water system $\alpha$ is thus modeled as the linear combination of three terms, as opposed to the two-term combination in Eqn. (4.3) for the ideal model:

$$\alpha = \sigma_{\text{Mb}} M_{\text{Mb}} + \sigma_{\text{Free H}_2\text{O}} M_{\text{Free H}_2\text{O}} + \sigma_{\text{Bio H}_2\text{O}} M_{\text{Bio H}_2\text{O}},$$

(4.10)

where $\sigma_{\text{Mb}}$, $\sigma_{\text{Free H}_2\text{O}}$, and $\sigma_{\text{Bio H}_2\text{O}}$ are the molar absorptivity of Mb, bulk free water, and biological bound water, respectively. $M_{\text{Mb}}$, $M_{\text{Free H}_2\text{O}}$, and $M_{\text{Bio H}_2\text{O}}$ are the fractional molar concentrations of Mb, bulk free water and biological bound molecules water, respectively. Eqn. (4.10) applied to a given frequency. For selected frequencies at 0.35, 0.5, and 0.8 THz, $\sigma_{\text{Free H}_2\text{O}}$ will take the values in Table 4.2 for liquid water and $\sigma_{\text{Bio H}_2\text{O}}$ will be calculated according to the assumptions made above as a function of water concentration. $M_{\text{Mb}}$ is calculated using Eqn. (4.6) and assumes the values in Table 4.3. Based on the first assumption, the first 0.6 grams of water per gram of Mb is taken as biological water and the free bulk water is the amount beyond that, therefore $M_{\text{Free H}_2\text{O}}$ and $M_{\text{Bio H}_2\text{O}}$ can be calculated and they should add up to the value calculated from Eqn. (4.7) as listed in Table 4.3.

$M_{\text{Free H}_2\text{O}}$ and $M_{\text{Bio H}_2\text{O}}$ as a function of water content are plotted in Figure 4.5. It can be seen that there is no free water for water contents up to around 40 wt%, and then the free water starts to increase. For Mb solutions, the free water molar concentration increases linearly with water content and the biological water molar concentration decreases linearly because of the decreasing amount of available Mb.
Figure 4-5. Molar concentration of free water and biological water as a function of water content. It can be seen that there is no free water for water contents up to around 40 wt%. For Mb solutions, the free water molar concentration increases linearly with water content and the biological water molar concentration decreases linearly because of the decreasing amount of available Mb.
Rearrangement of Eqn. (4.10) will give the molar absorptivity of Mb at a given frequency as a function of water content:

\[
\sigma_{\text{Mb}} = \frac{\alpha - \sigma_{\text{Free, } H_2O} M_{\text{Free, } H_2O} - \sigma_{\text{Bio, } H_2O} M_{\text{Bio, } H_2O}}{M_{\text{Mb}}}.
\]  

(4.11)

The resulting molar absorptivity of Mb, using the progressively increasing absorption of water described above, is plotted in Figure 4.6 as a function of water content and frequency. The baseline absorption of dry Mb (1.2 wt% water) is shown as the dashed line; note that adding water increases the Mb molar absorptivity by more than an order of magnitude compared with dry Mb.

Also shown in Figure 4.6 on each data point is the error bar. It is very important to understand how these error bars are obtained and the sources of uncertainty in \(\sigma_{\text{Mb}}\) since excessive error on the data point could potentially prevent the correct interpretation of the data and obscure the important results [158]. We will perform a detailed error analysis as follows.
Figure 4-6. Molar absorptivity of Mb as a function of water content at frequencies of 0.35, 0.5, and 0.8 THz. The closed circles with error bars connected by solid line are the actual data points. The dashed line denotes the absorptivity of dry Mb as a base level. An enhanced molar absorption is seen throughout, with a dramatic increase at the dilute limit.
In Eqn. (4.11), the free water absorption $\sigma_{\text{Free, H}_2\text{O}}$ and the biological water absorption $\sigma_{\text{Bio, H}_2\text{O}}$ assume the modeled values, so that any uncertainties in them will be systematic and tend to influence the entire data sets in a consistent way. The main sources of random error result from the measurements of $\alpha$, $M_{\text{Mb}}$, $M_{\text{Free, H}_2\text{O}}$, and $M_{\text{Bio, H}_2\text{O}}$. Based on the error analysis theory, the standard deviation of $\sigma_{\text{Mb}}$, $\Delta\sigma_{\text{Mb}}$, can be written as ($\sigma_{\text{Free, H}_2\text{O}}$ and $\sigma_{\text{Bio, H}_2\text{O}}$ are assumed to be constants):

$$\Delta\sigma_{\text{Mb}} = \sqrt{\left(\frac{\Delta\alpha}{M_{\text{Mb}}}\right)^2 + \left(\frac{\Delta M_{\text{Mb}}}{M_{\text{Mb}}}\right)^2 \sigma_{\text{Mb}}^2 + \left(\frac{\Delta M_{\text{Free, H}_2\text{O}}}{M_{\text{Mb}}}\right)^2 \sigma_{\text{Free, H}_2\text{O}}^2 + \left(\frac{\Delta M_{\text{Bio, H}_2\text{O}}}{M_{\text{Mb}}}\right)^2 \sigma_{\text{Bio, H}_2\text{O}}^2},$$

(4.12)

where $\Delta\alpha$, $\Delta M_{\text{Mb}}$, $\Delta M_{\text{Free, H}_2\text{O}}$, and $\Delta M_{\text{Bio, H}_2\text{O}}$ are the standard deviations of $\alpha$, $M_{\text{Mb}}$, $M_{\text{Free, H}_2\text{O}}$, and $M_{\text{Bio, H}_2\text{O}}$, respectively.

In this thesis work, the measurements showed that

$$\frac{\Delta\alpha}{M_{\text{Mb}}} \sim 15\% \sigma_{\text{Mb}},$$

(4.13)

and the last three terms all have the same order of magnitude,

$$\frac{\Delta M_{\text{Mb}}}{M_{\text{Mb}}} \sigma_{\text{Mb}} \sim \frac{\Delta M_{\text{Free, H}_2\text{O}}}{M_{\text{Mb}}} \sigma_{\text{Free, H}_2\text{O}} \sim \frac{\Delta M_{\text{Bio, H}_2\text{O}}}{M_{\text{Mb}}} \sigma_{\text{Bio, H}_2\text{O}} < 3\% \sigma_{\text{Mb}}.$$

(4.14)

Obviously, the uncertainty in $\sigma_{\text{Mb}}$ is dominated by the experimental error in the total absorption coefficient $\alpha$ (94% contribution), so that $\Delta\sigma_{\text{Mb}}$ can be approximated as:
which is the error bar shown on each data point in Figure 4.6.

In Figure 4.6, the error bars generally increase with frequency and water content. This is because the standard deviation in the absorption coefficient increases with frequency and the molar concentration of Mb decreases with water content.

In Figure 4.6, the absorption per Mb molecule shows two clear trends: a general enhanced absorption relative to dry Mb, and a sharp increase at high water contents.

The major part of an absorption process results from the motions of individual electric dipole moment [153]. It has previously been speculated that the presence of water might enhance the mobility of various amino acid side chains along the protein, enhancing their response to the applied electric field. Water is therefore believed to increase the local flexibility of the protein molecule, and is described as a “plasticizer” by R. Pethig et al [144, 145] in their studies of lysozyme and cytochrome c. This enhanced side chain flexibility could be related to the generally increased molar absorption seen in Figure 4.6.

The dramatic increases seen above 90 wt% water is especially surprising. THz absorption studies of solvated BSA at 1.56 THz by other researchers [158] showed results consistent with ours above 95 wt% water, which are compared in Figure 4.7 with our results. The results on BSA are divided by a factor of 8 and our results at 0.5 and 0.8 THz are multiplied by a factor of 1.3 and 1.6, respectively, so that we can put all the curves in one figure. The error bars for our results are only shown on the 0.5 THz curve for a representative. The large
experimental uncertainties in the measurements for BSA could not exclude the possibility of no change in BSA molar absorptivity.

Figure 4-7. Comparison between our results on solvated Mb absorption and the published results on BSA absorption [158]. Error bars on BSA results are copied from Ref. [158] and the error bars for our results are only shown on the 0.5 THz curve for a representative. The increase in BSA molar absorptivity above 95 wt% water is consistent with our results. However, the large experimental uncertainties on the data points could not exclude the possibility of no change in BSA molar absorptivity.
While being short of reliable information about the detailed behavior of biological water in Mb-water system, we can only offer a tentative explanation for the Mb molar absorptivity increase in the dilute limit. The average separation between Mb molecules at 90 wt% concentration is about 6.5 nm. Given that the protein radius is around 1.5 nm if its shape is approximated as a sphere and the shell of biological water is usually considered to be less than 0.5 nm thick [115], this average separation corresponds to a significant amount of bulk water already between adjacent proteins. It is likely, however, that nearby proteins interact with each other, perhaps forming an instantaneous network of weak bonds that are sufficient to suppress the full polarizability of a protein molecule in the dilute limit. The 90 wt% concentration may then mark the onset of there being a finite fraction of Mb molecules free of protein-protein interactions. This would be manifested as a decrease in the solute molecule orientational relaxation time due to low viscosity [159], which facilitates the alignment of Mb molecules with incoming electromagnetic field. This would presumably enhance the interaction between Mb molecules and incident radiation field and result in an increased Mb molar absorptivity with increasing water content.

The absorption enhancement of Mb molecule may also be understood in terms of Eqn. (4.2). In Eqn. (4.2), the total dipole correlation function $C_T(t) = \langle M(t) \cdot M(0) \rangle$ is composed of three contributions, namely the permanent dipole moment of a molecule, the dipole moment on a molecule induced by other dipoles, and the interference term between permanent and induced dipoles. Upon hydration, it is unlikely that the permanent dipole moment of Mb molecule will change by a large amount. Most possibly, the perturbation of biological water will result in induced dipole moment on Mb molecules and also enhances the interferences between permanent and induced dipoles, causing stronger absorption. While this is only a plausible explanation, further studies will be required to identify the microscopic mechanisms.
Since we used seemingly arbitrary values for the biological water absorption in the third and fourth assumptions of this interacting model, one may wonder how model-dependent our results are. In the last part of this chapter, we will inspect the model-dependence of the results.

Let's take the two extreme cases as an illustration. In one extreme, the protein-bound water, except the few water molecules entrained in the protein interior (below about 4 wt%), behaves normally and absorbs THz radiation the same as bulk water. In the other extreme, the water molecules, once bound to the protein, will not interact with THz radiation at all and have no absorption. The molar absorptivity of Mb as a function of water content can be calculated accordingly using Eqn. (4.11). The results at 0.5 THz are shown in Figure 4.8. Again the molar absorptivity of dry Mb is shown as a dashed black line for comparison. Also included in Figure 4.8 are the results copied from Figure 4.6 (the middle panel), shown as solid black line. The error bar on each data point is obtained from Eqn. (4.15).

From Figure 4.8, the two trends are still clearly identifiable: there is a general enhancement of the Mb absorption relative to dry Mb in the intermediate hydration levels, and a sharp increase in the dilute limit when the water content is above 90 wt%. For the extreme when protein-bound water is taken as non-absorbing (solid blue line), the biological water absorption is generally underestimated and the absorption enhancement of Mb upon hydration is exaggerated. For the extreme when protein-bound water is taken as normal free water (solid red line), the biological water absorption is generally overestimated and the absorption enhancement of Mb upon hydration is degraded. For any models in between, the Mb molar absorptivity curves will fall between the blue line and the red line, just as what we got for the black curve using the assumptions above. This investigation shows that our results are indeed quite
model-robust. The blue and red lines actually set the upper and lower limits for the possible Mb molar absorptivity enhancement upon hydration.

Figure 4-8. Model-dependence investigation of the results. The blue line is the Mb molar absorptivity when taking all the protein-bound water as non-absorbing and the red line is the results when taking all the protein-bound water as normal free water. It can be seen that any models between these two extremes will yield results between the blue and the red lines, just as what we got for the black curve. The two trends on the Mb molar absorptivity are still clearly identifiable.
4.5. Conclusions

Absorption of far-infrared radiation in the range of 0.1 to 1.2 THz has been measured for Mb-water mixtures with water contents ranging from 3.6 to 98 wt%, using THz-TDS. A non-interacting model where absorption depends linearly on the number of Mb and water molecules does not fit the data. The expected reduction in total absorption due to biological water that strongly interacts with the protein is not observed; on the contrary, enhanced absorption is seen over a broad range of compositions and frequencies. A simple model for treating the reduced absorption of biological water and using measured values for the remaining bulk water allows the extraction of the molar absorptivity of Mb, i.e. the absorption per Mb molecule as a function of water content. This shows generally enhanced absorption plus a sharp increase above 90 wt% water. This response may be associated with enhanced flexibility of the protein for intermediate water concentrations, plus greater rotational degrees of freedom at high water contents.

These results provide quantitative data for the response of a specific protein to electromagnetic radiation in the THz range. Understanding the microscopic mechanisms for enhanced absorption may be important for accurate modeling of possible deleterious effects of electromagnetic radiation on biological materials.
CHAPTER 5. THESIS SUMMARY

Terahertz time-domain spectroscopy, since its advent in the 1980s, has been utilized successfully in the far-infrared spectroscopic measurement of various materials, including semiconductors, superconductors, polar and non-polar liquids, organic compound, gases, and biological molecules like proteins and DNA.

In this thesis, the far-infrared absorption properties of Mb-water mixtures at a series of carefully controlled hydration levels were studied. The absorption spectra of the samples were obtained at hydration levels from 1.2 to 98 wt% in the frequency range from 0.1 to 2.0 THz. It was found that Mb absorption properties are dominated by the sample water content. Each spectrum is a smooth continuous curve over the entire frequency range measured, without any identifiable sharp lines intrinsic to the Mb molecule. It was proposed that inhomogeneous broadening plus the intrinsically high spectral density of vibrational and rotational modes in the region below 2.0 THz apparently combine to obscure the lowest frequency vibrational modes expected for protein molecules of this size.

For its superiority in spectroscopic applications including wide bandwidth, high resolution, and coherent electric field measurements, there has been continuing interest in using THz-TDS as a means to detect, identify, and differentiate biological macromolecules [90, 160-162]. However, our results showed no identifiable absorption features for Mb in the THz range. It is the author's opinion that the detection of biomolecules of similar size is doubtful. Another lesson is
that one should take the experimental uncertainties in measurements as well as
data analysis very seriously since the Fabry-Perot oscillation effects often
contaminate the spectrum and the spectral density of vibrational and rotational
modes in the region below 2.0 THz are intrinsically high. In our opinion, future
hope of biological identification with THz spectroscopy may rely on
measurements of small molecules or crystal samples at cryogenic temperatures.
In fact, this has been done by several researchers on smaller molecules and
promising results were demonstrated [96, 163, 164].

The existence of biological water in proteins and the protein-water interactions
are an interesting and important research topic. Since water’s large permanent
dipole moment causes it to dominate absorption of THz radiation in most
biological systems, this technique should be especially well-suited to observing
protein-induced polarization changes in biological water. Based on the
systematic absorption measurements of Mb-water mixtures, it is demonstrated
that the Mb absorption is enhanced over a broad range of compositions and
frequencies upon hydration. This is contrary to the expected reduction in
absorption due to biological water that strongly interacts with the protein. A
simple model for treating the reduced absorption of biological water and using
measured values for the remaining bulk water allows the extraction of the molar
absorptivity of Mb (the absorption per Mb molecule) as a function of water
content. This shows generally enhanced absorption plus a sharp increase above
90 wt% water. This response may be associated with enhanced flexibility of the
protein for intermediate water concentrations, plus greater rotational degrees of
freedom at high water contents. It will be important to determine if these THz
results for Mb are seen in other proteins, and to further pursue the underlying
mechanisms for enhanced absorption.


[121] ABB BOMEM Inc. 585 Charest BLVD East Suite 300 Quebec (Quebec) G1K 9H4 Canada.

[122] *Infrared Laboratories Inc. 1808 E 17th Street, Tucson, AZ 85719.*


C. L. Brooks and M. Karplus, "Solvent Effects on Protein Motion and Protein Effects on Solvent Motion - Dynamics of the Active-Site Region of Lysozyme," *Journal of Molecular Biology*, vol. 208, pp. 159-181, 1989.


Zhang, Chenfeng was born to a rural family in Luoyang, Henan, a province in the central region of China. He spent seventeen years in his hometown until he went to Hefei, Anhui for his undergraduate study in the University of Science and Technology of China (USTC). After getting his Bachelor of Science (B.S.) degree in physics from the Department of Modern Physics in USTC, he continued his graduate study in the same department under the supervision of Prof. Xu, Kezun and obtained his Master of Science (M.S.) degree in physics three years later. He then joined in the Department of Physics in Purdue University and pursued his Doctor of Philosophy (Ph.D.) degree under the supervision of Prof. Stephen M. Durbin, doing research in terahertz spectroscopy. He obtained his PhD degree in physics in the year of 2006.