2012-07-05

Liquid Dielectric Spectroscopy and Protein Simulation

Brett Lee Mellor
Brigham Young University - Provo

BYU ScholarsArchive Citation

Mellor, Brett Lee, "Liquid Dielectric Spectroscopy and Protein Simulation" (2012). *Theses and Dissertations*. 3661.
https://scholarsarchive.byu.edu/etd/3661

Liquid Dielectric Spectroscopy and Protein Simulation

Brett L. Mellor

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Brian A. Mazzeo, Chair
David D. Busath
Neal K. Bangerter
Gregory P. Nordin
Karl F. Warnick

Department of Electrical and Computer Engineering

Brigham Young University

August 2012

ABSTRACT

Liquid Dielectric Spectroscopy and Protein Simulation

Brett L. Mellor
Department of Electrical and Computer Engineering
Doctor of Philosophy

Protein electrical properties have been studied using dielectric relaxation measurements throughout the past century. These measurements have advanced both the theory and practice of liquid dielectric spectroscopy and have contributed to understanding of protein structure and function. In this dissertation, the relationship between permittivity measurements and underlying molecular mechanisms is explored. Also presented is a method to take molecular structures from the Protein Data Bank and subsequently estimate the charge distribution and dielectric relaxation properties of the proteins in solution. This process enables screening of target compounds for analysis by dielectric spectroscopy as well as better interpretation of protein relaxation data.

For charge estimation, the shifted pK_a values for amino acid residues are calculated using Poisson-Boltzmann solutions of the protein electrostatics over varying pH conditions. The estimated internal permittivity and estimated dipole moments through shifted pK_a values are then calculated. Molecular dynamics simulations are additionally used to refine and approximate the solution-state conformation of the proteins.

These calculations and simulations are verified with laboratory experiments over a large pH and frequency range (40 Hz to 110 MHz). The measurement apparatus is improved over previous designs by controlling temperature and limiting the electrode polarization effect through electrode surface preparation and adjustment of the cell's physical dimensions. The techniques developed in this dissertation can be used to analyze a wide variety of molecular phenomena experimentally and computationally, as demonstrated through various interactions amongst avidin, biotin, biotin-labeled and unlabeled bovine serum albumin, β-lactoglobulin, and hen-lysozyme.

Keywords: dielectric spectroscopy, proteins, dipole moment, electrostatics, capacitance, permittivity, molecular dynamics, molecular interactions, protein aggregation

ACKNOWLEDGMENTS

I wish to express my gratitude for my advisor, Brian Mazzeo, not only for his invaluable direction and advice regarding my research, but for his interest in my family's well-being. David Busath has also been an integral part of this research, providing biological expertise and introducing me to molecular dynamics. Many other students and collaborators made important contributions to this dissertation.

I am immensely grateful for my wife, Whitney, who has been my biggest supporter and motivator during both my undergraduate and graduate studies, and for our son, Lincoln, who was born during this research and has been a constant source of happiness for our family.

I also acknowledge the various funding sources that made this research possible, namely, the BYU Department of Electrical and Computer Engineering, the BYU graduate fellowship award, the Lincoln Laboratory fellowship award, and the BYU Office of Research and Creative Activities.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Biomolecules, such as proteins, have remarkable electrical properties [20]. Protein electrostatic models must comprise thousands of small discrete charges to accurately model their behavior and intricate charge distributions in solution. Proteins perform a wide range of important biological functions, many of which are driven primarily by electrostatic interactions [21]. In protein-DNA binding, the driving mechanism is the attractive force between the positively charged groups on a protein and the negatively charged DNA phosphate groups. Protein electrostatics also play a major role in protein-protein and protein-ligand interactions [22, 23].

For many proteins the link between electrostatics and molecular function is less clear. For example, the reason the major protein in cow's milk exhibits one of the largest dipole moments known has eluded investigators for decades [24]. Although it is apparent there is a strong connection between electrostatics and molecular function, this connection is, in some cases, tenuous. Nonetheless, the study of biological molecules through their electrical properties has found a useful role over the years. The technique of dielectric spectroscopy has been applied with success to diverse fields such as colloid science [25], polymer science [26], and electrochemistry [27], as well as to practical applications in pharmaceutical science [28], industrial material characterization [29], and medical imaging [30].

It has become commonplace among engineers to use the term "dielectric" to refer to a material with a large dielectric constant in which no significant conduction of electrons occurs. This designation perhaps downplays the remarkable amount of information that can be learned solely from a material's dielectric constant. Indeed, knowledge of frequency-dependent dielectric constant, or permittivity, can determine the nature of the molecular mechanisms responsible for the permittivity. In liquids, this translates to size, shape, flexibility, and charge information for every solute and solvent molecule present in solution. The applications of such a technique are essentially endless, and comprise the burgeoning field of liquid dielectric spectroscopy.

This dissertation seeks to strengthen the connection between macroscopic permittivity measurements, using dielectric spectroscopy, and underlying molecular mechanisms. Proteins in aqueous solution are the principal focus, as proteins are the building blocks in many important life processes. Furthermore, electrical measurements of proteins have broader applications that reach beyond engineering and biology, into the realms of medicine and health care.

1.1 Dielectric Spectroscopy of Polar Liquids

Permittivity, ε, is a measure of a material's ability to orient its charges under an applied electric field. Permittivity is influenced by many things and is particularly sensitive to changes in temperature and frequency and thus is most accurately expressed as a function of both. Despite the term "dielectric" referring to a restricted class of materials, all materials have dielectric properties, just as all materials have mass and temperature. Most solids have permittivities between 2 and 5 that are relatively insensitive to frequency and temperature (excluding frequencies greater than 100 GHz and extreme temperatures). Because of this, two solids are likely indistinguishable by measuring their low-frequency permittivity alone.

A practical definition of permittivity is the extent to which charge polarizes within a material under the influence of an electric field. All materials exhibit electronic polarization, which is the displacement of the electron cloud relative to the nucleus. This polarization, however, can be small in comparison to the polarization that stems from the rotation of permanent dipoles. This lends itself to a common partitioning of materials into two groups: those which contain permanent dipoles, called polar materials, and those which do not, called non-polar materials. A non-polar material that is easy to contemplate is methane gas (CH_4), which contains a central carbon atom and four hydrogen atoms spaced equidistantly around the carbon in a tetrahedral configuration. Because of its symmetry, there is no imbalance of charge and it will not rotate to polarize in an electric field. As a result, the permittivity of methane gas is small ($\varepsilon < 2$ at 1 atm [31]) and its dipole moment is approximately zero. Now

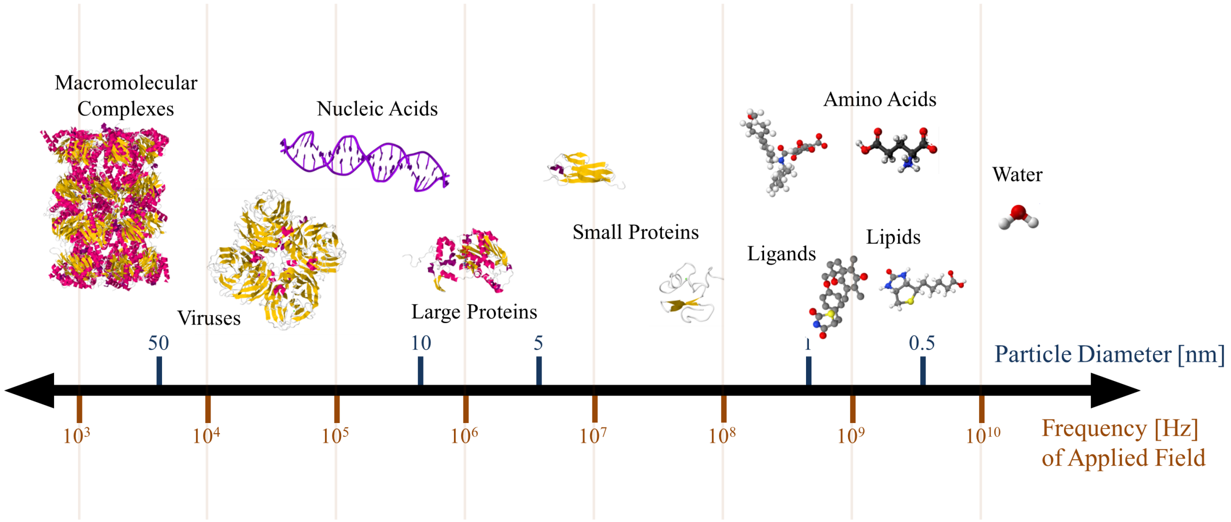


Figure 1.1: Spectrum of β-relaxations (tumbling motions constrained by hydrodynamic parameters) of molecular objects in solution that can be targeted in dielectric studies. Structures were taken from the Protein Data Bank and drawn using Jmol [2].

consider the polar molecule methanol, identical to methane with the only difference being one hydrogen atom is replaced by an OH group. Methanol has a much larger molecular dipole moment (1.68 Debye [32]) and a low-frequency permittivity over 10× larger than methane despite having just one additional atom ($\varepsilon \approx 33$ at 25 °C [33]), albeit the higher permittivity is also reflective of a higher molar density.

Some of the best materials for dielectric studies are liquids containing polar molecules. Polar molecules within solids are unable to rotate and contribute little to the material's permittivity, which is why dielectric studies of solids are focused on different molecular mechanisms [34]. The permittivities of polar liquids are of interest as they are strongly dependent on composition, temperature, frequency, and even pressure [35]. These factors make most polar liquids distinguishable from each other through their permittivity and excellent targets for dielectric studies.

1.2 Role of Dielectric Studies of Proteins

Liquid dielectric spectroscopy reveals molecular processes over a wide range of frequencies. Figure 1.1 shows various molecular objects and their approximate tumbling relaxation frequencies. At the upper end, measurements in the GHz range study the dynamics of water, lipids, and other small molecules less than 1 nm in diameter. At the lower end, which is limited by electrode polarization, proteins and larger structures such as nucleic acids can be studied in the kHz to low-MHz range. It is best to study a molecular object by measuring permittivity at frequencies at least an order of magnitude below and above its relaxation frequency. It is impossible to measure the entire frequency range with any single instrument. While measurements can range from 10^{-3} to 10^{12} Hz [36], low-frequency bridges (such as the Agilent 4294A Impedance Analyzer used in this work) cover a range from $\sim 10^1$ to $\sim 10^8$ Hz. Measurements above 10^8 Hz can be made with resonant circuits or waveguides.

Proteins are good targets for dielectric studies because their β-relaxation frequencies lie in the middle of most impedance bridge ranges. The tertiary structures of proteins typically have the property of tending to fold into a globular shape. The spherical shape decreases the spread of the relaxation frequency and decreases the required frequency range to be measured (as will be shown in Figure 2.3). Proteins are also zwitterions, being composed of both positive and negative charges resulting from titratable amino acid groups. Zwitterions with charge asymmetry have large, permanent dipole moments that give them large dielectric increments in the permittivity spectrum.

Another reason why proteins are good targets for dielectric studies is that charge and structure properties can be directly measured through permittivity, and many of these properties have important medical implications. Proteins are involved in nearly all cellular processes [37]. Disease treatment methods using protein-based therapies have been linked to coronary artery disease [38], atherosclerosis [39], and bone regeneration [40]. The functionality of the Influenza A virus is contingent on a protein (M2) binding to the cellular membrane and creating an ion channel [41]. Many neurodegenerative diseases with no available cures operate at the protein level. Both Parkinson's and Alzheimer's diseases are believed to be triggered by misfolding of synaptic proteins [42, 43]. Huntington's disease is caused by a protein mutation which is also followed by misfolding [44].

Dielectric spectroscopy of proteins has a long history, and most of the work until now has focused on measuring the protein's dipole moment [45], conformation [46], interaction with water [47], and interaction with other protein molecules [47, 48]. Broadening the applications for dielectric studies will require more sophisticated techniques and apparatus than currently are used.

1.3 Impedance-Based Biosensors

There are two major reasons why dielectric studies of proteins are performed: (1) to learn about the proteins from a theoretical standpoint, and (2) to measure biological targets in practical applications such as biosensors and in-vitro diagnostics. Biosensors are already thriving in the field of medicine. The recent, unprecedented growth of diagnostic technologies has been impressive. By the end of 2010, the estimated market value of in-vitro diagnostics was US$ 43 billion [49]. This is only expected to grow as the demand for accurate biosensors in laboratories and hospitals increases.

One important function of biosensors is pathogen detection, which is of great interest primarily for health and safety reasons. The majority of pathogen detection research is directed toward specific problems in the food industry, the field of medicine, and the Department of Defense. Over the past 20 years, biosensor technology ranked fourth behind polymerase chain reaction (PCR) [50], colony counting methods [51], and enzyme-linked immunosorbent assay (ELISA) [52] in terms of number of articles published for each pathogen detection technology [53]. Of these, biosensors are the fastest growing and are expected to soon surpass ELISA. The popularity of the traditional methods (PCR, colony counting methods, and ELISA) is explained by their sensitivity and reliability. Results from these methods, however, are usually available only after a few hours (or in the case of colony counting methods, several days [54]). Biosensor technology comes with the promise of matching the sensitivity of the traditional methods in much shorter capture times.

Biosensors must transduce a biological signal into an electrical one. The transduction method may be optical, electrochemical, thermometric, piezoelectric, magnetic, or micromechanical. Impedance is attractive for biosensors because the transduction of electrical signals is straightforward for integrated devices. It has already been used in various configurations such as monitoring bacteria or counting cells [55, 56]. Notably, impedance spectroscopy is a low-resolution technique when compared to techniques such as NMR or X-ray crystallography. However, its advantage lies in its simplicity, low cost, and speed at which experiments can be performed with the potential for highly-parallelized assays. Additionally, most biosensors require a label attached to a target, whereas impedance biosensors can be label-free [57]. This is especially advantageous for detecting binding as labeling can alter the binding characteristics of a molecule.

Due to the mathematical relationship between impedance and permittivity, it seems promising to build impedance sensors that exploit the permittivity changes of a specific reaction. As an example, consider the binding of two proteins in solution. Impedance measurements at frequencies f_1 (below both protein relaxation frequencies) and f_2 (above both protein relaxation frequencies) would be used to compute the permittivity at both frequencies, $\varepsilon(f_1)$ and $\varepsilon(f_2)$. Upon binding, the value of $\varepsilon(f_1) - \varepsilon(f_2)$ would increase or decrease as determined by the change in dipole moment and effective mass of the resulting structure. The assay could be improved by taking additional impedance measurements at surrounding frequencies to determine a confidence level that binding occurred.

1.4 Dissertation Outline

Following this introduction, the physics and theory of liquid dielectric spectroscopy will be explained in Chapter 2. This is then followed in Chapter 3 by a discussion of what is perhaps the biggest limitation in current dielectric spectroscopy measurements, electrode polarization, and how we have mitigated it. Chapter 4 describes a method for obtaining a quantitative estimate for the protein dipole moment from permittivity measurements. Chapter 5 introduces a novel technique for calculating a protein dipole moment from its three-dimensional structure. The rationale for introducing the dipole moment theory *after* the measurements is that because there are several different ways the theory can be applied, it is convenient to analyze each alongside experimental data. In Chapter 6, the role of molecular dynamics in dielectric studies is discussed. Chapter 7 then examines several molecular interaction experiments using the concepts from Chapters 2 to 6. Finally, the accomplishments of this dissertation and future research directions are given in Chapter 8.

Chapter 2

Theory of Liquid Dielectric Spectroscopy

Of the three common states of matter, our understanding of the liquid state has historically lagged behind the solid and gaseous states. Physical processes in solids and gases, in which either intermolecular forces or thermal forces generally dominate, is complicated in liquids, which exhibit a dynamic balance between the two. Thus, it is not surprising that scientists have relied heavily on experimental data to understand the physical properties of liquids, and in particular their dielectric and electronic properties.

Dielectric spectroscopy has been used for over a century to probe a broad variety of interesting and important phenomena arising from physical and chemical processes occurring within solutions [58]. New developments in electrical equipment and measurement cells over the years have greatly enhanced the ability, accuracy, and speed of researchers to measure important dielectric properties [47]. One of the more fruitful applications of dielectric studies has been protein measurements, which historically have been used to compute the electrical properties of proteins in varying solution conditions [59].

This chapter examines the underlying theory associated with liquid dielectric measurements. First, a brief history will be discussed with an emphasis on the important historical developments dealing with measurements of proteins and macromolecules. Next, the concepts of frequency-dependent permittivity and polarization will be developed. This will be important in later chapters which build off of these ideas. Last, the apparatus used to take measurements will be presented.

2.1 A Brief History

The dielectric properties of liquids have been the subject of numerous investigations over the last two centuries. Before 1900, most dielectric studies were done to measure the permittivity of water and solutions with low conductivities [60, 3]. Perhaps the first of these was done by Kohlrausch, who in 1869 used alternating current to measure the electrical properties of water without disturbing it [61]. By adding small amounts of various ions and measuring conductance, he was able to make several deductions about the mechansim for conductivity in electrolytes. The experiments done by Kohlrausch and others were important in determining the temperature dependence of the permittivity and conductivity of water, and provided the foundation for subsequent biological measurements in aqueous solution.

In the 1920s, Peter Debye posited that molecules possessed large permanent dipole moments, and made the important link between molecular dipole moment and the permittivity of solution. Much of his work is compiled in his 1929 book, *Polar Molecules* [62], for which he later received the Nobel Prize. In his honor, Debye is the unit commonly used for molecular dipole moment (1 Debye $= 3.33564 \times 10^{-30}$ C·m).

On the experimental side, the first dielectric constant measurements above 10^6 Hz were done by Jeffries Wyman at Harvard who in 1930 built resonator circuits which extended the frequency range up to 10^8 Hz [60], surpassing the performance of the well-known bridge method which had been developed by Nernst three decades earlier [63]. With this new probe, Wyman made some of the first measurements on amino acids and proteins in solution [64, 65].

In the 1930s and 1940s, John Oncley and Lars Onsager, who researched alongside Debye for several years, made improvements to Debye's original work and derived a relationship between the molecular dipole moment and the dielectric increment that is still in use today [66, 67]. Oncley additionally improved the bridge method and measured the dielectric dispersions of the proteins β-lactoglobulin, egg albumin, and carboxyhemoglobin [68, 69, 70].

Since the work of these pioneers in the field of dielectric measurements of protein solutions, various new techniques and applications have emerged. More recent advances in dielectric spectroscopy are a result of modern network and impedance analyzers that can cover extremely broad ranges of frequency, usually several orders of magnitude [47]. Advances in cell construction have led to increased stability and accuracy of measurements, reducing drastically the sample volumes in cells [71, 72].

2.2 Dielectric Theory

When a block of material is inserted between two electrodes in a parallel-plate arrangement, the capacitance between the plates increases by a factor known as the relative permittivity, denoted ε_r. The frequency and temperature dependence of ε_r constitutes the dielectric properties of that material.

The permittivity of a material is traditionally defined using the electric field, $\mathbf{E}$, expressed in V/m, and the electric displacement field, $\mathbf{D}$, expressed in C/m^2, through the equation,

$$\mathbf{D} = \varepsilon_r \varepsilon_0 \mathbf{E} \tag{2.1}$$

where ε_0 is the permittivity of free space defined as 8.854×10^{-12} farad/m. The electric field and electric displacement field are related to polarization, $\mathbf{P}$, by

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{2.2}$$

where $\mathbf{P}$ is defined as the induced dipole moment per unit volume of the material. Setting these two equations equal to each other and solving for ε_r yields

$$\varepsilon_r = \frac{|\mathbf{P}|}{\varepsilon_0 |\mathbf{E}|} + 1. \tag{2.3}$$

Thus, ε_r is a function of the ratio of how much a material polarizes in response to an electric field. For the limiting cases, $\varepsilon_r = 1$ for $|\mathbf{P}| = 0$ (vacuum), and $\varepsilon_r \approx |\mathbf{P}| / \varepsilon_0 |\mathbf{E}|$ for $|\mathbf{P}| \gg \varepsilon_0 |\mathbf{E}|$.

Polarization of a material in response to an electric field induces an opposing electric field, causing the overall field to be reduced within the material. A practical interpretation of permittivity could therefore be a material's ability to "permit" the creation of an (opposing) electric field. The net decrease in electric field inside a material is directly related to its ability to store energy. Consider a parallel-plate arrangement of two electrodes with area, A, and separation, d, connected to a battery with fixed voltage, V. The capacitance, C, of this configuration is defined both in terms of the permittivity as

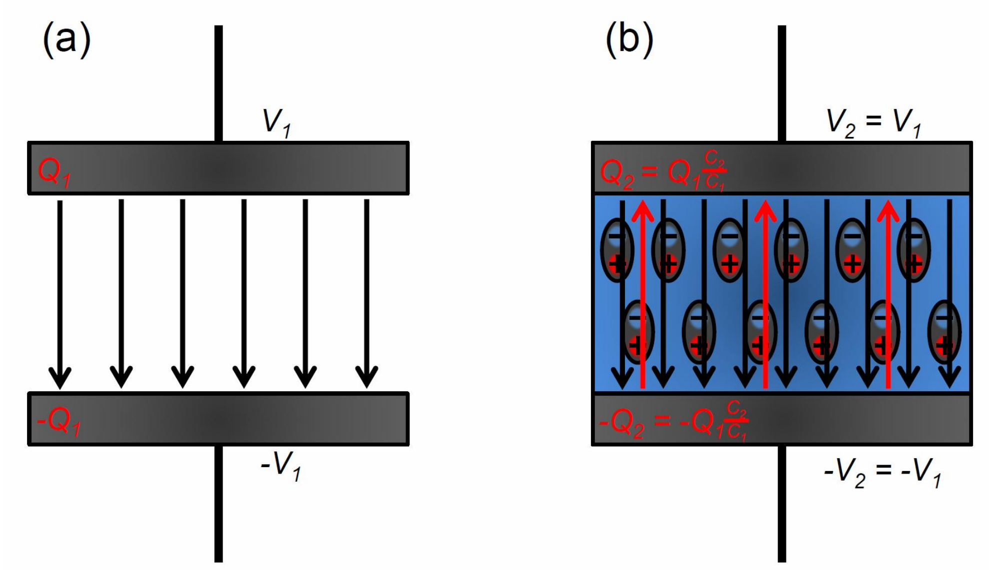


Figure 2.1: Electric field lines inside parallel-plate electrodes (a) before and (b) after a block of dielectric material is inserted between them. Polarization of the inserted material induces counter electric field lines (red arrows). Additional charge accumulates on both electrodes to maintain the net electric field constant at $2V_1/d$.

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \tag{2.4}$$

and, using the first-principles expression for capacitance, in terms of how much charge is held on each capacitor plate, Q, as

$$C = \frac{Q}{V}. \tag{2.5}$$

With V, A, d, and ε_0 all constants, the charge on each plate is proportional to the permittivity, i.e.,

$$Q \propto \varepsilon_r. \tag{2.6}$$

Because the net electric field must remain equal to V/d, when the permittivity of the material is raised and opposing electric fields are induced, additional charge must accumulate at the plates to keep the net electric field constant, as illustrated in Figure 2.1. The energy stored in the capacitor,

$$E_C = \frac{1}{2}CV^2 \tag{2.7}$$

is therefore also proportional to the permittivity, i.e.,

$$E_C \propto \varepsilon_r. \tag{2.8}$$

An important aspect of a material's dielectric properties is its ability to store electric charge, which as shown here is influenced primarily by its ability to polarize under the influence of an electric field. What follows is a discussion of the dielectric theory at the molecular level that gives a quantitative basis for permittivity and polarization.

2.2.1 The Origin of Permittivity

In Equation 2.3, permittivity is expressed as a relationship between the polarization of a material, $\mathbf{P}$, and an imposed electric field, $\mathbf{E}$. Polarization in solids, liquids, and gases come from three different molecular mechanisms: electronic polarization, $\mathbf{P}_\mathrm{e}$, atomic polarization, $\mathbf{P}_\mathrm{a}$, and orientational polarization, $\mathbf{P}_\mathrm{d}$. The total polarization is the sum of the three,

$$\mathbf{P} = \mathbf{P}_\mathrm{e} + \mathbf{P}_\mathrm{a} + \mathbf{P}_\mathrm{d}. \tag{2.9}$$

A system of two charges of $+q$ and $-q$ separated by distance d has a dipole moment of $m = qd$. Each type of polarization is a density of dipole moments, and can be expressed as a sum of dipole moments divided by a sample volume,

$$\mathbf{P}_\mathrm{e} = \frac{1}{V_s}\sum_i \mathbf{m}_{\mathrm{e},i}, \tag{2.10}$$

$$\mathbf{P}_\mathrm{a} = \frac{1}{V_s}\sum_j \mathbf{m}_{\mathrm{a},j}, \tag{2.11}$$

$$\mathbf{P}_\mathrm{d} = \frac{1}{V_s}\sum_k \mathbf{m}_{\mathrm{d},k} \tag{2.12}$$

where V_s is the sample volume. The units of polarization are then C · m / m^3, or C / m^2. Here, $\mathbf{m}_\mathrm{e}$ and $\mathbf{m}_\mathrm{a}$ are induced electronic and atomic dipole moments which by definition are always aligned with the electric field. $\mathbf{m}_\mathrm{d}$ arises from a permanent dipole moment $\boldsymbol{\mu}_\mathrm{d}$, and is the portion of $\boldsymbol{\mu}_\mathrm{d}$ that is aligned with the electric field. $\mathbf{m}_\mathrm{d}$ has magnitude

$$|\mathbf{m}_{\mathrm{d},k}| = \left|\boldsymbol{\mu}_{\mathrm{d},k}\right| \cos\theta_k \tag{2.13}$$

where θ is the angle formed by $\boldsymbol{\mu}_\mathrm{d}$ and the electric field.

Dipole moments are related to the electric field by

$$\begin{aligned} \mathbf{m}_{\mathrm{e},i} &= \alpha_{\mathrm{e},i}\mathbf{E}, && (2.14)\\ \mathbf{m}_{\mathrm{a},j} &= \alpha_{\mathrm{a},j}\mathbf{E}, && (2.15)\\ \mathbf{m}_{\mathrm{d},k} &= \alpha_{\mathrm{d},k}\mathbf{E} && (2.16) \end{aligned}$$

where α_e, α_a, and α_d are the electronic, atomic, and orientational polarizabilities respectively. Combining Equations 2.9-2.16, the total polarization can be expressed,

$$\begin{aligned} \mathbf{P} &= \frac{1}{V_s}\sum_i \mathbf{m}_{\mathrm{e},i} + \frac{1}{V_s}\sum_j \mathbf{m}_{\mathrm{a},j} + \frac{1}{V_s}\sum_k \mathbf{m}_{\mathrm{d},k}, && (2.17)\\ |\mathbf{P}| &= \frac{1}{V_s}\sum_i |\mathbf{m}_{\mathrm{e},i}| + \frac{1}{V_s}\sum_j |\mathbf{m}_{\mathrm{a},j}| + \frac{1}{V_s}\sum_k \left|\boldsymbol{\mu}_{\mathrm{d},k}\right| \cos\theta_k && (2.18)\\ &= \frac{1}{V_s}\sum_i \alpha_{\mathrm{e},i}\,|\mathbf{E}| + \frac{1}{V_s}\sum_j \alpha_{\mathrm{a},j}\,|\mathbf{E}| + \frac{1}{V_s}\sum_k \alpha_{\mathrm{d},k}\,|\mathbf{E}| && (2.19)\\ &= \frac{|\mathbf{E}|}{V_s}\left(\sum_i \alpha_{\mathrm{e},i} + \sum_j \alpha_{\mathrm{a},j} + \sum_k \alpha_{\mathrm{d},k}\right). && (2.20) \end{aligned}$$

In summary, the total polarization of a material is the sum of all dipole moments, both permanent and induced, of which arise from three different phenomena. Induced dipole moments are proportional to the electric field, and the component of permanent dipole moments aligned with the field is determined by $\cos\theta$. Equation 2.20 allows us to express $|\mathbf{P}| / |\mathbf{E}|$ as a ratio and expand the expression for ε_r, that is,

$$\frac{|\mathbf{P}|}{|\mathbf{E}|} = \frac{1}{V_s}\left(\sum_i \alpha_{\mathrm{e},i} + \sum_j \alpha_{\mathrm{a},j} + \sum_k \alpha_{\mathrm{d},k}\right) \tag{2.21}$$

and combined into Equation 2.3,

$$\varepsilon_r = \frac{1}{\varepsilon_0 V_s}\left(\sum_i \alpha_{\mathrm{e},i} + \sum_j \alpha_{\mathrm{a},j} + \sum_k \alpha_{\mathrm{d},k}\right) + 1. \tag{2.22}$$

2.2.2 Polarization

Electronic Polarization

Electronic polarization is caused by the displacement of an atom's electron cloud relative to its nucleus. An electric field exerts a force of $+\mathbf{E}Ze$ on the positively charged nucleus and a force of $-\mathbf{E}Ze$ on the negatively charged electron cloud (Z is the number of electrons). By using a simple model of a spherical electron cloud of radius, r, surrounding a nucleus of charge, Ze, it can be shown that the electronic polarizability of an atom is proportional to the radius cubed, i.e., $\alpha_{\mathrm{e}} \propto r^3$ [32]. Electronic polarization falls off in the ultraviolet frequency range ($\sim 10^{14} - 10^{15}$ Hz).

Atomic Polarization

Atomic polarization arises from changes in interionic spacings and bond angles of a molecule under an applied field. Typically this polarization falls off in the infrared frequency range ($\sim 10^{12}$ Hz).

Orientational Polarization

Orientational polarization is the polarization caused by the rotation of permanent dipoles in a material. Liquids and gases have the necessary spacing to allow for this rotation, and this is evident by the generally higher permittivities in liquids compared to solids. A molecule has a permanent electric dipole moment if it possesses spatial imbalance between positive and negative charge. Quantitatively, the dipole moment is found by integrating the charge distribution, ρ, and the position vector, $\mathbf{r}$, over all space,

$$\boldsymbol{\mu}_\mathrm{d} = \int_V \rho\left(\mathbf{r}\right)\mathbf{r}dV. \tag{2.23}$$

For biomolecules such as proteins and polypeptides, $\rho\left(\mathbf{r}\right)$ is dependent on several factors. The computation of $\boldsymbol{\mu}_\mathrm{d}$ will be the subject of Chapter 5.

It will now be shown how the orientational polarization $\mathbf{P}_\mathrm{d}$ for a group of identical dipole moments in solution is computed, following the development in Pethig [73]. For a single type of molecule, Equation 2.12 can be reduced to

$$\begin{aligned} \left|\mathbf{P}_\mathrm{d}\right| &= \frac{1}{V_s}\sum_k \left|\boldsymbol{\mu}_{\mathrm{d},k}\right| \cos\theta_k \qquad (2.24)\\ &= \frac{1}{V_s}\mu_\mathrm{d}\sum_k \cos\theta_k \qquad (2.25)\\ &= \frac{N_d}{V_s}\mu_\mathrm{d}\left\langle\cos\theta_k\right\rangle \qquad (2.26) \end{aligned}$$

where N_d is the number of dipoles $\boldsymbol{\mu}_\mathrm{d}$ in volume V_s, and $\left\langle\cos\theta_k\right\rangle$ is the ensemble average of $\cos\theta$. In reality, $\boldsymbol{\mu}_\mathrm{d}$ and $\mathbf{E}$ will never be perfectly aligned due to thermal disruptions that randomize the orientation of molecules. To find $\left\langle\cos\theta_k\right\rangle$, the energy of the dipole first needs to be expressed in terms of θ and then the Boltzmann distribution can be used.

The energy, U, of the dipole moment in an electric field can be derived by taking the integral of the force acting upon it over the rotation angle, from a reference angle, $\theta_0 = 90^\circ$, where the maximum force is exterted ($U = 0$), to an angle θ. The force exerted on a dipole moment by an electric field is a torque, Γ, given by

$$\begin{aligned} \boldsymbol{\Gamma} &= \boldsymbol{\mu}_\mathrm{d} \times \mathbf{E}, \qquad (2.27)\\ \left|\boldsymbol{\Gamma}\right| &= \mu_\mathrm{d} E \sin\theta \qquad (2.28) \end{aligned}$$

and the potential energy is then

$$
\begin{aligned}
U &= \int_{\theta_0}^{\theta} F(\theta)\, d\theta && (2.29)\\
&= \int_{\theta_0}^{\theta} \mu_{\mathrm{d}} E \sin\theta d\theta && (2.30)\\
&= \left[-\mu_{\mathrm{d}} E \cos\theta\right]_{90^\circ}^{\theta} && (2.31)\\
&= -\mu_{\mathrm{d}} E \cos\theta && (2.32)\\
&= -\boldsymbol{\mu}_{\mathrm{d}} \cdot \mathbf{E}. && (2.33)
\end{aligned}
$$

The Boltzmann distribution describes the probability that a system is in a particular energy state. The ensemble average of a function, $f(x)$, at a temperature, T, is expressed

$$
\langle f(x) \rangle = \frac{\int \mathrm{e}^{-\frac{U(x)}{k_B T}} f(x)\, dx}{\int \mathrm{e}^{-\frac{U(x)}{k_B T}} dx}. \tag{2.34}
$$

If we set $f(x) = \cos\theta$, $U(x) = -\mu_d E \cos\theta$, and set the limits of integration from $\theta = 0^\circ$, to $\theta = 180^\circ$, the probability of finding a dipole of potential energy U in a solid angle $dA = 2\pi \sin\theta d\theta$ is

$$
\begin{aligned}
\langle \cos\theta \rangle &= \frac{\int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} \cos\theta dA}{\int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} dA} && (2.35)\\
&= \frac{2\pi \int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} \cos\theta \sin\theta d\theta}{2\pi \int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} \sin\theta d\theta} && (2.36)\\
&= \frac{\int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} \cos\theta \sin\theta d\theta}{\int_{0^\circ}^{180^\circ} \mathrm{e}^{\frac{\mu_{\mathrm{d}} E \cos\theta}{k_B T}} \sin\theta d\theta}. && (2.37)
\end{aligned}
$$

This equation is simplified using the substitution $x = \mu_{\mathrm{d}} E / k_B T$ and $y = \cos\theta$,

$$
\begin{aligned}
\langle\cos\theta\rangle &= \frac{\int_{-1}^{1} \mathrm{e}^{xy} y\,dy}{\int_{-1}^{1} \mathrm{e}^{xy}\,dy} &(2.38)\\
&= \frac{d}{dx}\ln\int_{-1}^{1} \mathrm{e}^{xy}\,dy &(2.39)\\
&= \frac{d}{dx}\ln\left(\mathrm{e}^{x}-\mathrm{e}^{-x}\right) - \frac{d}{dx}\ln x &(2.40)\\
&= \coth x - \frac{1}{x} &(2.41)\\
&= L\left(x\right) &(2.42)
\end{aligned}
$$

where $L(x)$ is the Langevin function. The hyperbolic cotangent can be expressed as a series,

$$\coth x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} - \ldots, \tag{2.43}$$

from which the Langevin function can be expressed as a series,

$$L\left(x\right) = \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} - \ldots. \tag{2.44}$$

An approximation of $L(x) \approx x/3$ can be made if $x \ll 1$. This approximation holds if $\mu_{\mathrm{d}}E \ll k_B T$, that is, if the field strength-dipole interaction energy is sufficiently small with respect to the thermal energy. For a typical laboratory experiment, $T = 25$ °C and $\mu_{\mathrm{d}} = 500$ D. This corresponds to an electric field of

$$E = \frac{k_B T}{\mu_{\mathrm{d}}} = \frac{1.38\times10^{-23}\mathrm{J}\cdot\mathrm{K}^{-1}\times 298.15\,\mathrm{K}}{500\times3.336\times10^{-30}\mathrm{C}\cdot\mathrm{m}} = 2\times10^{6}\,\mathrm{V\,m^{-1}}. \tag{2.45}$$

This is several orders of magnitude larger than fields actually used in experiments, therefore, the approximation is reasonable for most protein dielectric measurements [74].

Using the Langevin approximation, the average alignment of the dipole moment with the applied field can now be expressed as

$$\langle\cos\theta\rangle = \frac{\mu_{\mathrm{d}}E}{3k_B T}. \tag{2.46}$$

The orientational polarization then becomes

$$|\mathbf{P}_{\rm d}| = \frac{N_d \mu_{\rm d}^2 E}{3V_s k_B T} \tag{2.47}$$

and a linear dependence on the electric field is obtained. An important development to note here is that polarization increases with the square of the permanent dipole moment. Also, at high temperatures the polarization goes to zero, as expected.

2.2.3 Dielectric Relaxation and Loss

The polarization expressions obtained in the previous sections assumed a DC electric field, which are the maximum values. As the frequency of the applied field is increased, some polarizations will no longer be able to attain these values. The roll-off of a polarization from its maximum value at low frequencies to a lesser value at high frequencies is called dielectric relaxation. As macromolecules are quite large, it is not surprising that orientational polarization is the slowest polarization mechanism and relaxes at the lowest frequency. Molecules can have relaxation frequencies in the GHz frequency range, where water molecules relax (see Figure 1.1). The polarization above this range essentially drops from a value of $\mathbf{P}_{\rm e}+\mathbf{P}_{\rm a}+\mathbf{P}_{\rm d}$ to $\mathbf{P}_{\rm e}+\mathbf{P}_{\rm a}$.

The common mathematical treatment to derive complex permittivity from dielectric relaxation assumes that polarization with maximum value, $\mathbf{P}_{\rm DC} = (\varepsilon_r - 1)\,\varepsilon_0\mathbf{E}$ (Equation 2.3), relaxes to zero at a rate proportional to the difference between its maximum value and its instantaneous value, i.e.,

$$\frac{d\mathbf{P}}{dt} = [\mathbf{P}_{\rm DC} - \mathbf{P}]\,\frac{1}{\tau} = [(\varepsilon_r - 1)\,\varepsilon_0\mathbf{E} - \mathbf{P}]\,\frac{1}{\tau} \tag{2.48}$$

where τ is the relaxation time (inverse of the relaxation frequency, or relaxation rate). For a field $\mathbf{E} = \mathbf{E}_0 \mathrm{e}^{j\omega t}$, the solution to Equation 2.48 is

$$\mathbf{P} = \frac{(\varepsilon_r - 1)\,\varepsilon_0\mathbf{E}}{1 + j\omega\tau}. \tag{2.49}$$

Now consider $\mathbf{P}$ as the sum of two polarizations $\mathbf{P}_1$ and $\mathbf{P}_2$, with maximum polarizations $(\varepsilon_1 - 1)\,\varepsilon_0\mathbf{E}$ and $(\varepsilon_2 - 1)\,\varepsilon_0\mathbf{E}$. If $\mathbf{P}_1$ is the combination of atomic and electronic polarizations and $\mathbf{P}_2$ is orientational polarization, then at frequencies below 10^{12} Hz

$$\begin{aligned} \mathbf{P}_1 &= (\varepsilon_1 - 1)\,\varepsilon_0\mathbf{E}, & (2.50) \\ \mathbf{P}_2 &= \frac{(\varepsilon_2 - 1)\,\varepsilon_0\mathbf{E}}{1 + j\omega\tau} & (2.51) \end{aligned}$$

and

$$\begin{aligned} \mathbf{P} &= (\varepsilon_1 - 1)\,\varepsilon_0\mathbf{E} + \frac{(\varepsilon_2 - 1)\,\varepsilon_0\mathbf{E}}{1 + j\omega\tau} & (2.52) \\ &= \left[\varepsilon_1 + \frac{\varepsilon_2 - 1}{1 + j\omega\tau} - 1\right]\varepsilon_0\mathbf{E}. & (2.53) \end{aligned}$$

Upon substituting $\varepsilon_s = \varepsilon_1 + \varepsilon_2 - 1$ and $\varepsilon_\infty = \varepsilon_1$, we can define complex permittivity as

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + j\omega\tau}. \tag{2.54}$$

As $\omega \longrightarrow 0$, $\varepsilon_r^* \longrightarrow \varepsilon_s$ where ε_s is the low-limiting permittivity. At high frequencies, as $\omega \longrightarrow \infty$, $\varepsilon_r^* \longrightarrow \varepsilon_\infty$ where ε_∞ is the permittivity at a sufficiently high frequency for orientational polarization to have fallen off. Equation 2.54 is typically expressed in terms of the change in permittivity $\Delta\varepsilon$,

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + j\omega\tau}. \tag{2.55}$$

Equation 2.55 is known as the Debye dispersion equation.

The real and imaginary components of complex permittivity are

$$\varepsilon_r^* = \varepsilon' - j\varepsilon''. \tag{2.56}$$

After some algebraic manipulation, the real component of the Debye dispersion equation is

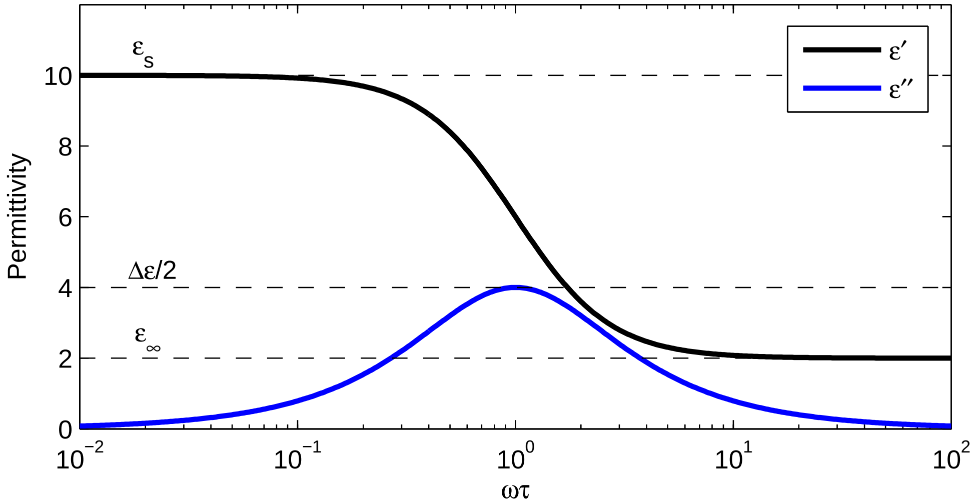


Figure 2.2: Ideal dielectric relaxation described by the Debye dispersion formulas (Equations 2.57 and 2.58). The peak height of the imaginary component is one half the difference between ε_s and ε_∞ .

$$\varepsilon^{'} = \varepsilon_\infty + \frac{\Delta\varepsilon}{1+\omega^2\tau^2} \tag{2.57}$$

and the imaginary component is

$$\varepsilon^{''} = \frac{\Delta\varepsilon\omega\tau}{1+\omega^2\tau^2}. \tag{2.58}$$

$\varepsilon^{'}$ and $\varepsilon^{''}$ are shown as functions of frequency in Figure 2.2.

2.2.4 Cole-Cole Equation

The Debye dispersion equation (Equation 2.55) was formulated for an ideal sphere, which rotates at a single relaxation time. In reality, dielectrics with a single relaxation time are rare. Proteins generally are better described by ellipsoids, which in the ideal case rotate at 2 or 3 relaxation times. In the nonideal case, ellipsoids have distributions of relaxation times. A general equation to expand Equation 2.55 to account for distributions of relaxation times is given by

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\Delta\varepsilon}{\left[1+(j\omega\tau)^{1-\alpha}\right]^\beta} \tag{2.59}$$

where exponents α and β account for symmetric and nonsymmetric distributions of relaxation times, respectively.

Equation 2.59 comprises well-known limiting cases: the Debye dispersion equation ($\alpha = 0$, $\beta = 1$ as in Equation 2.55), the Cole-Cole equation ($0 \leq \alpha \leq 1$, $\beta = 1$) [75],

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\Delta\varepsilon}{1+(j\omega\tau)^{1-\alpha}}, \tag{2.60}$$

and the Cole-Davidson equation ($\alpha = 0$, $0 \leq \beta \leq 1$),

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\Delta\varepsilon}{\left[1+j\omega\tau\right]^\beta}. \tag{2.61}$$

The variation of ideal permittivity with α and β is shown in Figure 2.3. The upper plot corresponds to the Cole-Cole equation, which retains symmetry about $\omega\tau = 1$. The Cole-Cole equation provides an adequate model for many permittivity spectra containing protein relaxation. It will be used heavily in this dissertation as a standard interpretation of the observed primary relaxation.

2.3 Measurement Apparatus

2.3.1 Two-Electrode Theory

Consider again a solution inserted between two parallel electrodes of area A and separation d. While the capacitance of the solution, C_S, is given by Equation 2.4, the conductance of the solution, G_S, is expressed

$$G_S = \frac{A\sigma}{d} \tag{2.62}$$

where σ is the electrical conductivity of the solution. The impedance of the solution, Z_S, and the admittance of the solution, Y_S ($= 1/Z_S$), are dependent on C_S and G_S and are defined

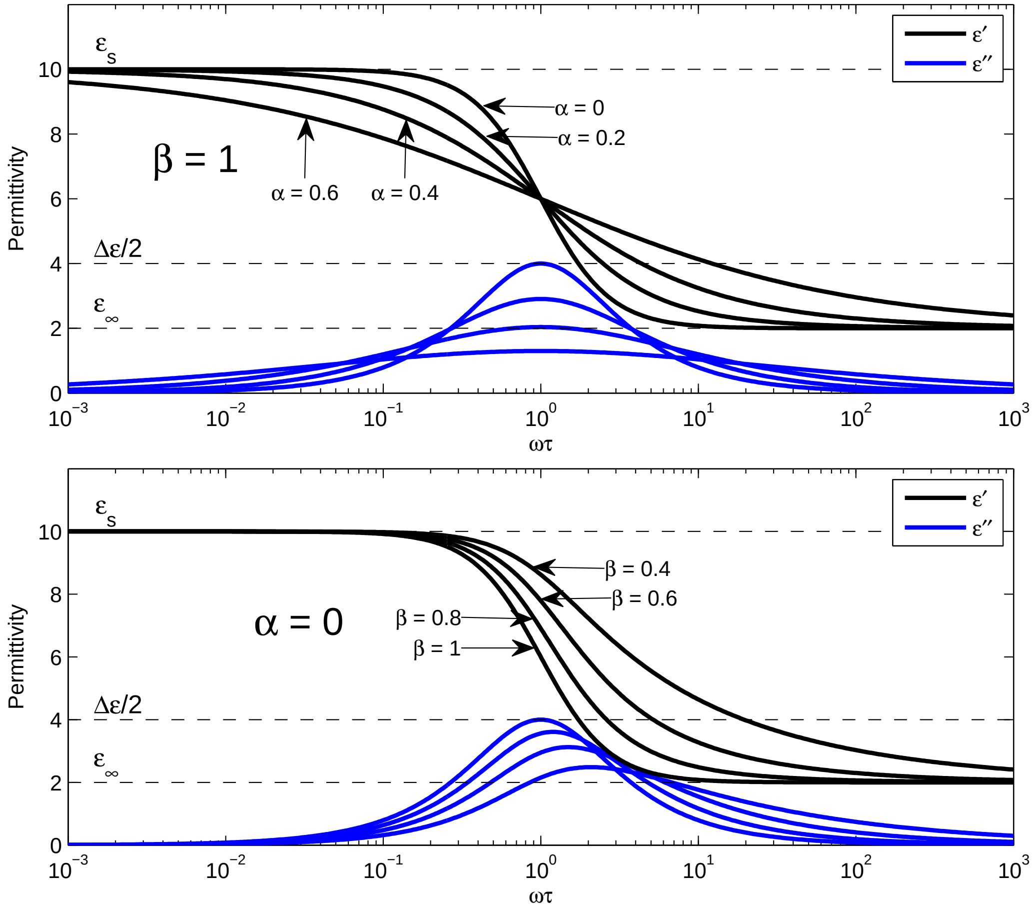


Figure 2.3: Variation of ideal dielectric relaxation with α and β as described by Equation 2.59. The upper plot is the Cole-Cole equation and the lower plot is the Cole-Davidson equation. The Debye dispersion equation is observed in both plots where $\alpha = 0$ and $\beta = 1$.

as

$$Z_S = \frac{1}{j\omega C_S + G_S} \tag{2.63}$$

and

$$Y_S = j\omega C_S + G_S. \tag{2.64}$$

Thus, the admittance must be computed first from impedance before the capacitance and conductance can be solved for.

When $\sigma = 0$, $Y_S = j\omega C_S$, and the relative permittivity remains ε_S:

$$Y_S = j\omega \frac{A\varepsilon_0\varepsilon_S}{d}. \tag{2.65}$$

When σ is non-zero, the permittivity has additional loss that is dependent on σ and ω:

$$\begin{aligned} Y_S &= G_S + j\omega C_S \quad (2.66) \\ &= \frac{A}{d}\left(\sigma + j\omega\varepsilon_0\varepsilon_S\right) \quad (2.67) \\ &= j\omega\frac{A\varepsilon_0}{d}\left(\varepsilon_S + \frac{\sigma}{j\omega\varepsilon_0}\right). \quad (2.68) \end{aligned}$$

If ε_S is the complex permittivity of a single relaxation described by the Cole-Cole equation (Equation 2.60), the measured complex relative permittivity is given by

$$\varepsilon_r^* = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma}{j\omega\varepsilon_0}. \quad (2.69)$$

Equation 2.69 is used to describe dielectric relaxation in an electrically conductive system.

Modern impedance analyzers are typically limited in their ability to resolve small phase angles between the real and imaginary parts of the impedance. Equation 2.68 demonstrates why solutions of high conductivity have very small real impedance components. To resolve solution permittivity at low frequencies requires sufficiently low conductivity. This is one reason why most dielectric spectroscopy experiments are performed in solutions with an ionic strength less than 1 mM. At high frequencies, the problem is lessened as the phase angle is larger due to the frequency-dependent impedance of the capacitance. These are some of the theoretical considerations that were considered in designing several workable cells for protein measurements.

2.3.2 Cell Calibration

When the electric field lines between two electrodes extend outside the sample solution (called fringing fields) as depicted in Figure 2.4, capacitance contributions other than C_S are present. Cell calibration is required to eliminate the parasitic capacitance, C_P, from the measurement. From the measured complex impedance, Z_M, which is received directly from the impedance analyzer, the measured capacitance, C_M, and the measured conductance, G_M, are computed through the equations

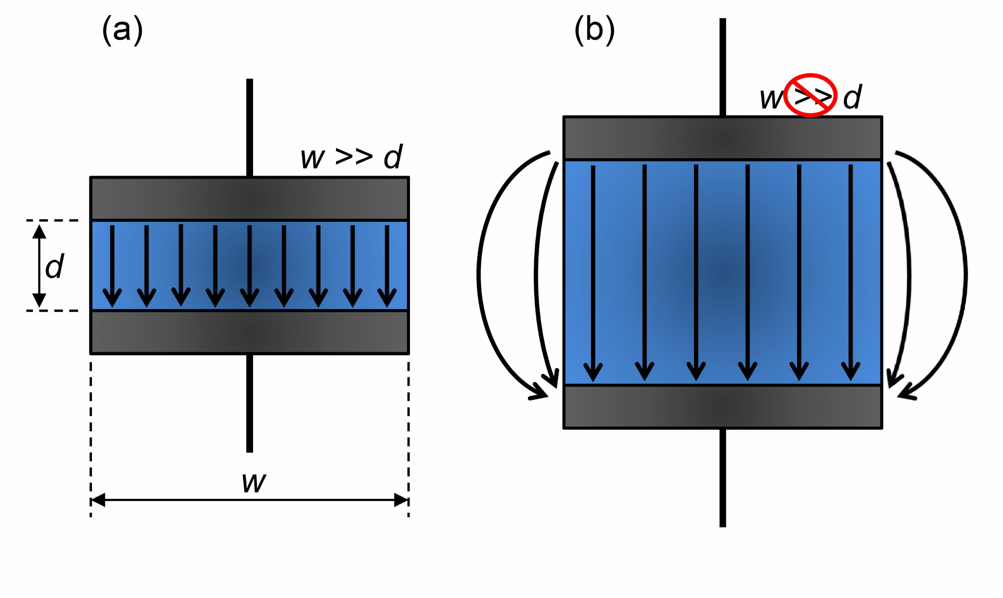


Figure 2.4: Parallel-plate electrodes (b) with and (a) without fringing fields.

$$C_M = \frac{1}{\omega}\text{imag}\left\{\frac{1}{Z_M}\right\} \tag{2.70}$$

and

$$G_M = \text{real}\left\{\frac{1}{Z_M}\right\}. \tag{2.71}$$

The measured permittivity, ε_M, is then obtained through the formula

$$C_M = k\varepsilon_M + C_P \tag{2.72}$$

where k is the cell constant. k and C_P are determined through calibration using permittivity values of 1 and 78.368 for air and de-ionized (DI) water (at 25 °C), respectively [3]. The relationship between impedance and permittivity is easily obtained through combination of Equations 2.70 and 2.72. The distinction made here between ε_M and ε_S is important. It will be shown in Chapter 3 that the two are equal if and only if $\sigma = 0$.

2.3.3 Temperature-Stable Parallel-Plate Dielectric Spectrometer

In precision protein experiments, the dielectric increment associated with proteins is an order of magnitude smaller than the background permittivity of the solution. The stability of the medium is critical so the small permittivity signal can be extracted. The permittivity of water is noticeably dependent on temperature [3]. Additionally, the relaxation characteristics of proteins are also dependent on temperature [73]. Because of this dependence, strict controls are necessary to keep the solution stable. Many commercially available cells do this by complete immersion into a bath. However, this may prevent the introduction of reaction agents into the cell without disturbing the finely tuned characteristics of the cell. Alternative methods of temperature stabilization, such as flow-through electrodes, may be employed [76].

Titration capability is important for dielectric measurements because it allows the baseline to be established for the electrode polarization and solvent permittivity. Removing and adding liquids then are perturbations that can be measured relative to the baseline. This differential method can resolve lower concentrations of protein by suppressing background parasitic contributions. Computer control is necessary to take repeated measurements and to plot time-resolved studies of protein interaction.

Dielectric cells are constructed of materials that can be autoclaved and cleaned with solvents to remove organic contaminants. Teflon (Polytetrafluoroethylene) is a popular choice for these applications due to thermal stability and chemical resistance. A cell was designed with these considerations in mind to provide a stable platform for dielectric titration of proteins. It is a significant improvement over past cells used for this purpose [48].

Cell Design

Given the above requirements, a new cell was constructed that allows the measurement of protein characteristics over the entire upper range of the impedance analyzer, up to 110 MHz. The base material for the cell was Teflon (polytetrafluoroethylene), which has low thermal conductivity 0.26 W/(K·m) [77]. The electrodes themselves were milled from 3/8" diameter stainless steel, ASTM Grade 304. A slight chamfer at the electrode edge was added to ease insertion and tight sealing against the Teflon. The surfaces were ground with 600,

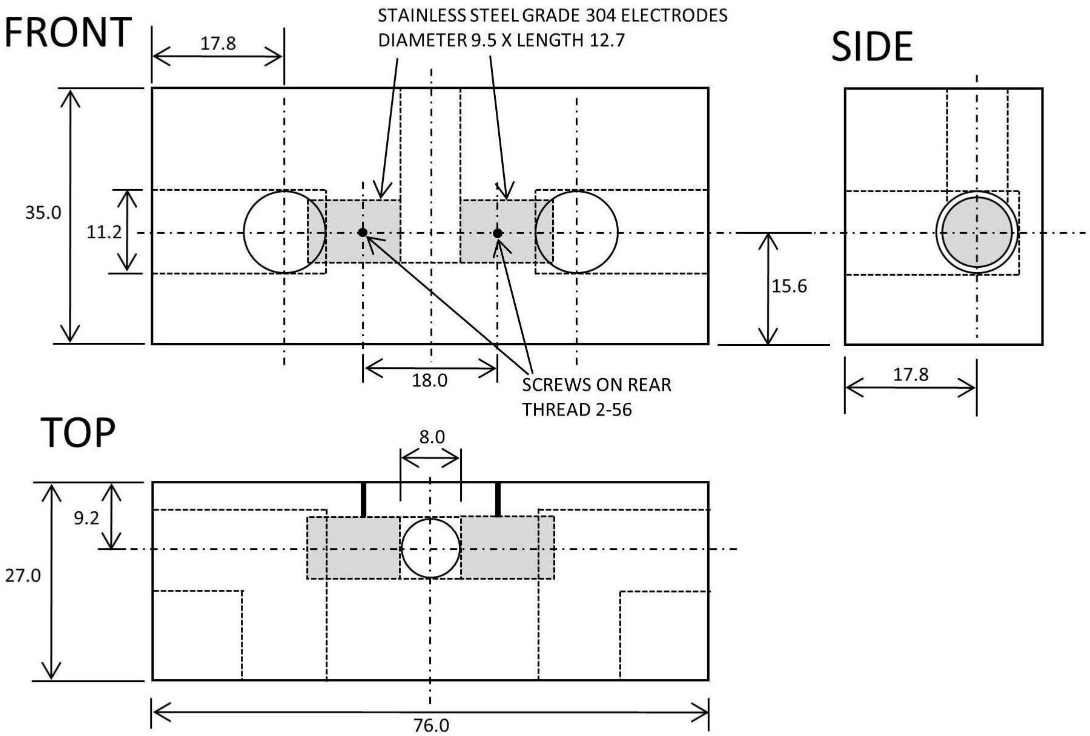


Figure 2.5: Schematic of dielectric cell designed for protein titration. Dimensions are in mm. DI water flows through ports in the front and sides. This flow by the interior electrodes maintains constant temperature in the liquid cavity.

800, and 1200 grit silicon carbide abrasive discs in a Spectrum System 2000 (LECO Corp.). DI water flowed from a NESLAB RTE-40 thermal bath to provide temperature stability. The design takes advantage of the high thermal conductivity and electrical conductivity of the stainless steel for stabilization and sensing. A rubber bung is used to seal the inner cavity. A small air gap between the liquid and the bung is sufficient to accommodate liquid expansion and contraction during thermal cycling.

Screws through tapped holes of size 2-56 were fastened through the Teflon to provide electrical connection to the inner stainless-steel electrodes. The inner, parallel-plate cavity of the dielectric cell holds 800-900 μl of fluid. The reduction of metal usage to only parallel electrode surfaces is notable in this design to reduce the parasitic capacitance of this cell. The relatively large spacing between the electrodes also reduces electrode polarization in series with the solution. The schematic for this cell is shown in Figure 2.5.

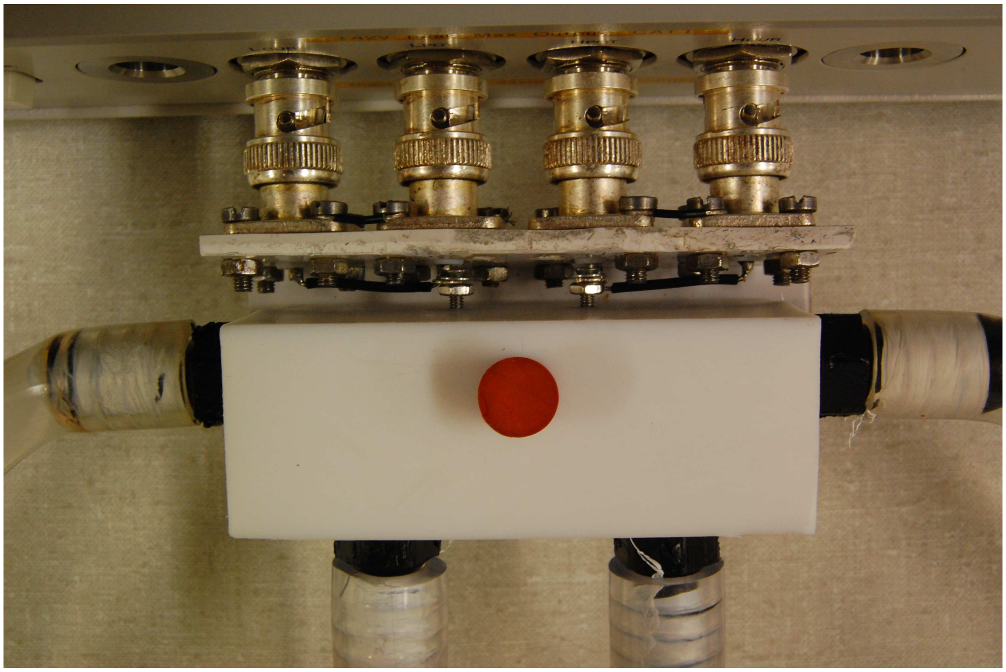


Figure 2.6: Photograph of dielectric cell connected to the 4294A Precision Impedance Analyzer. The tubes carrying DI water to and from the NESLAB RTE-40 circulating bath are visible.

The entire cell was mounted on a plastic plate connected to the four terminals of the 4294A impedance analyzer. The screws had washers and bolts that held wire that was soldered to crossbars between the low current and potential terminals and the high current and potential terminals. In this fashion, the effective cable length from the terminals of the analyzer to the cell was minimized, which is important to reduce parasitic inductance [76, 72]. Figure 2.6 shows a photograph of the completed cell mounted to the Agilent 4294A.

A LABVIEW computer interface controlled the impedance analyzer through a GPIB connection. The RTE-40 thermal bath was also controlled via the LABVIEW computer interface through a serial connection. In this fashion, the solution temperature could be accurately controlled over the entire range of the bath and impedance measurements could be taken at selected temperatures. Generally, sweeps were performed over 601 logarithmically spaced points from 40 Hz to 110 MHz with settings of a bandwidth of 5 (highest resolution) and oscillator strength of 500 mV.

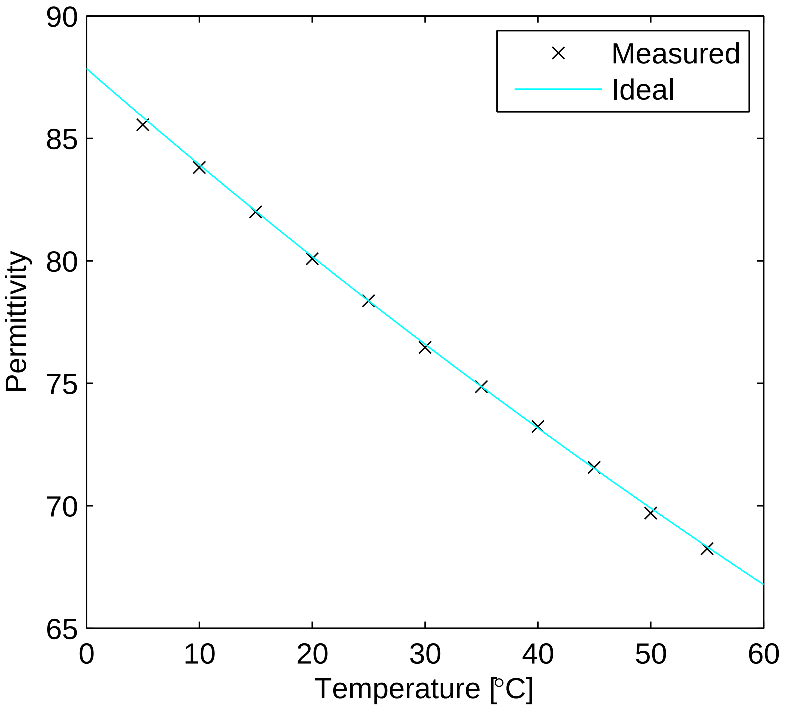


Figure 2.7: Permittivity of water versus temperature from 5 to 55 °C. The plot of the ideal temperature is taken from [3]. The dielectric cell has stable temperature readings through this range, with small departure at the temperature extremes.

Temperature Stability Test

To verify the accurate temperature control, multiple experiments were performed on water. To establish a baseline cell constant, the following approach was used. The water bath was set to 25 °C and was allowed to stabilize for 15 minutes. The capacitance was measured with the cell empty and then with 800 μl of DI water added followed by 15 minutes of stabilization time. Using these values of capacitance, the cell constant and the parasitic capacitance were determined through Equation 2.72. The measured cell constant was found to be 0.0494 pF and the parasitic capacitance was calculated to be 0.712 pF at 1 MHz.

To test the ability of the apparatus to scale through a range of temperatures, the temperature was scaled from 5 to 55 °C in 5° increments. The bath was held at the set temperature for 15 minutes before multiple frequency sweeps were performed. The permittivity of the water at 1 MHz as measured and compared to standards is shown in Figure 2.7. Indeed, the agreement is very good, particularly within the 15° to 40° range, which covers most physiological temperatures. Departure from the ideal measurements at the temperature extremes is due to the stronger thermal gradients from the interior to the exterior of the cell. At low temperatures, this is most noticeable.

Chapter 3

Electrode Polarization Effects in Dielectric Spectroscopy

Applications of dielectric spectroscopy are limited by electrode polarization, an electrochemical phenomenon that takes place at the interface between the electrode and ion-containing liquid [78]. Ions accumulate at the interface and form an electrical double layer which modifies the electrical properties of the interface [79]. Electrode polarization is a major obstacle because it is a large impedance in series with the solution impedance and thus can obscure bulk relaxation phenomena in the solution, particularly at low frequencies [35]. The impedance associated with electrode polarization, Z_{EP}, is often described phenomenologically by a constant-phase element (CPE) of the form

$$Z_{EP} = \kappa \left(j\omega\right)^{-\nu} \tag{3.1}$$

where $\kappa \geq 0$ and $0 \leq \nu \leq 1$. The phase of the CPE is described by ν, and its magnitude is described by κ. The modeling of the electrode-electrolyte interface as a CPE is attributed to the numerous measurements made over the past century indicating that the electrode polarization impedance possesses a frequency-independent proportionality between real and imaginary parts [80, 81, 82, 83].

It will be shown in Section 3.1 that as $\kappa \to 0$ and $\nu \to 1$, Z_{EP} decreases, resulting in wider bandwidth. Minimizing κ and increasing ν is thus a primary objective in dielectric spectroscopy. κ decreases at higher conductivities [84, 85, 86]. κ also decreases with larger electrode surface area [87]. ν has been related to the fractal dimension of the surface roughness [88, 89], where $\nu = 1$ is believed to correspond to a perfectly smooth surface. In some experiments, ν has also been observed to be ion concentration independent [85, 86]. Other experiments show ν to be dependent on ion type [90] and electrode material [91]. ν is not only a function of the surface characteristics of the electrode but also the chemistry of the interface.

Based on this evidence, it is not immediately clear that increasing surface roughness will reduce the electrode polarization effect; while κ is reduced, ν is also reduced, and other factors, such as increased conductance, may be more important. Other proposed techniques in the literature to reduce electrode polarization are the use of multiple electrodes [92] and blocking electrodes [93]. Electrode polarization may also be reduced through adjusting the cell's physical dimensions [94, 95]. Often, electrode polarization effects are removed analytically [96, 85].

Different groups construct parallel-plate dielectric cells for examination of liquids in specific frequency ranges. Because electrode preparation varies between different groups, analysis of surface preparation techniques will be useful to understand how to mitigate electrode polarization. In this chapter, multiple dielectric spectroscopy experiments with varying electrode properties are carried out to quantify the electrode polarization influence in the permittivity spectrum. In the process, the following three questions are examined:

1. Where does the $\omega^{-\text{const.}}$ (const. > 1) dependency of electrode polarization in the permittivity spectrum originate?
2. Can a liquid metal form an atomically smooth surface ($\nu = 1$) with a conducting liquid?
3. Does increasing surface roughness through sanding and/or depositing a conductive polymer reduce electrode polarization effects in dielectric spectroscopy measurements?

Question 1 addresses the steep electrode polarization frequency roll-off commonly observed in permittivity spectra [96, 97, 98]. It will be answered in Section 3.1 in the form of a derivation. Questions 2 and 3 will be analyzed in the context of experimental results in Sections 3.3 and 3.4, respectively.

3.1 Electrode Polarization Contribution to the Real-Valued Permittivity Spectrum

In this section, the CPE modeled electrode polarization contribution to the real-valued permittivity spectrum is derived. To understand how electrode polarization affects permittivity measurements, it is important to consider how permittivity is calculated from the measured impedance, Z_M. First, C_M is calculated from Z_M using Equation 2.70; and second, ε_M is calculated from C_M using Equation 2.72. What follows will make the important connection between Z_M and ε_S.

In the presence of electrode polarization, the measured impedance is given by

$$\begin{aligned} Z_M &= Z_{EP} + Z_S && (3.2) \\ &= \kappa \left(j\omega\right)^{-\nu} + \frac{1}{j\omega C_S + G_S}. && (3.3) \end{aligned}$$

In the absence of electrode polarization, the measured capacitance C_M is equal to the solution capacitance C_S (computed using Equation 2.70). Extraction of C_S from C_M in the presence of electrode polarization is less straightforward and is performed as follows,

$$\begin{aligned} C_M &= \frac{1}{\omega}\text{imag}\left\{\frac{1}{Z_M}\right\} && (3.4) \\ &= \frac{1}{\omega}\text{imag}\left\{\frac{1}{\kappa\left(j\omega\right)^{-\nu} + \frac{1}{j\omega C_S + G_S}}\right\} && (3.5) \\ &= \frac{1}{\omega}\text{imag}\left\{\frac{1}{\kappa j^{-\nu}\omega^{-\nu} + \frac{G_S - j\omega C_S}{\omega^2 C_S^2 + G_S^2}}\right\} && (3.6) \\ &= \frac{1}{\omega}\text{imag}\left\{\frac{1}{\kappa e^{-j\nu\frac{\pi}{2}}\omega^{-\nu} + \frac{G_S}{\omega^2 C_S^2 + G_S^2} - j\frac{\omega C_S}{\omega^2 C_S^2 + G_S^2}}\right\} && (3.7) \\ &= \frac{1}{\omega}\text{imag}\left\{\frac{1}{\kappa\left[\cos\left(-\nu\frac{\pi}{2}\right) - j\sin\left(\nu\frac{\pi}{2}\right)\right]\omega^{-\nu} + \frac{G_S}{\omega^2 C_S^2 + G_S^2} - j\frac{\omega C_S}{\omega^2 C_S^2 + G_s^2}}\right\} && (3.8) \end{aligned}$$

$$= \frac{1}{\omega}\text{imag}\left\{\frac{1}{\left[\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{G_S}{\omega^2C_S^2+G_S^2}\right]-j\left[\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{\omega C_S}{\omega^2C_S^2+G_S^2}\right]}\right\} \tag{3.9}$$

$$= \frac{1}{\omega}\text{imag}\left\{\frac{\left[\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{G_S}{\omega^2C_S^2+G_S^2}\right]+j\left[\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{\omega C_S}{\omega^2C_S^2+G_S^2}\right]}{\left[\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{G_S}{\omega^2C_S^2+G_S^2}\right]^2+\left[\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{\omega C_S}{\omega^2C_S^2+G_S^2}\right]^2}\right\} \tag{3.10}$$

$$= \left(\frac{1}{\omega}\right)\frac{\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{\omega C_S}{\omega^2C_S^2+G_S^2}}{\left[\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{G_S}{\omega^2C_S^2+G_S^2}\right]^2+\left[\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}+\frac{\omega C_S}{\omega^2C_S^2+G_S^2}\right]^2} \tag{3.11}$$

$$= \tag{3.12}$$

$$= \frac{\frac{C_S}{\omega^2C_S^2+G_S^2}+\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu-1}}{\kappa^2\omega^{-2\nu}+\frac{2\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}G_S}{\omega^2C_S^2+G_S^2}+\frac{G_S^2}{\left(\omega^2C_S^2+G_S^2\right)^2}+\frac{2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu+1}C_S}{\omega^2C_S^2+G_S^2}+\frac{\omega^2C_S^2}{\left(\omega^2C_S^2+G_S^2\right)^2}} \tag{3.13}$$

$$= \tag{3.14}$$

$$= \frac{C_S+\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu-1}\left(\omega^2C_S^2+G_S^2\right)}{\kappa^2\omega^{-2\nu}\left(\omega^2C_S^2+G_S^2\right)+2\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}G_S+2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu+1}C_S+1} \tag{3.15}$$

$$= \frac{C_S+\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu+1}C_S^2+\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu-1}G_S^2}{\kappa^2\omega^{-2\nu+2}C_S^2+\kappa^2\omega^{-2\nu}G_S^2+2\kappa\cos\left(-\nu\frac{\pi}{2}\right)\omega^{-\nu}G_S+2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-\nu+1}C_S+1} \tag{3.16}$$

$$= \frac{C_S+A_1+A_2}{A_3+A_4+A_5+A_6+1} \tag{3.17}$$

where

$$\begin{aligned}
A_1 &= \kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu-1)}C_S^2, \qquad (3.18)\\
A_2 &= \kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu+1)}G_S^2,\\
A_3 &= \kappa^2\omega^{-2\nu+2}C_S^2,\\
A_4 &= \kappa^2\omega^{-2\nu}G_S^2,\\
A_5 &= 2\kappa\cos\left(\nu\frac{\pi}{2}\right)\omega^{-\nu}G_S,\\
A_6 &= 2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu-1)}C_S.
\end{aligned}$$

For experimental conditions ($\kappa \approx 10^5$, $\nu \approx 0.8$, $C_S \approx 10^{-12}$, $G_S \approx 10^{-4}$, $2\pi \cdot 10^4 < \omega < 2\pi \cdot 10^8$), A_1 through A_6 become

$$\begin{aligned}
5\cdot 10^{-19} \lesssim & \quad A_1 & \lesssim 2\cdot 10^{-18}, \quad (3.19)\\
10^{-19} \lesssim & \quad A_2 & \lesssim 2\cdot 10^{-12},\\
8\cdot 10^{-13} \lesssim & \quad A_3 & \lesssim 3\cdot 10^{-11},\\
8\cdot 10^{-13} \lesssim & \quad A_4 & \lesssim 2\cdot 10^{-6},\\
5\cdot 10^{-7} \lesssim & \quad A_5 & \lesssim 9\cdot 10^{-4},\\
2\cdot 10^{-6} \lesssim & \quad A_6 & \lesssim 10^{-5}.
\end{aligned}$$

Simplifying due to $A_3 + A_4 + A_5 + A_6 \ll 1$ and $A_1 \ll C_S$ yields

$$\begin{aligned}
C_M &\approx C_S + A_2 && (3.20)\\
&= C_S + G_S^2 \kappa \sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu+1)}. && (3.21)
\end{aligned}$$

When calibration is performed using Equation 2.72, the measured permittivity using the approximation in Equation 3.20 is given by

$$\begin{aligned}
\varepsilon_M &= \frac{\left(C_S + G_S^2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu+1)}\right) - C_p}{k} && (3.22)\\
&= \frac{C_S - C_p}{k} + \frac{G_S^2\kappa\sin\left(\nu\frac{\pi}{2}\right)\omega^{-(\nu+1)}}{k} && (3.23)\\
&= \varepsilon_S + \frac{G_S^2\kappa\sin\left(\nu\frac{\pi}{2}\right)}{k}\omega^{-(\nu+1)} && (3.24)\\
&= \varepsilon_S + \overline{\kappa}\omega^{-\overline{\nu}} && (3.25)
\end{aligned}$$

where

$$\begin{aligned}
\overline{\kappa} &= \frac{G_S^2\kappa\sin\left(\nu\frac{\pi}{2}\right)}{k} && (3.26)\\
\overline{\nu} &= \nu + 1. && (3.27)
\end{aligned}$$

We thus see that the electrode polarization contribution to the real-valued permittivity spectrum is

$$\varepsilon_{EP} = \frac{G_S^2 \kappa \sin\left(\nu \frac{\pi}{2}\right)}{k} \omega^{-(\nu+1)}. \tag{3.28}$$

For a parallel-plate arrangement, $k = A/d$ and $G_S = \sigma A/d$. Equation 3.28 then becomes

$$\varepsilon_{EP} = \sigma^2 \left(\frac{A}{d}\right) \kappa \sin\left(\nu \frac{\pi}{2}\right) \omega^{-(\nu+1)}. \tag{3.29}$$

Equations 3.28 and 3.29 clearly explain the origin of the $\omega^{-\text{const.}}$ (const. > 1) dependency that is observed in real-valued permittivity measurements, as $\overline{\nu}$ would be expected to lie between 1 and 2. Note that ε_{EP} decreases faster with increasing frequency than Z_{EP} ($\nu < \overline{\nu}$) and its magnitude increases with the square of the solution conductivity and thus becomes significant in solutions of high ionic strength. We can use Equation 3.29 to understand the basic mechanisms through which electrode polarization influences permittivity measurements. The electrode polarization effect is essentially comprised of three parts: first, the contribution from the solution conductivity (σ^2); second, the contribution from the physical dimensions of the measurement cell (A/d) (consistent with previous studies showing that electrode polarization effects decrease with increased electrode spacing [94]); and third, the contribution from the electrode-electrolyte interface $\left(\kappa \sin\left(\nu \frac{\pi}{2}\right) \omega^{-(\nu+1)}\right)$. Equations 3.28 and 3.29 also permit a quantifiable lowest frequency where electrode polarization effects appear, f_{on} ("onset of EP"), which is defined in this dissertation as the frequency such that $\varepsilon_{EP} = 1$ (constrast with [78]). This is expressed as

$$f_{on} = \frac{1}{2\pi} \left[\frac{G_S^2 \kappa \sin\left(\nu \frac{\pi}{2}\right)}{k}\right]^{\left(\frac{1}{\nu+1}\right)}. \tag{3.30}$$

Equation 3.30 is a single frequency lower bound from which electrode polarization effects in dielectric spectroscopy can be evaluated as it combines information from both κ and ν.

3.2 Model Equation for Salt Solution with Electrode Polarization

The solution impedance, Z_S, for a salt solution is well characterized by a constant capacitance, C, (constant permittivity) and conductance, G, (constant conductivity) at all

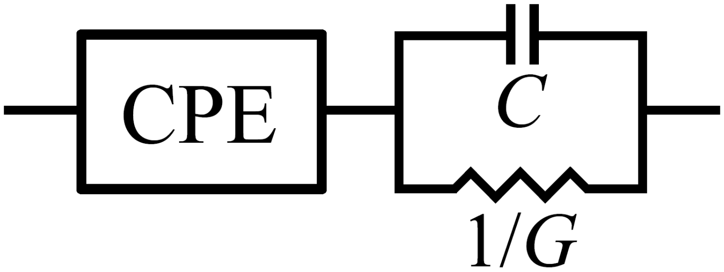


Figure 3.1: Model circuit for CPE in series with salt solution (Equation 3.3 with C_S and G_S constant over frequency).

frequencies ($C_S = C$ and $G_S = G$ in Equation 3.3). The model equation for a salt solution in series with a CPE is thus represented by the circuit in Figure 3.1. In the next two sections, all least-squares fittings of Z_M to this equation are performed using a complex fit routine in MATLAB.

3.3 Liquid Metal Electrode Experiments

An atomically smooth surface formed by a liquid metal could yield a reproducible and ideal electrode surface from both theoretical and practical perspectives [79, 99, 83]. If a liquid electrode could yield a ν close to 1, it would have huge implications for electrode polarization mitigation. Thus, the use of a non-toxic liquid metal, Galinstan, motivated a series of experiments to determine its viability as an electrode surface.

Galinstan was developed by Geratherm to replace mercury in thermometers and is a eutectic mixture of gallium, indium, and tin [100]. Galinstan is similar to eutectic Gallium-Indium (EGaIn), which has been the subject of many recent studies [101, 102, 103], and exhibits properties of a non-Newtonian liquid due to formation of a thin oxide layer [101]. It is liquid above -19 °C and has an electrical resistivity of 0.435 Ω-μm [100]. Previous applications have included antennas [103], mirrors [104], and microfluidic structures [102, 101]. Here, Galinstan is used as an electrode for liquid impedance spectroscopy [105]. Polished ASTM grade 304 stainless steel is used comparatively [91].

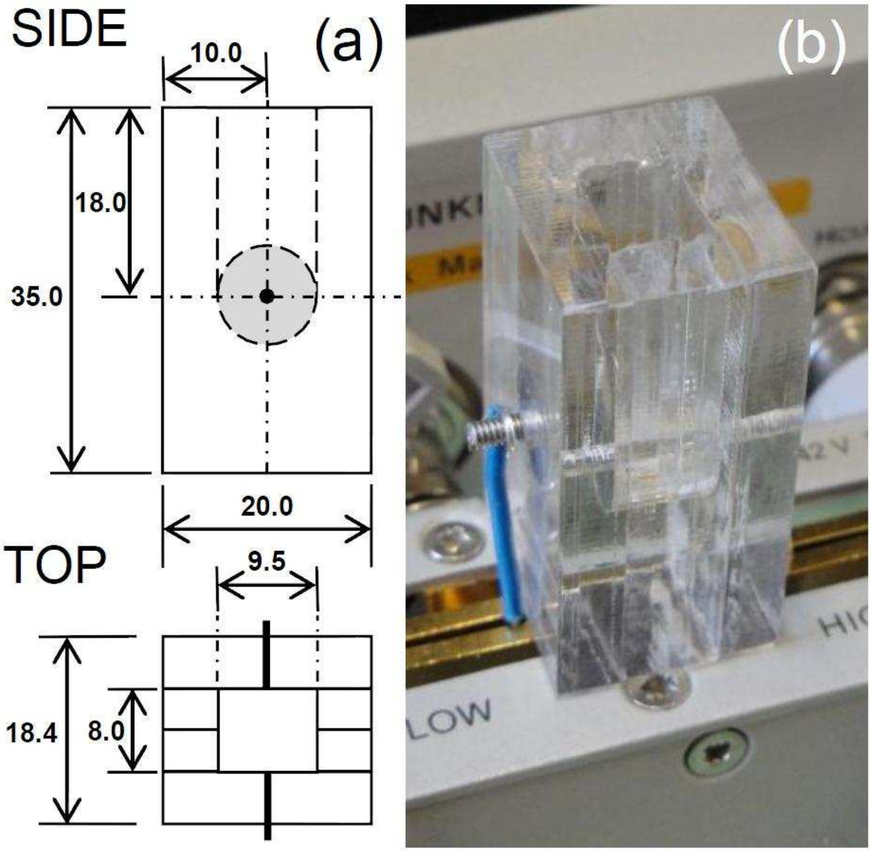


Figure 3.2: Parallel-plate dielectric cell with Galinstan electrodes (a) schematic and (b) picture. Cell is mounted on the Agilent 16047E test fixture. Dimensions are in mm.

3.3.1 Cell Design

Two 800 μl volume dielectric cells, one with stainless steel electrodes and one with Galinstan electrodes, were constructed to measure the impedance of various liquids. The construction of the stainless steel cell was described in Section 2.3.3. The Galinstan cell was built with a similar geometry to the stainless steel cell.

The Galinstan cell schematic and picture are shown in Figure 3.2. Four acrylic plates, measuring 20 mm by 35 mm, were cut out of 5.2 mm thick acrylic stock using a Rayjet Laser Engraver (Trotec). Holes for screws were cut into the outside plates. 9.5 mm wide channels and 9.5 mm diameter circles were cut out of the two middle plates. An electrode spacing of 8 mm was attained by thinning the middle plates through a process of laser engraving and sanding. Steel zinc-plated 2-56 pan-slotted-head 1/2” screws were ground to a flat tip, then screwed into the end plates and left slightly recessed. Galinstan, which adheres to acrylic

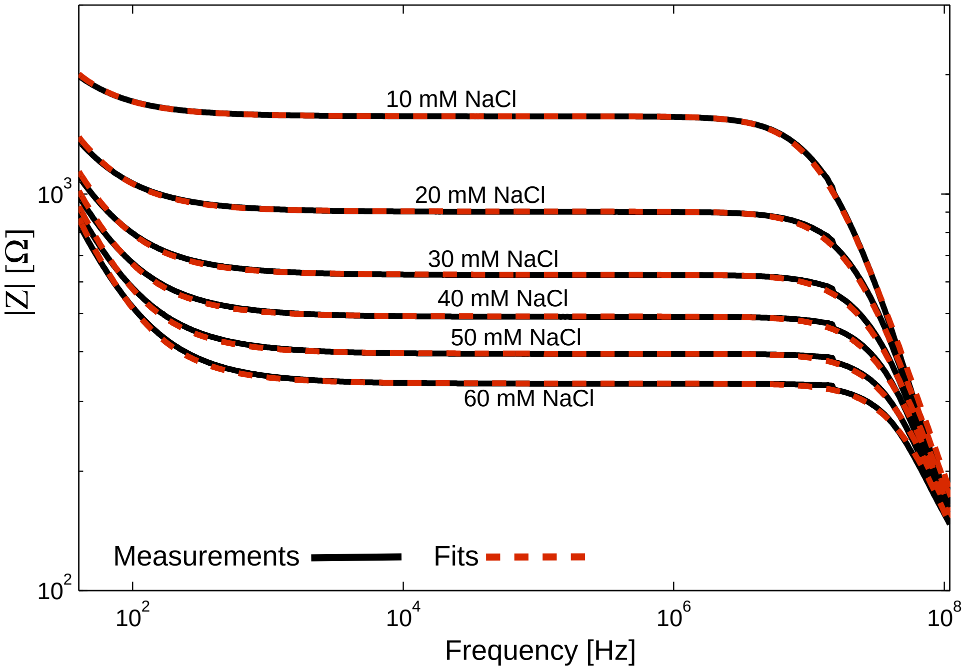


Figure 3.3: Measurements of impedance magnitude using the Galinstan cell (solid) and least-squares fits (dashed) of different salt solutions. The effect of electrode polarization is visible at low frequencies, and is more pronounced for higher NaCl concentrations. The small deviation at high frequencies is due to the parasitic inductance of the cell.

surfaces, was pipetted onto the outer plates in 300 μl droplets through a auxiliary plate with a 9.5 mm diameter circular hole. Excess Galinstan was then drawn off, leaving a thin layer of $\sim$ 50 μl forming the circular electrodes. The acrylic plates were then thermally bonded together following Kelly and Woolley's technique [106].

The stainless steel cell was connected directly to the measurement ports of the 4294A Impedance Analyzer [98]. The Galinstan cell was connected through wires from the cell screws to the test fixture (Agilent 16047E). Complex impedance was measured with 601 logarithmically-spaced points from 40 Hz to 110 MHz with a bandwidth setting of 5 and an oscillator level of 500 mV. The apparatus and liquids were kept at room temperature. Capacitance measurements at 20 MHz in the Galinstan cell indicated that the temperature of the liquids was relatively stable.

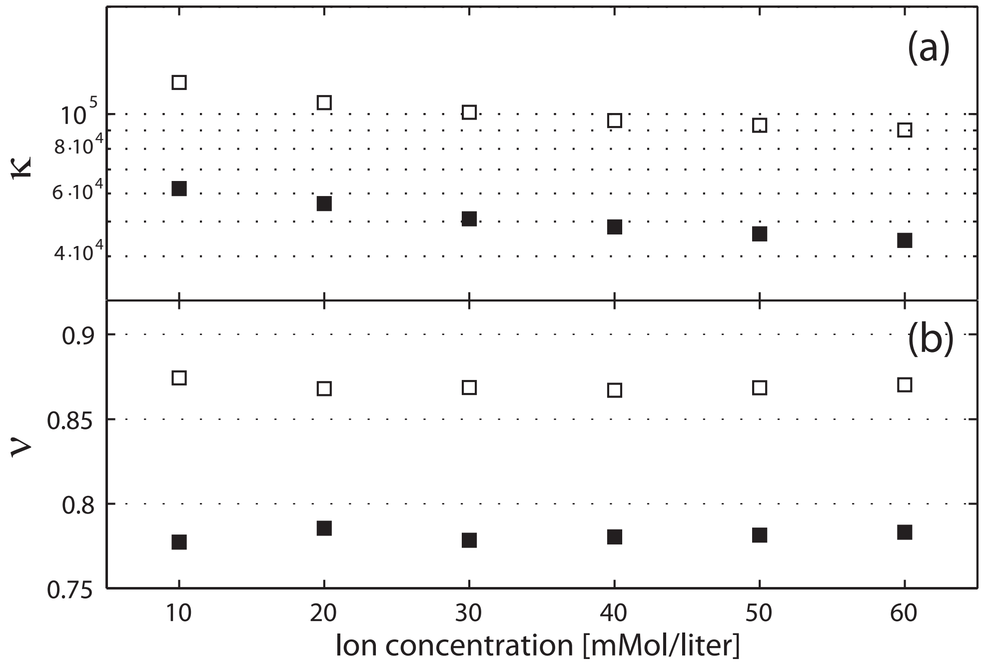


Figure 3.4: CPE parameters obtained by least-squares fitting for (a) κ and (b) ν from measured complex impedance of 10 mM to 60 mM NaCl solutions using stainless steel (□) and Galinstan (■) electrodes.

Salt solutions of 10 mM to 60 mM NaCl in 10 mM increments were prepared from 5 M NaCl stock from Sigma (S6546) and DI water. 800 μl volumes of each concentration were pipetted into both cells.

3.3.2 Results and Discussion

Fits were quite good at all frequencies using this model and are plotted in Figure 3.3. Fits for κ and ν in Figure 3.4 show that the CPE magnitude and phase are smaller for the Galinstan cell at all salt concentrations. Notably, ν is not close to 1 for Galinstan as predicted. The average CPE phase for the polished stainless steel electrode was $\nu = 0.87$, much higher than the Galinstan electrode at $\nu = 0.78$, indicating that although κ is smaller for Galinstan, the influence of electrode polarization will be much larger at high frequencies in that cell. This is confirmed by solving for f_{on} in both setups. Therefore, we expect the Galinstan cell to have a smaller frequency bandwidth for permittivity measurements of proteins.

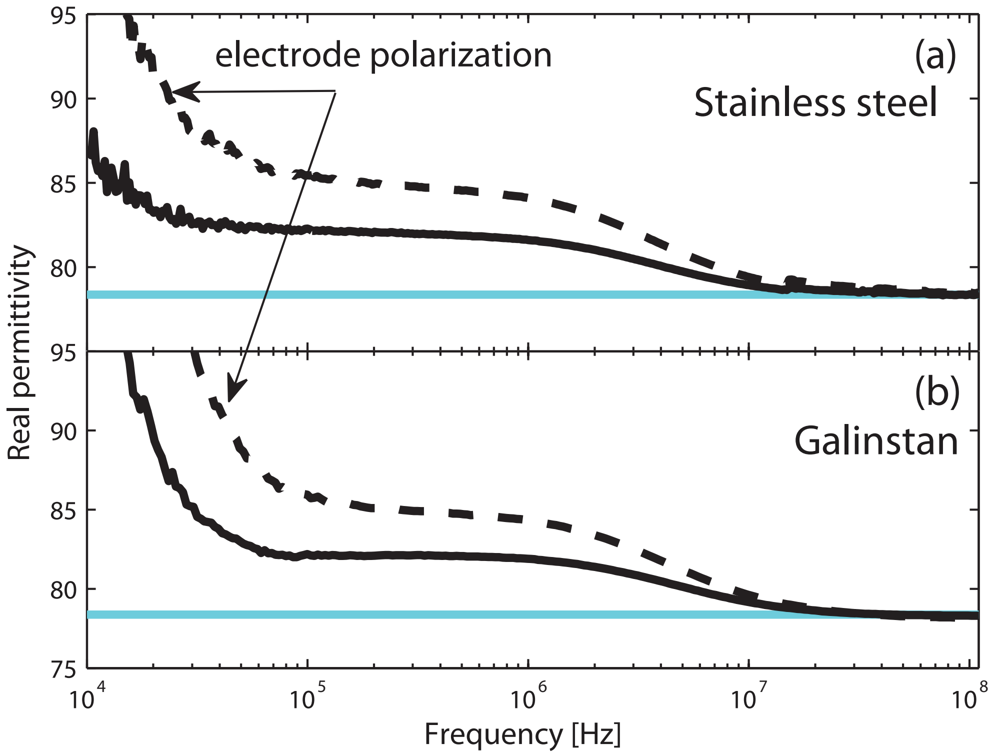


Figure 3.5: Measured permittivity of β-Lg using (a) stainless steel and (b) Galinstan cells. Solid horizontal line is the permittivity of water at 25 °C. Protein concentrations were 2 mg/ml (solid black line) and 3.8 mg/ml (dashed line). Electrode polarization is more significant in the measurements using Galinstan electrodes.

Next, to demonstrate increased electrode polarization for the Galinstan cell, measurements were made of a protein relaxation. Solutions of 2 mg/ml and 3.8 mg/ml β-lactoglobulin (β-Lg) in DI water were prepared and pipetted into both cells. Permittivity measurements for β-Lg using both cells are shown in Figure 3.5a and 3.5b. Both plots show typical protein relaxations and are nearly identical at high frequencies where electrode polarization is negligible. Electrode polarization is clearly visible at low frequencies and more significantly influences measurements of protein relaxation in the Galinstan cell. This can be confirmed by examining the frequency visually at which the permittivity curves cross the $\varepsilon = 95$ line. Although the β-Lg relaxation in this case is still accessible, for other biomolecules with relaxations occurring at larger time scales, the dielectric spectra would be partially obscured and more difficult to extract. It is interesting to note that at high frequencies the spectrum measured by the Galinstan cell appears to be smoother than the stainless steel. This is attributed to the automatic switching of the impedance bridge and is not a function of the electrode materials themselves.

Post-experiment optical microscopy of the electrodes (with solution removed) revealed Galinstan's surface to be uncommonly rough. It is likely the roughness is due to the oxide layer which Galinstan forms spontaneously at its surfaces [104, 101]. The electrochemical nature of the metal-oxide system of a similar liquid metal was reported by previous investigators to have a strong dependence on the applied voltage and solution conductivity [103]. Particularly, the mechanical stability of the oxide layer is reduced with strong voltages, high conductivities, or low pH, under which conditions the liquid metal could possibly form a smoother interface. However, in the conditions of this experiment (low salt, low voltage, and physiological pH), the oxide layer persists in the presence of the electrolyte medium. Future experiments will address this.

While some liquid metals may be able to form atomically smooth surfaces, this experiment did not find it true for Galinstan electrodes. Other electrode materials or additional surface treatment may produce better surface conditions to reduce electrode polarization in liquid impedance spectroscopy experiments.

3.4 Electrode Surface Preparation Experiments

The next set of experiments were aimed to limit electrode polarization through different surface preparations of the stainless steel electrodes. The most common technique used in the literature is polishing with microgrit and macrogrit sandpapers. Electrodeposition of a conductive polymer onto the stainless steel electrodes was also investigated.

3.4.1 Materials and Methods

Cell Design

Dielectric cells were constructed out of acrylic as shown in Figure 3.6. Electrodes were cut from ASTM grade 304 stainless steel 1/4" diameter rod, then press-fit into opposite sides of the acrylic. The cells were connected to the impedance analyzer through the Agilent 16047E test fixture.

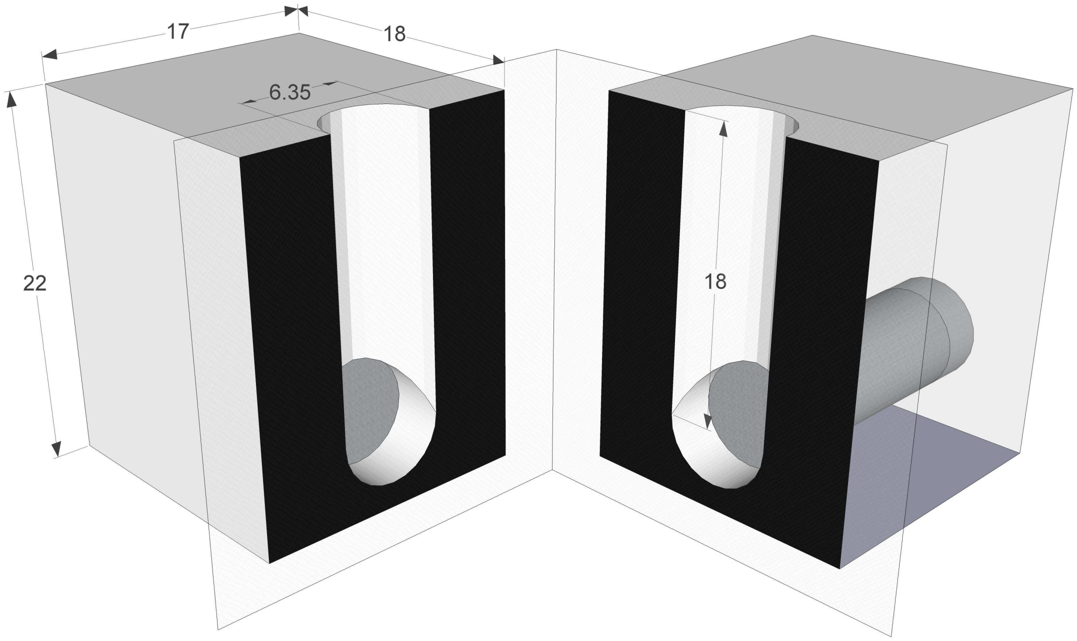


Figure 3.6: Cross-sectional view of measurement cell (dimensions in mm). Liquid is pipetted into central cavity through the top port. The actual cells are constructed from a single block of acrylic [4].

Polishing

Five different levels of polishing were applied to the stainless steel electrodes. One electrode set was sanded using 60 grit silicon carbide (SiC) sandpaper and the other four were sanded with microgrit SiC sandpapers. The four electrode sets that received microgrit sanding were first flattened on both ends with a lathe and then ground with a 600 grit SiC abrasive disc. Additional stages of polishing using 800 grit SiC paper, 1200 grit SiC paper, and diamond paste were then applied sequentially to the electrodes, resulting in four levels of microgrit polishing. The diamond paste was applied in 9, 3, and 1 μm sizes. After the polishing process, the electrodes were rinsed with DI water and inspected with a scanning electron microscope (SEM) and a profilometer (Tencor Instruments) before insertion into the acrylic cells. Several SEM images are shown in Figure 3.7. It is clear that microgrit sandpapers remove some of the irregularities in the surface structure.

The roughness of the microgrit sanded electrodes were quantified with profilometer measurements in 0.04 μm steps over an 80 μm sample length (Figure 3.8). The vertical path

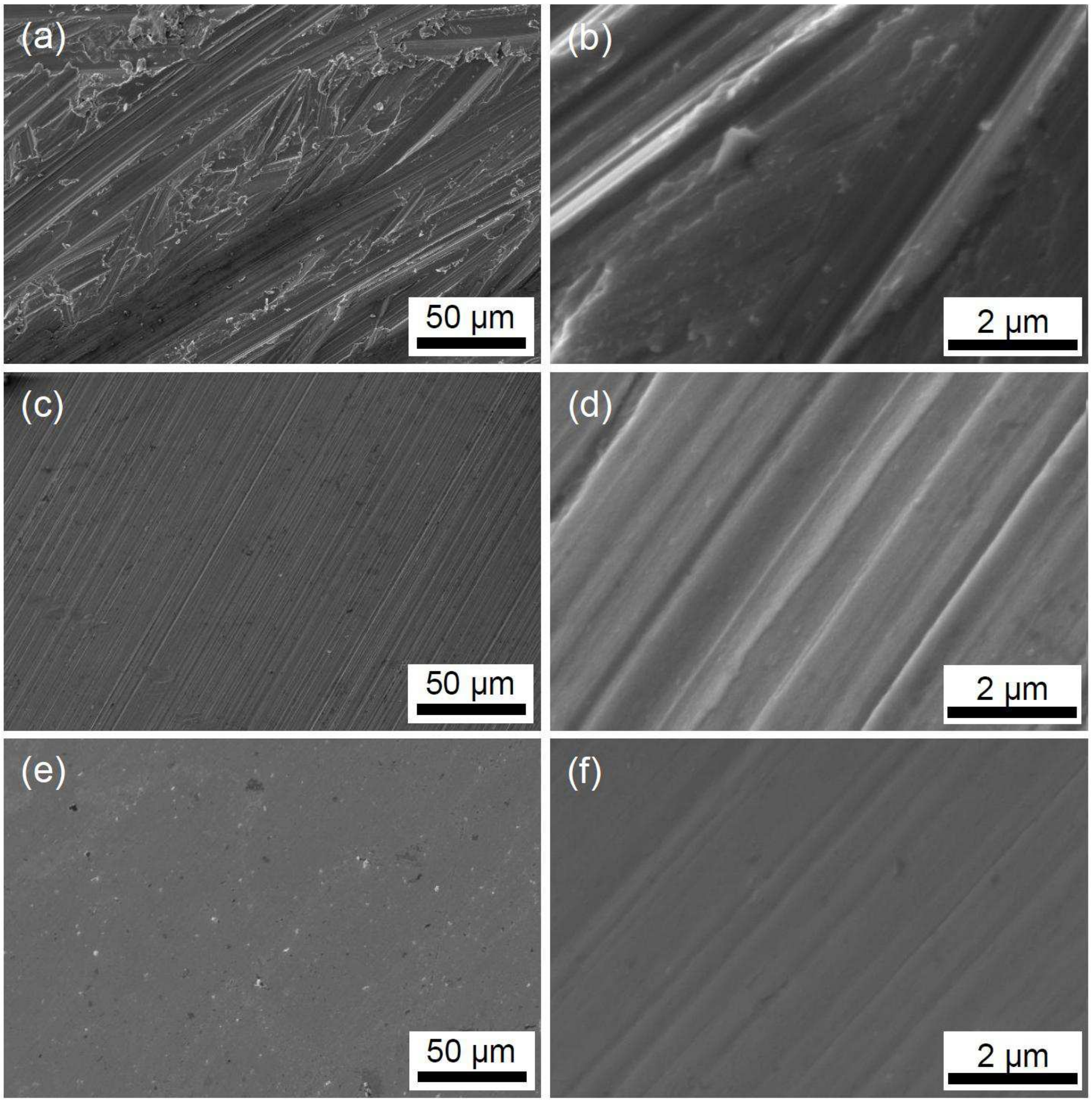


Figure 3.7: SEM images of electrode surfaces with (a,b) 60 grit, (c,d) 600 grit, and (e,f) 1200 grit sanding.

length, L, and the standard deviation of the surface heights, σ, can be indicative of roughness for certain sample lengths. For 80 μm sample lengths, the higher levels of polishing clearly demonstrate increased smoothness from 800 grit to 1200 grit and 1200 grit to diamond paste. The difference between 600 grit and 800 grit cannot easily be discerned due to their large particle diameters (25.8 and 21.8 μm respectively). Profilometry at coarser roughness levels was not instructive as grooves were on the order of the scribe movement.

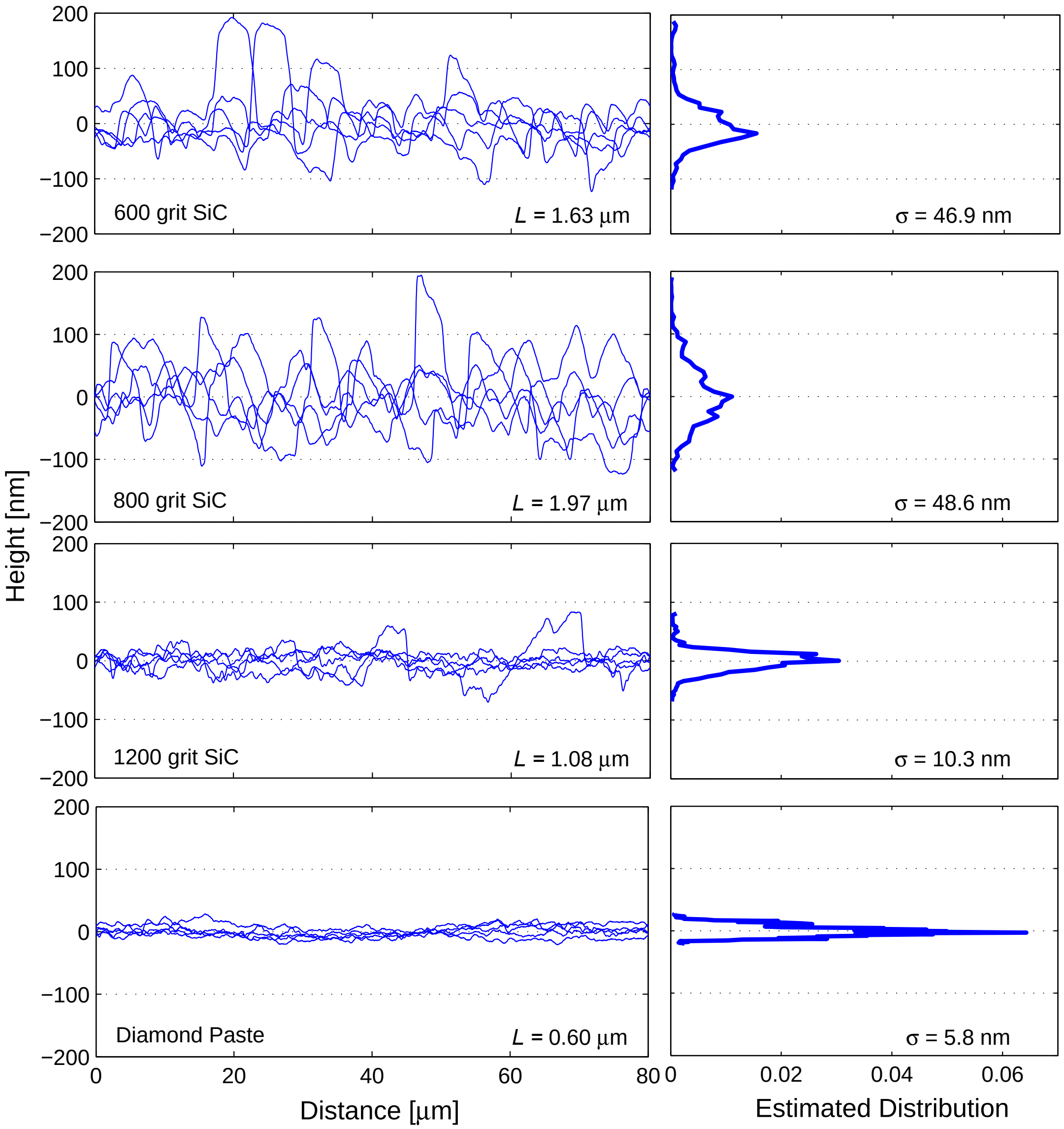


Figure 3.8: Profilometer measurements of polished electrode surfaces. Each surface was measured over a linear distance of 80 μm in 0.04 μm steps. Subplots on the left side contain superpositions of five sampled areas of the electrode surface. Subplots on the right side show the corresponding estimated distribution of the surface heights. The vertical path length, L, and standard deviation of the surface heights, σ, are displayed for each level of polishing.

PPy/PSS Electrodeposition

Polypyrrole/polystyrenesulfonate (PPy/PSS) is a conductive polymer that has been investigated previously for its use in neural probe applications [107, 108]. Malleo et al. showed that PPy/PSS has a remarkably small CPE magnitude compared to other common materials used for dielectric spectroscopy and does not suffer from poor reproducibility and biotoxicity like the platinum black treatment [87]. For electrodeposition on our electrodes, 0.2 M Pyrrole reagent (Sigma 131709) and 0.2 M Poly (4-styrenesulfonic acid) solution (Sigma 561223) were stirred for 15 minutes and then pipetted into the measurement cell. The cell electrodes were connected to the positive terminal of a DC power supply and the negative terminal was connected to a copper wire immersed in the solution through the open top hole. The power supply voltage (∼ 2 V) was adjusted to allow a current density of 1.5 mA/cm^2. Three different deposition times, 30 seconds, 3 minutes, and 6 minutes, were applied to three different measurement cells to access the role of film thickness. Slightly better performance was achieved with a 3 minute deposition time and therefore was used as the standard preparation of PPy/PSS electrodes in this work.

3.4.2 Results and Discussion

A typical fit to the measurements is shown in Figure 3.9. The fitting is very good even at the lowest and highest frequencies where the impedance analyzer is less accurate. Parameter extraction of κ and ν for various salt solutions is shown in Figure 3.10. Both parameters appear to be generally correlated with surface roughness. The values obtained using the diamond paste deviate somewhat from the trend, suggesting that the relationship between electrode surface conditions and the CPE parameters may be more complicated than simply the roughness profile left by sanding [89]. The macrogrit sandpaper exhibited κ values almost an order of magnitude smaller than the microgrit sandpapers, and PPy/PSS had κ values an additional 1 to 2 orders of magnitude below those. The small PPy/PSS CPE magnitude κ is consistent with the literature [87]. Because κ is small for the conductive polymer, indicating a small series electrode polarization contribution to the impedance relative to the solution impedance, measurements of ν show large variations at high salt concentrations.

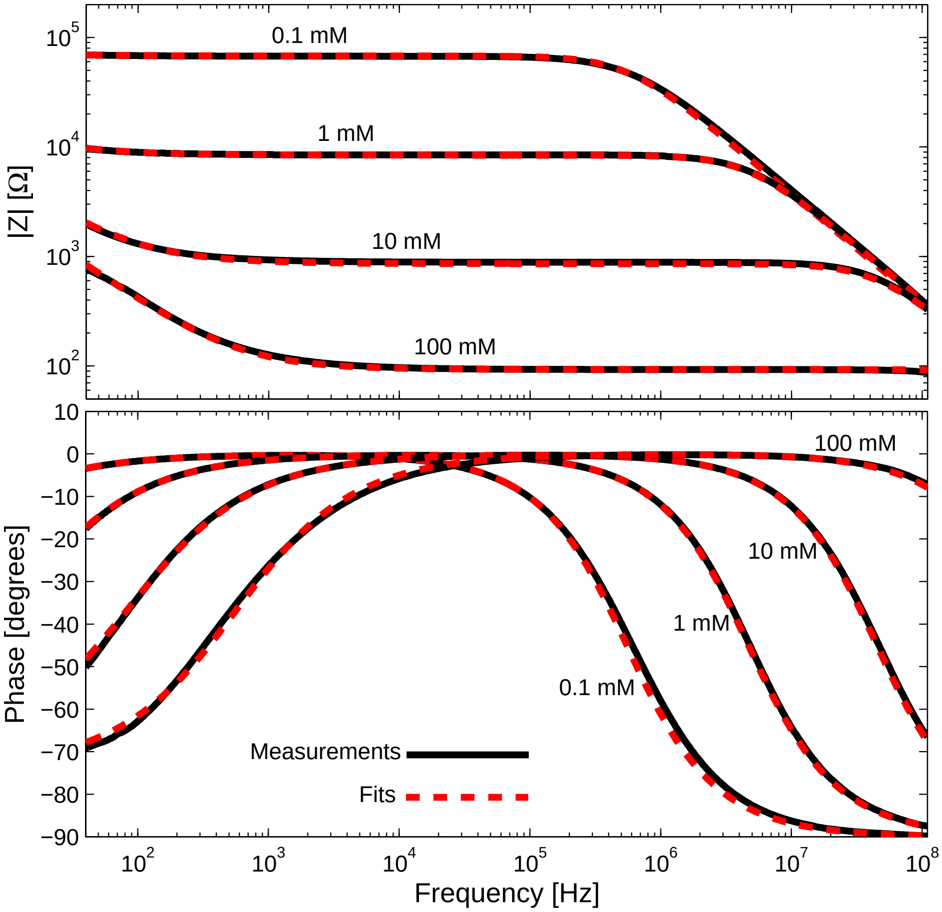


Figure 3.9: Impedance measurements (solid) and least-squares fits (dashed) for different NaCl concentrations. Measurements were made using the cell with 800 grit SiC polishing.

The net influence of electrode polarization on permittivity is evaluated from the onset frequency, f_{on} (Equation 3.30) and not from κ and ν alone. In Figure 3.11, f_{on} is plotted as a function of NaCl concentration for the various surface preparations. PPy/PSS coated electrodes increased the usable frequency bandwidth up to an order of magnitude from the sanded electrodes. The 60 grit sandpaper achieved the next lowest f_{on}, followed by the microgrit sandpapers in somewhat random order due to the complex relationship amongst κ, ν, and f_{on}. The full parameterization of κ, ν, C, G, and f_{on} is shown in Table 3.1. The solution conductivity σ strongly influences f_{on} ($f_{on} \propto \sigma^2$), and increased conductance does seem to be the single factor causing electrode polarization effects to increase at higher

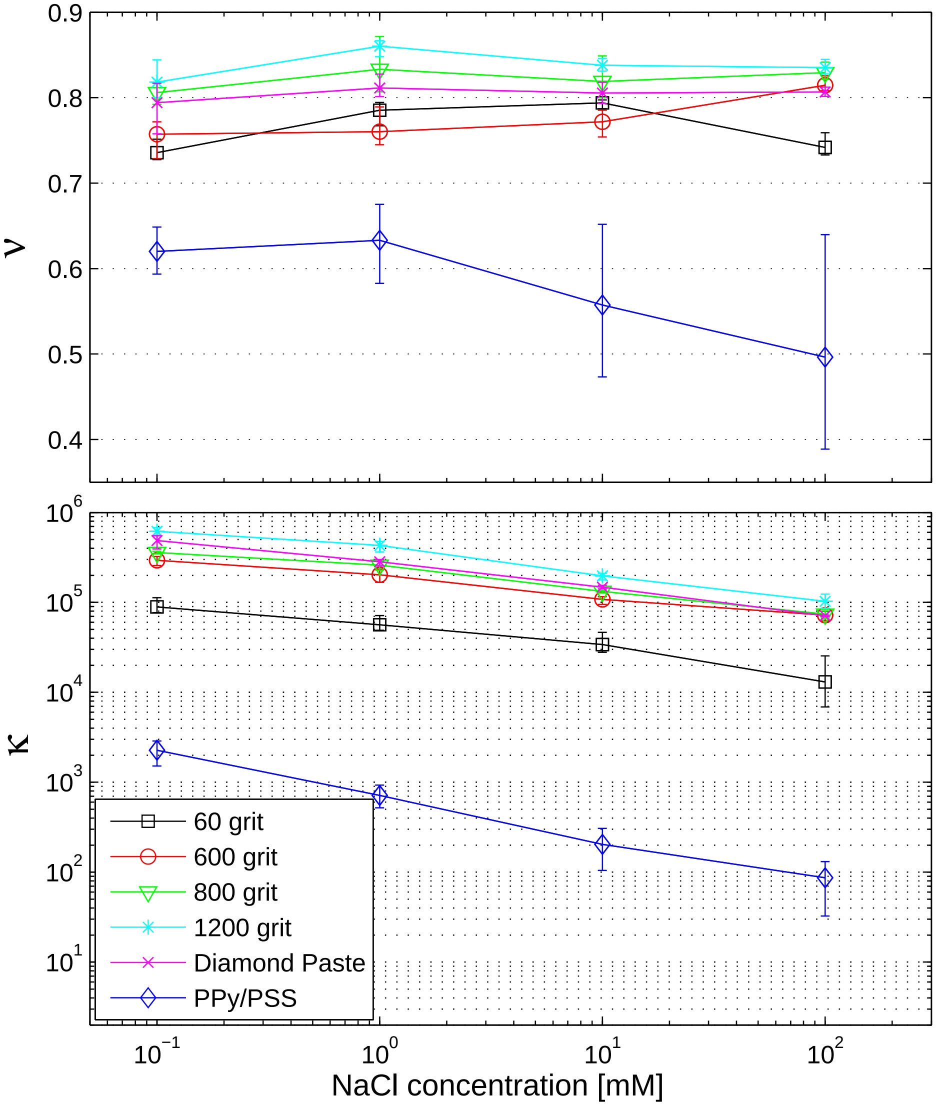


Figure 3.10: CPE parameters extracted from measured impedance spectra for solutions of increasing NaCl concentration. The legend indicates different electrode surface preparations. Error bars denote minimum and maximum values over three trials.

salt concentrations. Apart from surface preparation, the next important step for designing electrode polarization-limiting setups may be through manipulating the cell's geometry [95].

A protein experiment was conducted to verify the functionality of PPy/PSS electrodes on an understood biological system. β-Lg powder was dissolved at a concentration of 4 mg/ml in DI water. Starting with a dielectric cell with 60 grit sanded electrodes, permittivity spectra before and after electrodeposition of PPy/PSS clearly show bandwidth improvement

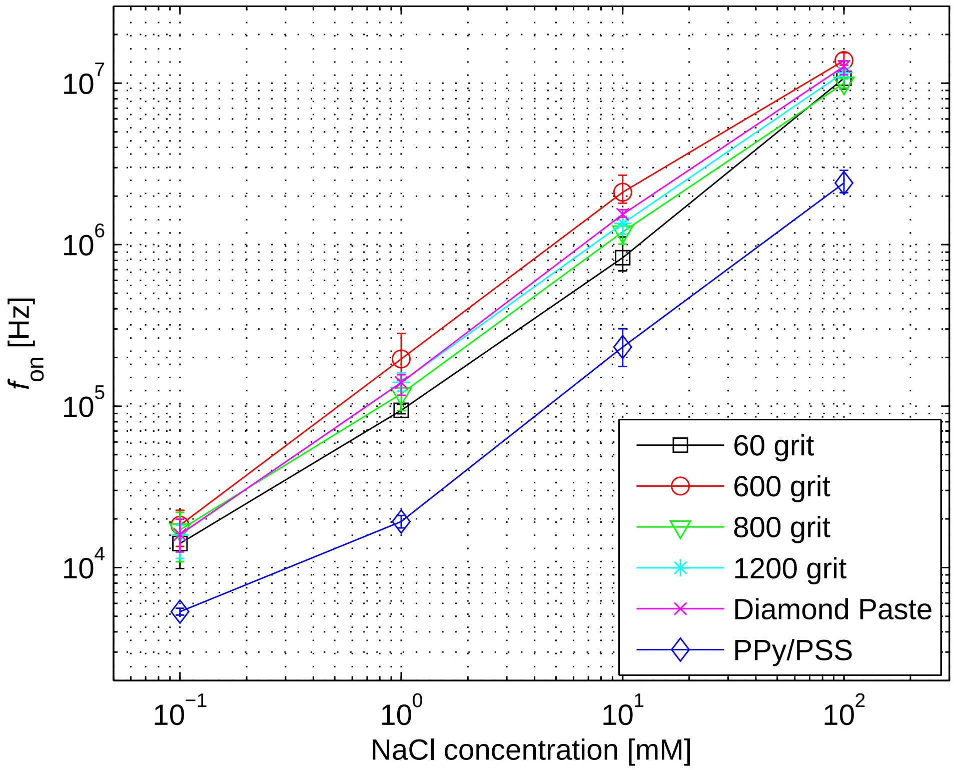


Figure 3.11: Onset frequency of electrode polarization f_{on} (Equation 3.30) for NaCl solutions. The legend indicates different electrode surface preparations. Error bars denote minimum and maximum values over 3 trials.

(Figure 3.12). The protein relaxation appears unaffected by the change in surface material, suggesting good biocompatibility of PPy/PSS for dielectric spectroscopy of protein solutions.

3.5 Summary

Many new measurement paradigms make use of microelectrodes for interrogating liquids and biological tissues. Double-layer formation at the electrode-electrolyte interface contributes to the measured impedance of the electrode surface in series with the substance that is being measured. In the permittivity spectrum, the electrode polarization contribution was derived quantitatively (Equation 3.28). This representation elucidates the $\omega^{-\text{const.}}$ (const. > 1) dependency in the real-valued permittivity spectrum and its proportionality to the CPE magnitude κ. A mathematical representation of the electrode polarization effect also allows it to be accounted for in model fittings and partially subtracted off. However, an-

Table 3.1: Parameterization of NaCl Solutions using model circuit in Figure 3.1.

0.1 mM NaCl	κ [10^5]	ν	C [pF]	G [μS]	f_{on} [kHz]
60 grit	0.89	0.74	5.15	16.8	14.1
600 grit	2.94	0.76	5.57	13.0	18.3
800 grit	3.58	0.81	4.26	14.5	17.1
1200 grit	6.16	0.82	4.60	11.3	15.9
Diamond Paste	4.89	0.79	4.74	11.1	16.0
PPy/PSS	0.023	0.62	4.82	26.5	5.3
1 mM NaCl	κ [10^5]	ν	C [pF]	G [μS]	f_{on} [kHz]
60 grit	0.56	0.79	5.10	152	94
600 grit	2.03	0.76	5.01	126	196
800 grit	2.60	0.83	4.08	117	119
1200 grit	4.30	0.86	4.15	127	140
Diamond Paste	2.83	0.81	4.40	113	140
PPy/PSS	0.0071	0.63	5.34	148	19
10 mM NaCl	κ [10^5]	ν	C [pF]	G [mS]	f_{on} [MHz]
60 grit	0.34	0.79	5.12	1.42	0.83
600 grit	1.08	0.77	5.30	1.54	2.11
800 grit	1.32	0.82	4.12	1.22	1.20
1200 grit	1.97	0.84	4.19	1.29	1.35
Diamond Paste	1.48	0.81	4.54	1.31	1.54
PPy/PSS	0.002	0.56	12.05	1.32	0.23
100 mM NaCl	κ [10^5]	ν	C [pF]	G [mS]	f_{on} [MHz]
60 grit	0.13	0.74	0.58	14.2	10.7
600 grit	0.72	0.81	1.47	13.1	13.8
800 grit	0.74	0.83	2.01	11.0	10.1
1200 grit	1.03	0.84	1.76	11.4	11.6
Diamond Paste	0.71	0.81	1.29	11.5	12.7
PPy/PSS	0.0009	0.50	2.70	11.0	2.4

alytical removal of electrode polarization can only be justified when the solution permittivity ε_S is known with confidence and is not completely obscured by polarization permittivity ε_{EP}.

Increasing electrode surface roughness has a considerable effect on the impedance of the electrode-electrolyte interface. Notably, both CPE magnitude κ and phase ν tend to decrease with increased roughness which perform counterproductive roles in the frequency value for the onset of electrode polarization. Nonetheless, the improvement gained through smaller κ exceeds the detriment of smaller ν, which encourages the general practice of roughening electrodes. In this work, electrode polarization effects in permittivity measurements were

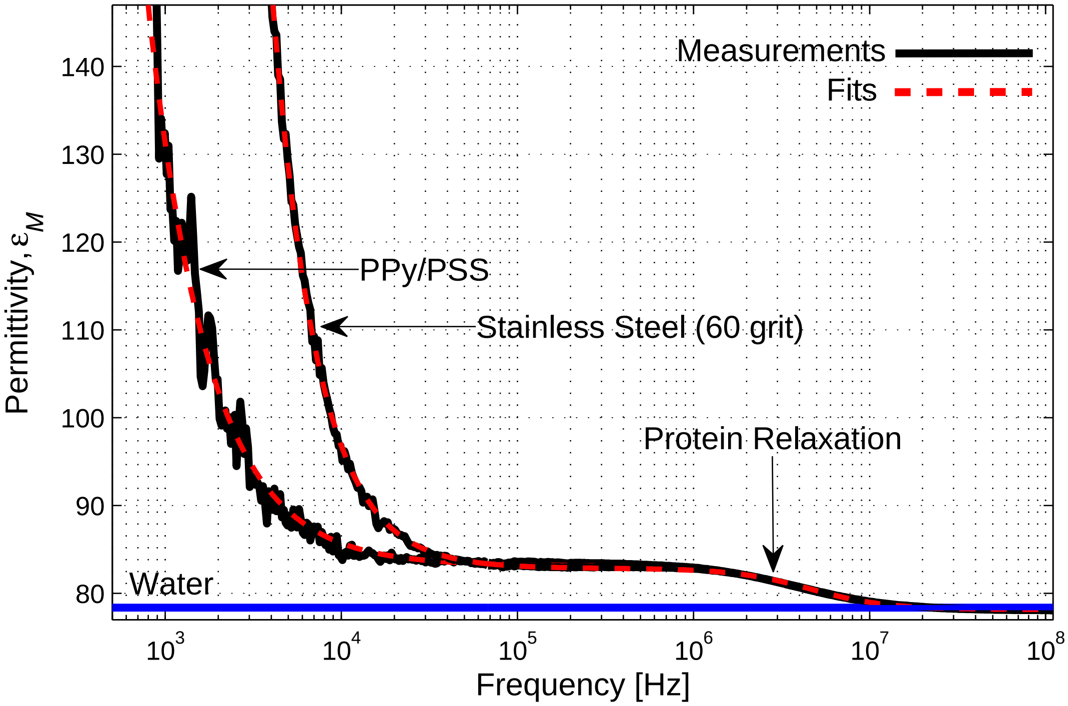


Figure 3.12: Measured permittivity of 4 mg/ml aqueous β-Lg using 60 grit polished stainless steel and PPy/PSS electrodes. Onset frequency of electrode polarization is visibly lower using the conductive polymer. Fits were made using Equation 3.25 (ε_S as defined in Equation 2.55).

minimized through a simple and robust process of depositing PPy/PSS onto the electrode surface. This procedure had the overall effect of greatly increasing the available frequency bandwidth for dielectric spectroscopy of liquids containing proteins and other biological objects.

Chapter 4

Measurement of the Protein Dipole Moment

The dipole moment is a low-resolution description of electrical charge distribution, which in practice has been used to identify DNA-binding proteins [109], characterize docking for protein-protein complexes [110], and predict protein aggregation rates [111]. These applications rely on the availability of accurate dipole moment data, often in various solution conditions. Experimentally, dipole moments are most commonly measured using dielectric relaxation spectroscopy [112, 113] and electrooptical techniques [114].

In Chapter 2, the various sources of polarization in liquids were discussed. Among these, orientational polarization was the slowest mechanism to respond to external fields. Orientational polarization was also shown to be proportional to the square of the dipole moment of the rotating molecule. This concept made the important connection between a microscopic mechanism, the dipole moment, and an experimentally obtainable quantity, permittivity.

In this chapter, the relationship between permittivity and the molecular dipole moment is clearly defined and explored. Additionally, it is shown that an effective hydrodynamic radius can be obtained from the permittivity spectrum. A case study of β-Lg is provided and the effects of temperature, concentration, and pH on the molecular parameters are analyzed. Before this can be accomplished, two important equations must be introduced: the Oncley equation and the Debye equation.

4.1 Oncley Equation

In Equation 2.47, the orientational polarization $\mathbf{P}_\mathrm{d}$ of a liquid in response to an applied electric field $\mathbf{E}$ was related to the square of the dipole moment. It is then possible

[35, 115] to arrive at the relationship between permittivity and dipole moment known as the Oncley equation,

$$\mu = \sqrt{\frac{2Mk_bT\varepsilon_0\delta}{Ng}} \tag{4.1}$$

where μ is the dipole moment, M is the protein molecular weight in kilodaltons, N is Avogadro's number, and g is the Kirkwood correlation parameter assumed to be 1 for dilute protein solutions [73]. The dielectric increment, δ, is defined as

$$\delta = \lim_{c \to 0} \frac{\Delta\varepsilon}{c} \tag{4.2}$$

where c is the molar concentration of the protein. The dielectric increment is evaluated as concentration approaches zero to eliminate interprotein effects.

4.2 Debye Equation

In 1897, Drude measured a large dispersion in the permittivity spectrum within the GHz range for certain groups of molecules such as OH and NH_2 [116]. Debye hypothesized that the decrease in permittivity Drude observed was due to orientational polarization and he set out to derive a mathematical model for the dielectric relaxation time and its reciprocal, the dielectric relaxation frequency. Debye posited a very simple model using spherical dipoles whose rotation is opposed by the viscosity of the solution. The relaxation time, τ, for such a system is given by

$$\tau = \frac{\zeta}{2k_bT} \tag{4.3}$$

where ζ is a molecular friction constant relating the torque applied to the molecule to the angular velocity. If the molecule is assumed to be a rigid sphere of radius a, Stokes' law gives a solution of

$$\zeta = 8\pi\eta a^3 \tag{4.4}$$

where η is the viscosity of the solution.

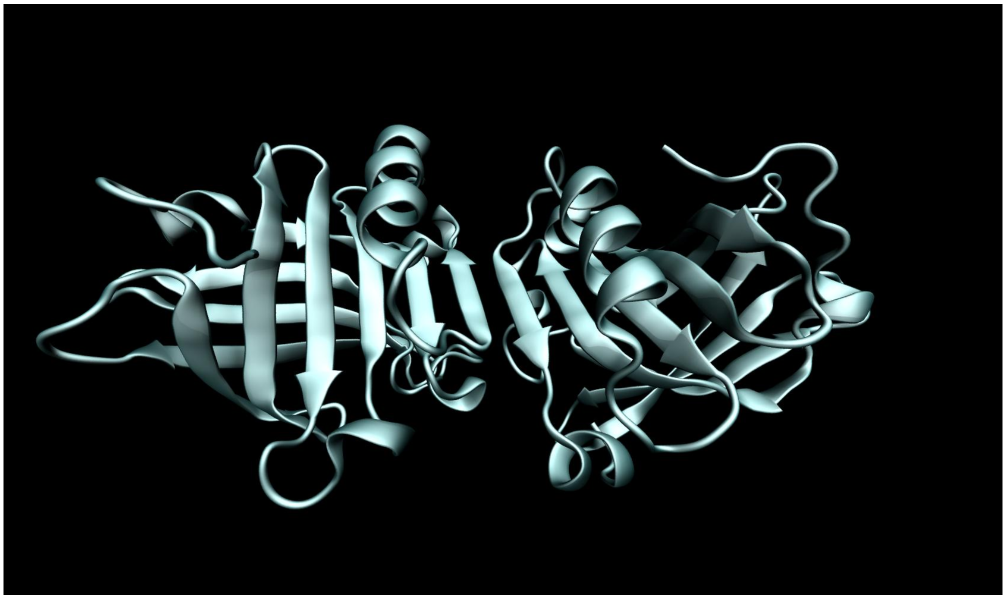


Figure 4.1: β-Lg dimer captured using X-ray diffraction [5], drawn using VMD [6].

Thus, assuming a protein is roughly spherical, its effective hydrodynamic radius can be estimated from the relaxation time taken from the fitted permittivity spectrum by

$$\tau = \frac{4\pi\eta a^3}{k_b T}. \tag{4.5}$$

Equation 4.5 is often referred to as the Debye equation [62]. It is noteworthy that this equation provides a very accurate prediction for the relaxation time of water. Using one-half the distance between adjacent oxygen molecules at 20 °C as the radius a, the predicted value for τ is 8.5 ps, in good agreement with the measured value of 9.3 ps [117].

4.3 Dielectric Relaxation of β-lactoglobulin

The Cole-Cole equation, Oncley equation, and Debye equation provide the necessary tools to examine the dipole moment and radius of a protein species using dielectric spectroscopy. For the first measurements, the dominant protein in cow's milk, β-Lg, was selected (Figure 4.1). β-Lg is an ideal candidate because it possesses a large permanent dipole moment and has interesting binding properties [98, 118]. Several genetic variants of β-Lg exist. The two most common, A and B, have similar molecular weights and differ only by

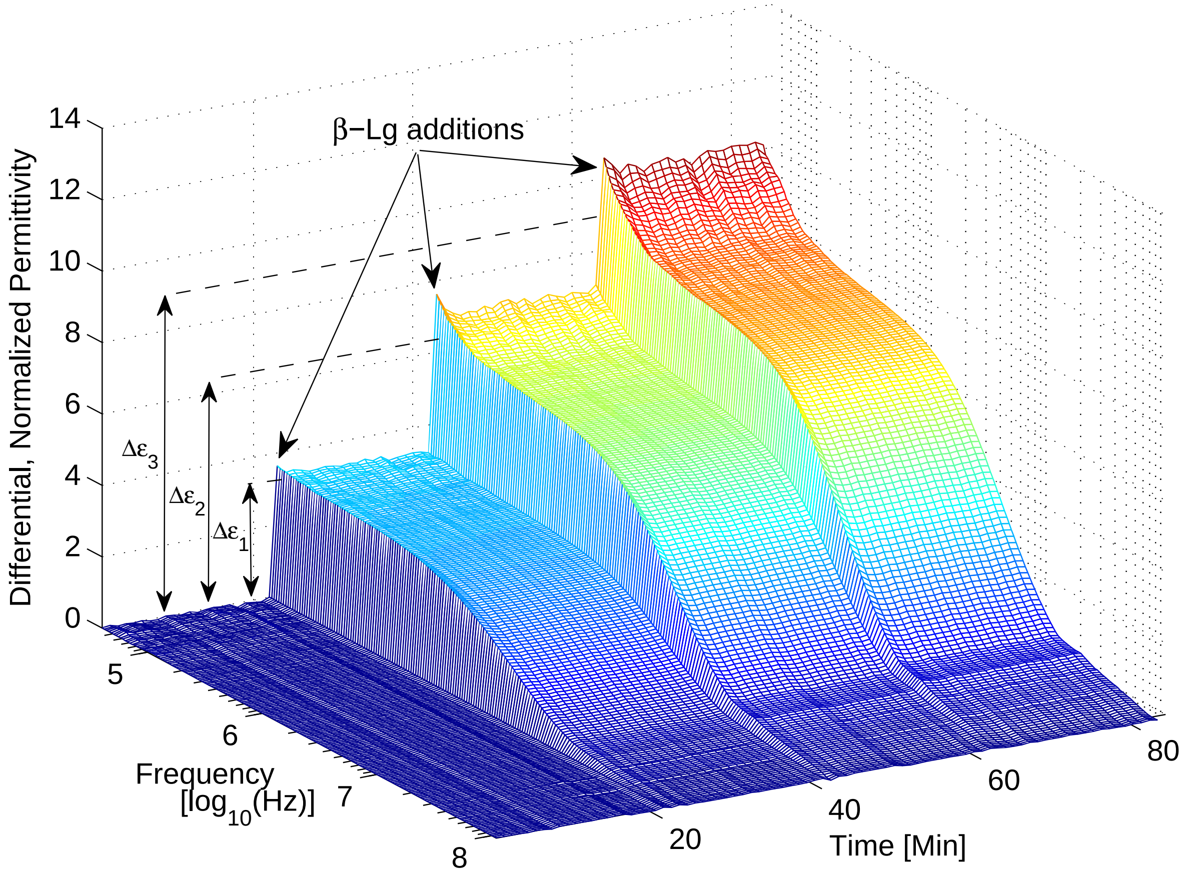


Figure 4.2: Plot of dielectric titration of β-Lg at 25 °C measured by dielectric spectroscopy. Differential, normalized permittivity was taken at multiple frequencies and frequency sweeps were taken at regular time intervals. $\Delta\varepsilon_1$, $\Delta\varepsilon_2$, and $\Delta\varepsilon_3$ represent least-squares fits of $\Delta\varepsilon$ (Equation 2.60) at three concentrations of β-Lg: 2 mg/ml, ∼4 mg/ml, and ∼6 mg/ml. The pH of the solution was ∼6.5. The relaxation time τ was ∼40 ns, which increased slightly at each β-Lg addition.

two amino acids (one of which is charged) and therefore can be expected to exhibit similar dielectric dispersions [119]. Previous treatments on the dielectric relaxation of β-Lg were limited and did not address the protein's dynamic electrical properties in acidic and basic solutions. The most noteworthy of these was the pioneering work done by Ferry and Oncley, who measured the dielectric increment of β-Lg in 0.25 and 0.50 M glycine at presumably neutral pH [70, 120].

Relaxation data for β-Lg at neutral pH is shown in Figure 4.2 along with least-squares fits of $\Delta\varepsilon$ (Equation 2.60) to the equilibrium frequency sweeps at each of three β-Lg

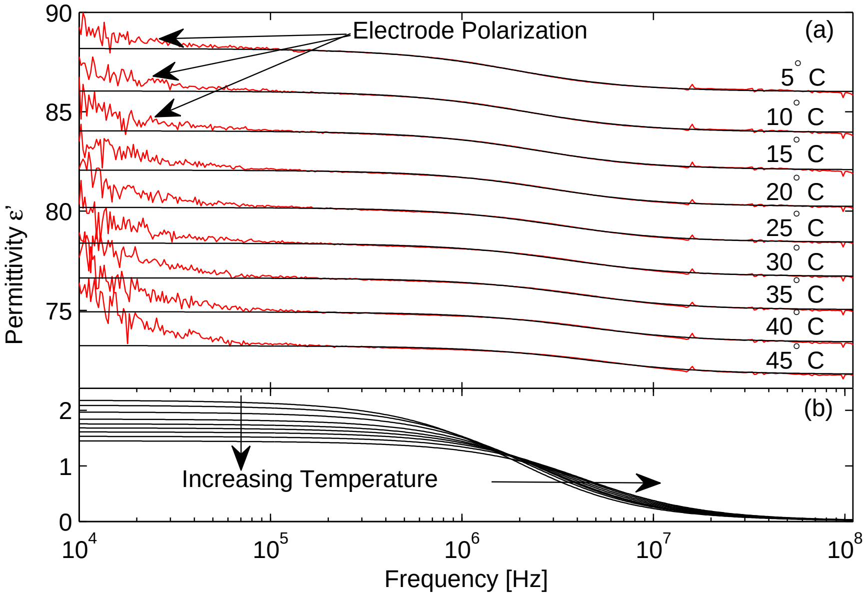


Figure 4.3: Temperature dependence of the dielectric relaxation of β-Lg. (a) Results from nine identical experiments conducted at temperatures from 5 °C and 45 °C are plotted with fits using Equation 2.60. (b) Superposition of fits clearly show temperature dependence of $\Delta\varepsilon$ and τ.

concentrations. This data was recorded using the titration capabilities of the apparatus. Every 20 minutes, 80 μl of 20 mg/ml β-Lg was pipetted into the 0.8 ml volume cell, which was initially filled with DI water. Before each addition, 80 μl was also removed to keep the overall volume constant. Visible at each addition are permittivity contributions from the protein and effects of electrode polarization at low frequencies.

4.3.1 Temperature and Concentration Dependence

It was desired to measure the effects of temperature on the dielectric relaxation of β-Lg. The Oncley equation predicts an inverse relationship between $\Delta\varepsilon$ and temperature and the Debye equation predicts an inverse relationship between τ and temperature. This was tested by conducting identical experiments at 5 °C intervals between 5 °C and 45 °C. The results are shown in Figure 4.3 and clearly support the theory. At high temperatures thermal agitation disrupts molecular movement, causing permanent dipoles to align less with the applied field.

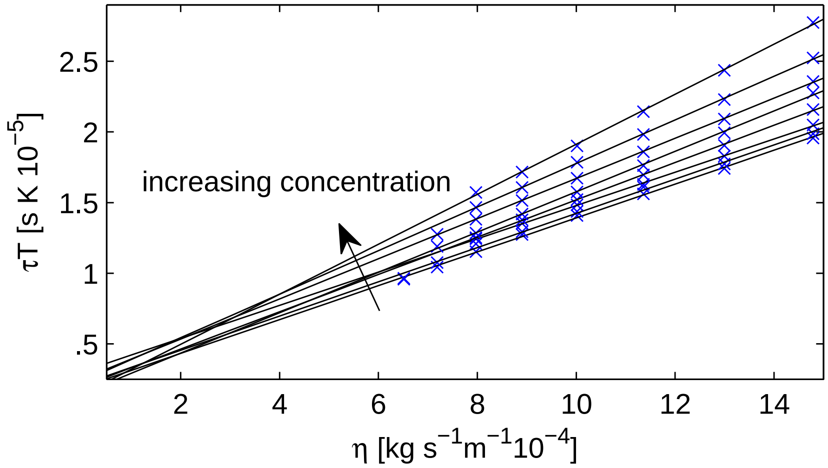


Figure 4.4: Relaxation time τ (Equation 4.5) multiplied by temperature T against estimated solvent viscosity η for β-Lg in 0.1 mM HCl. Measured values (×) are plotted with linear least-squares fits (solid lines). Concentrations of β-Lg used were 1, 2, 3, 6, 8, 10, 15, and 20 mg/ml.

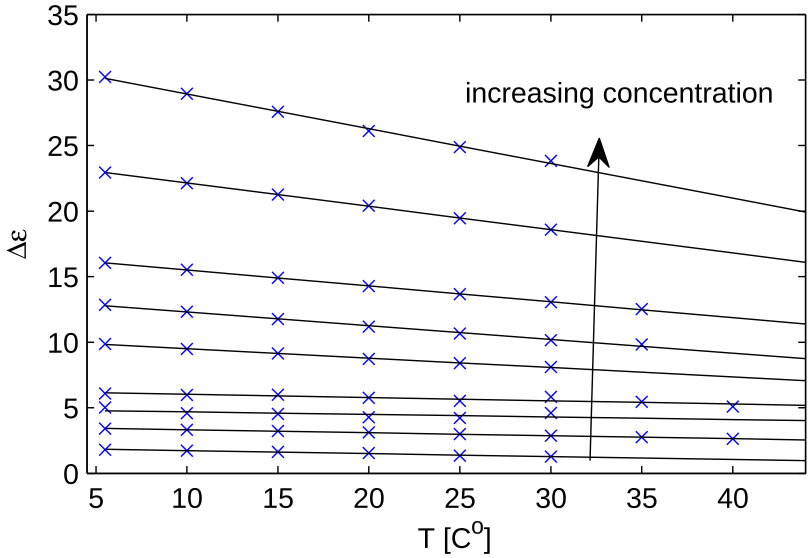


Figure 4.5: Change in permittivity $\Delta\varepsilon$ (Equation 4.1) against temperature T for β-Lg in 0.1 mM HCl. Measured values (×) are plotted with linear least-squares fits (solid lines). Concentrations of β-Lg used were 1, 2, 3, 4, 6, 8, 10, 15, and 20 mg/ml.

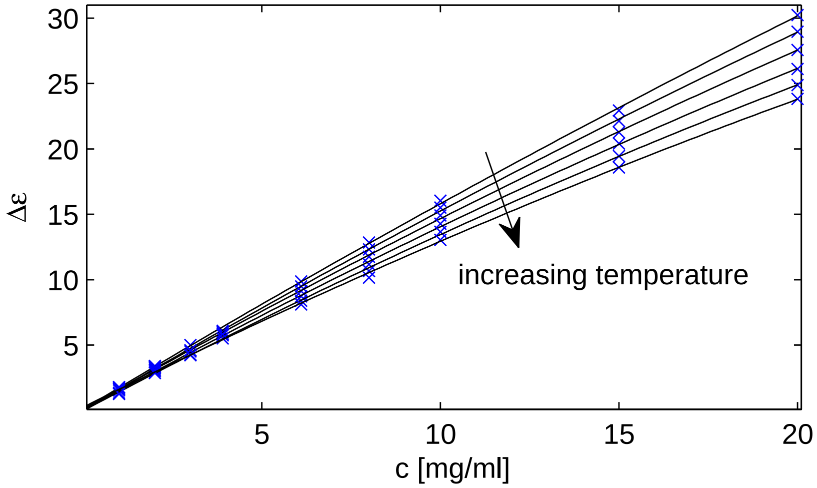


Figure 4.6: Change in permittivity $\Delta\varepsilon$ (Equation 4.1) against concentration c for β-Lg in 0.1 mM HCl. Measured values (×) are plotted with least-squares fits of second-order polynomials (solid curves, Equation 4.6). Temperature was adjusted from 10 to 35 °C in 5 °C increments.

Figures 4.4-4.6 show further experiments demonstrating temperature and concentration dependence also obtained in a 0.1 mM HCl solution (constant pH). In Figure 4.4 the relationship between (relaxation time × temperature) and solvent viscosity was linear for all concentrations as predicted by Equation 4.5. Values for solvent viscosity were taken from [121]. Figure 4.5 likewise shows a linear relationship between the change in permittivity and temperature as predicted by Equation 4.1. The increasing slope of the linear fits reflects protein-protein effects at increasing concentration.

Figure 4.6 shows the nonlinear relationship between change in permittivity and concentration at various temperatures. To evaluate the dielectric increment as solute concentration approaches zero, measurements of $\Delta\varepsilon$ and c were fitted to a second-order interpolation equation of the form

$$\Delta\varepsilon = k_2 c^2 + k_1 c + k_0 \tag{4.6}$$

where k_1 is the value used for $\Delta\varepsilon/c$ extrapolated to infinite dilution (for Equation 4.1). The $k_2 c^2$ term accounts for the decrease in permittivity at high concentrations and is a measure of the protein-protein interaction caused by interfering electric fields of neighboring β-Lg molecules [70]. Thus, k_2 typically takes on a negative value, as interference of molecular rotation reduces permittivity.

Using Equation 4.6, the value for k_2 of β-Lg in 0.1 mM HCl at 25 °C was found to be -0.010. This value, along with the value found for k_0 ($\sim$0), was substituted into Equation 4.6 and rearranged to form the equation

$$\delta = k_1 = \frac{\Delta\varepsilon}{c} + 0.01c. \tag{4.7}$$

For the dimeric β-Lg, Equation 4.1 can be reduced, assuming $g = 1$, and expressed in Debye units as

$$\mu = 36.3\sqrt{T\delta}. \tag{4.8}$$

Combining Equations 4.7 and 4.8 and assuming a constant temperature of 25 °C in the dielectric cell yields

$$\mu = 626.8\sqrt{\frac{\Delta\varepsilon}{c} + 0.01c}. \tag{4.9}$$

In this fashion, a single spectral measurement of β-Lg at a known concentration was sufficient to compute an electrical dipole moment μ at a given pH.

4.3.2 pH Dependence

To modify the solution pH during the experiment, small volumes of 1 M HCl or 1 M NaOH were pipetted into the cell reservoir, thus maintaining the concentration of protein essentially constant. To lessen electrode polarization effects, the ionic strength of the solution was minimized by reconstituting the protein in DI water and subsequently adding either HCl or NaOH instead of using buffer solutions to achieve a desired pH. The pH was measured by inserting a 3 mm diameter Accumet MicroProbe into the cell cavity immediately after data collection.

Measurements of the effective hydrodynamic radius at various pH were made using Equation 4.5 and are shown in Figure 4.7. Values are in the range of $\sim$24$-$30 Å. The changes in apparent radius at different pHs may be due to structural fluctuations and changes in the

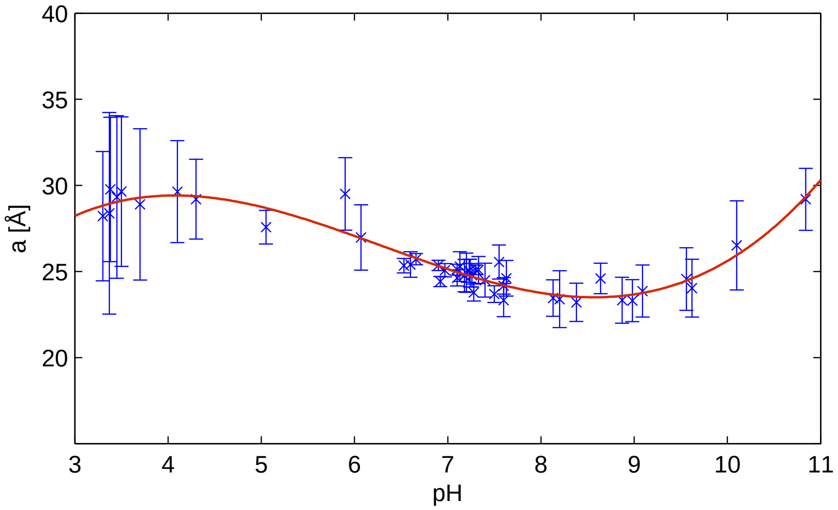


Figure 4.7: Hydrodynamic radii a (Equation 4.5) of β-Lg at various pH. The error bars represent the minimum and maximum values of several different least-squares fits of the measured data. The solid red spline curve is displayed to show the general trend of the measurements.

thin layer of bound water associated with the macromolecule. Larger fluctuations in radii of ±5 Å were observed by Bonincontro et al., who measured the relaxation parameters of lysozyme at various pH [122].

Dipole moment measurements at various pH are shown in Figure 4.8. The dipole moment fluctuates to ~ 550 D at low pH from its maximum value of ~ 850 D around pH 7. At low and high pH, the greater uncertainty in the measurement is due to an increase in the ionic strength of the solution. The relaxation parameters in Figures 4.7 and 4.8 demonstrate the dynamic behavior typical of proteins dissolved in solution. These measurements will provide an important reference for the electrostatic computational work done in the next chapter.

4.4 Analytical Removal of Electrode Polarization Effects from Protein Relaxation Spectra

The derivation of ε_M in Chapter 3 provides a simple method to remove electrode polarization through post-processing of measured permittivity data. It can be applied to a wide number of different solution permittivities. However, here the focus is proteins dis-

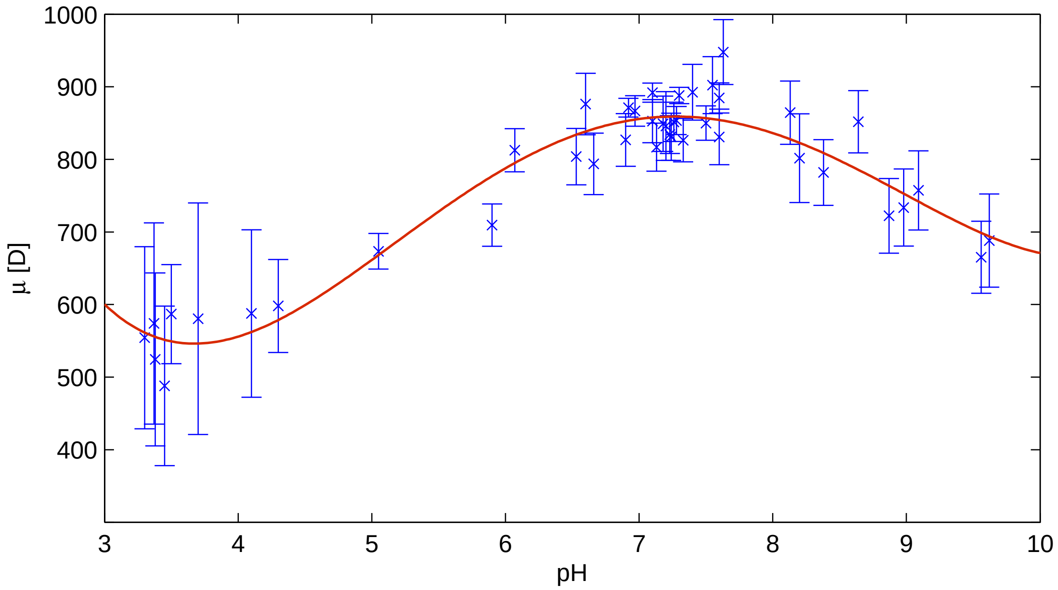


Figure 4.8: Measured dipole moment μ of β-Lg at various pH. Error bars represent the minimum and maximum values of several different least-squares fittings of the measured data. The solid red line is displayed to show the general trend of the measurements.

solved in solution where ε_S is described by the Cole-Cole equation (Equation 2.60). The full expression for ε_M is then given by

$$\varepsilon_M = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + (j\tau\omega)^{1-\alpha}} + \overline{\kappa}\omega^{-\overline{\nu}}. \tag{4.10}$$

For β-Lg in a less than 1 mM ionic strength buffer, use of ε_M (in place of ε_S) is not necessary for accurate parameterization due to the large SNR of the sample. For other proteins, electrode polarization effects are more pronounced and omission of ε_{EP} in the fitting function may lead to large errors. Figure 4.9 shows one such protein, avidin, which will be investigated in depth in Chapter 7. The two red lines represent permittivity before and after electrode polarization effects are removed using a fitting of Equation 4.10. The overlap of the two lines indicates the frequency range where electrode polarization effects are negligible and the original Cole-Cole equation could be fitted. For this sample, reliable fits of $\Delta\varepsilon$ and τ require use of the wider frequency range and fitting equation 4.10.

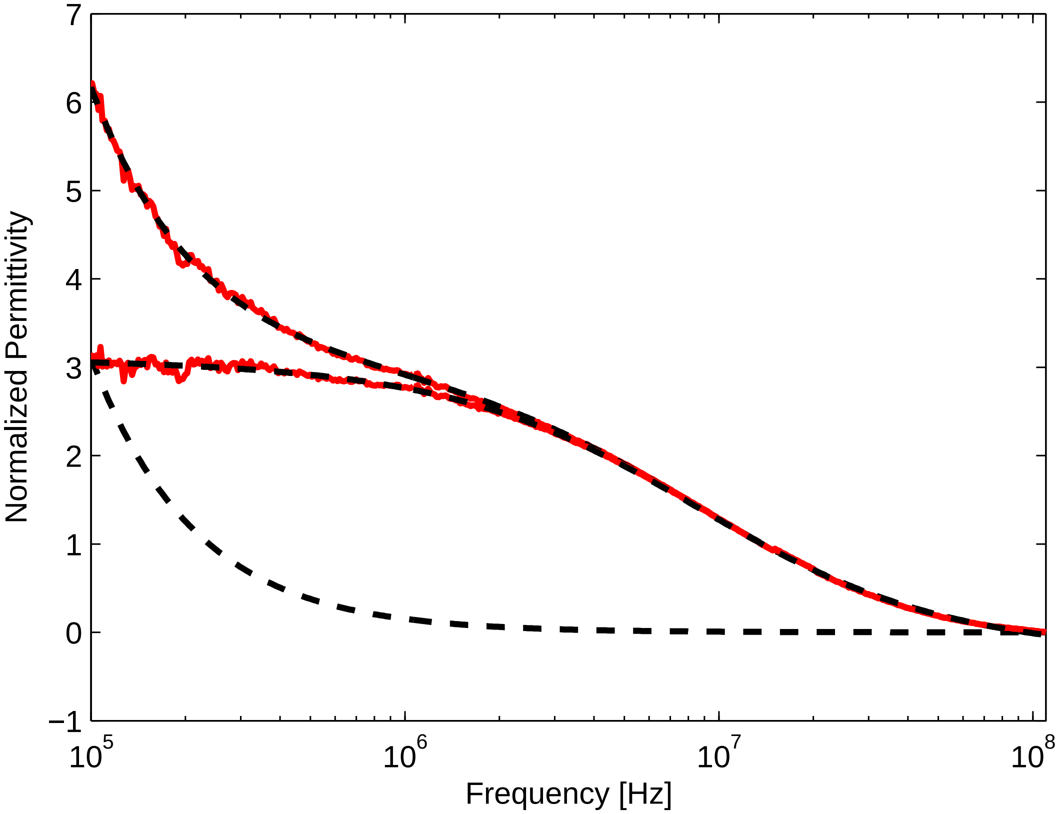


Figure 4.9: Removal of electrode polarization from permittivity spectrum of 3 mg/ml avidin. Upper red line is the raw measured permittivity and lower red line is the same with ε_{EP} subtracted. Dashed black lines represent fits of ε_M, ε_S, and ε_{EP} (in order of decreasing permittivity at 200 kHz).

Chapter 5

Theoretical Method to Calculate Protein Charge Moments

Charge distributions are linked to form, function and stability of biological macromolecules [21, 123, 124, 125], and consequently are dependent on the molecule's internal shape and composition, as well as the external influences of the surrounding solution. While different measures such as molecular weight, isoelectric point, and overall charge are commonly quoted, perhaps no single parameter is more indicative of the spatial charge distribution of a protein than its electric dipole moment. In the previous chapter, a technique to extract a measured dipole moment from permittivity measurements was discussed. In this chapter, a theoretical method to calculate the protein dipole moment and charge moments in general is presented. This technique takes the basic laws of electrostatics and applies them to simple molecular systems under the assumption that proteins and polypeptides can be modeled sufficiently well by discrete charge distributions. It will be shown that this method is in fact quite accurate and its application reveals new information about the electrical properties of proteins and biomolecules.

5.1 Charge Moments for Continuous and Discrete Charge Distributions

Charge moments are statistics on charge distributions. The first-order charge moment of a system is the net charge, or sum of all charge. For a continuous charge density $\rho(\mathbf{r})$, the overall charge, Q, is given by

$$Q = \int_V \rho(\mathbf{r})\, dV. \tag{5.1}$$

The second-order charge moment of a system is its dipole moment. The general dipole moment equation for a continuous charge density is

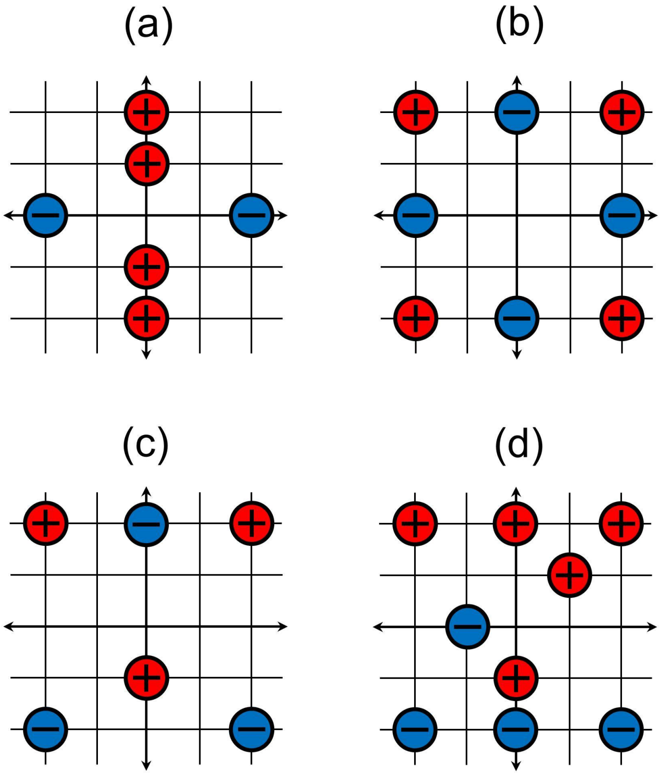


Figure 5.1: Discrete charge distributions with (a) $Q > 0$, $|\boldsymbol{\mu}| = 0$; (b) $Q = 0$, $|\boldsymbol{\mu}| = 0$; (c) $Q = 0$, $|\boldsymbol{\mu}| > 0$ (pointing upward); and (d) $Q > 0$, $|\boldsymbol{\mu}| > 0$ (pointing upward and right). The origin is used as the point of reference.

$$\boldsymbol{\mu} = \int_V \rho\left(\mathbf{r}\right)\left(\mathbf{r} - \mathbf{r}_r\right) dV \tag{5.2}$$

where $\mathbf{r}_r$ is the vector from the origin to the reference point. By convention (in physics), the dipole moment vector points from the direction of negative charge towards the direction of positive charge.

The discrete analogues of Equations 5.1 and 5.2 are

$$Q = \sum_i q_i \tag{5.3}$$

and

$$\boldsymbol{\mu} = \sum_i q_i \left(\mathbf{r}_i - \mathbf{r}_r \right) \tag{5.4}$$

where $\mathbf{r}_i$ is the vector from the origin to point charge q_i. Some examples of simple discrete charge distributions and their first- and second- order charge moments are displayed in Figure 5.1.

In a neutrally charged system, the dipole moment is not dependent on the reference point. In such a case, $\boldsymbol{\mu}$ reduces to

$$\begin{aligned} \boldsymbol{\mu} &= \sum_i q_i \left(\mathbf{r}_i - \mathbf{r}_r \right) && (5.5) \\ &= \sum_i q_i \mathbf{r}_i - \mathbf{r}_r \sum_i q_i && (5.6) \\ &= \sum_i q_i \mathbf{r}_i . && (5.7) \end{aligned}$$

In a non-neutrally charged system, where the net charge is nonzero, two popular reference points are the center of mass, $\mathbf{r}_{\mathrm{cm}}$, expressed as

$$\mathbf{r}_{\mathrm{cm}} = \frac{\sum\limits_i m_i \mathbf{r}_i}{\sum\limits_j m_j} \tag{5.8}$$

and the center of charge, $\mathbf{r}_{\mathrm{cq}}$, expressed as

$$\mathbf{r}_{\mathrm{cq}} = \frac{\sum\limits_i \left| q_i \right| \mathbf{r}_i}{\sum\limits_j \left| q_j \right|}. \tag{5.9}$$

From the literature, it appears the best reference point might be the center of diffusion [126, 127]. However, this would require additional simulations and is not expected to improve the calculation beyond the error of the measurements. The center of charge has been used by other investigators at pH close to the isoelectric point [128].

It was discovered early on that the center of mass in dipole moment calculations yields dipole moments in better agreement with measurements. By using the general dipole moment equation with the center of mass as the reference point, we focus on the torque about that point produced by applying an electric field to a molecule, and disregard the translational movement of the molecule.

5.2 Dipole Moment Calculation Using the Protein Data Bank

The proposed method for calculating a protein's dipole moment uses a snapshot of the molecule's three-dimensional structure. The dipole moment is then calculated as the vector sum of two components: (1) the core dipole moment which arises from the unequal sharing of electrons in covalent bonds; and (2) the dipole moment resulting from partial charges at amino acid side chains. The former is a neutral charge system and can be computed with Equation 5.7 and the latter is only neutral at pI and thus must use Equation 5.4.

The Protein Data Bank (PDB) is an online repository for structural data of biological molecules (www.pdb.org [129]). The vast majority of the molecules are proteins and nucleic acids. Structural biologists typically obtain a molecule's three-dimensional structure by one of two methods, X-ray crystallography or NMR spectroscopy. Both methods provide atomic-level detail for individual conformations of proteins. Figures 5.2 and 5.3 show families of β-Lg and HU-Protein backbone structures consistent with NMR observables.

Typically, only NMR structures have enough resolution to accurately identify hydrogen atom locations. When working with proteins in solution, NMR data is preferred over X-ray crystallographic data because it is taken in solution and is thus more similar to experimental conditions. Additionally, NMR data contains different variations of structures so statistics of charge organization can be computed. Currently about 85% of the structures in the PDB are from X-ray crystallography, although the percentage of NMR structures is slowly increasing.

5.2.1 Dipole Moment from Covalent Bonds

Due to the difference in electronegativity between different atoms, permanent dipoles exist at covalent bonds throughout the molecular structure. Although individually weak, in

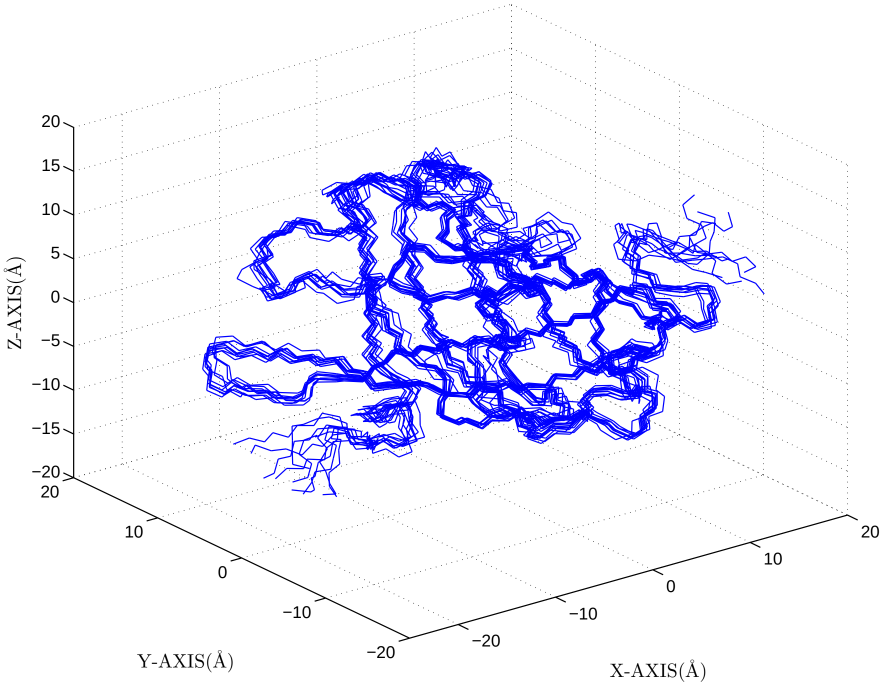


Figure 5.2: β-Lg (PDB 1cj5) protein backbone of the monomer at pH 2.0 [7]. Structural variability is apparent and shows the instabilities of the solved structure. The plot is a superposition of the ten variations in the PDB file.

a macromolecule their contributions can become significant. Their net dipole is referred to as the core dipole moment.

The core dipole moment is calculated from the locations of N–H, C=O, and C–N bonds in both side chains and protein backbone, using dipole moments of 1.31 D, 2.31 D, and 0.2 D, respectively [130]. For a system of symmetric dipole moments with an arbitrary reference point, the general dipole moment equation (Equation 5.4) can be rewritten:

$$\boldsymbol{\mu}_{\text{core}} = \sum_i \boldsymbol{\mu}_i \tag{5.10}$$

where the length of the summation is the total number of N-H, C=O, and C-N bonds in the structure.

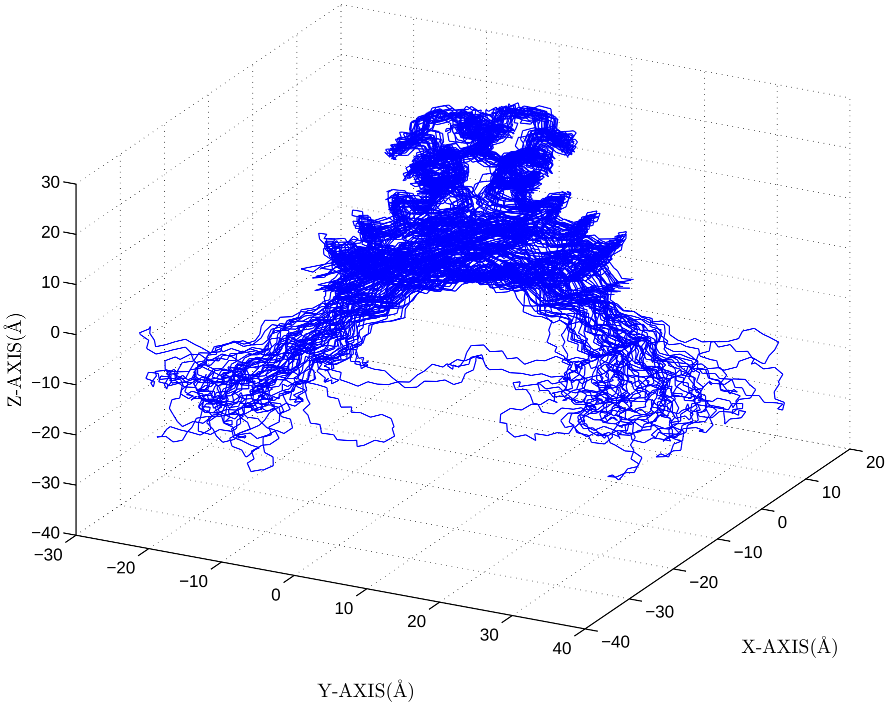


Figure 5.3: HU-Protein (PDB 1hue) backbone at pH 4.6 [8]. The visible flexibility of the homodimeric protein is critical for its function packaging high molecular weight DNA in prokaryotic cells. The protein exhibits the unique characteristic of being partially rigid around the head while the β-arms are variable in solution. The plot is a superposition of the 25 models in the PDB file.

5.2.2 Calculation of pK_a Shifts

Partial charges that exist at side chains of ionizable residues are the main contributors to the charge moments of proteins. These charges are governed by the Henderson-Hasselbalch equations [131], i.e.,

$$Q_+ = \frac{1}{1 + 10^{pH - pK_a}} \tag{5.11}$$

for positively charged residues such as arginine, histidine, lysine, and N-terminal, and

$$Q_- = \frac{10^{pH - pK_a}}{1 + 10^{pH - pK_a}} \tag{5.12}$$

for negatively charged residues such as aspartic acid, glutamic acid, tyrosine and C-terminal, where Q_+ and Q_- are the fraction of the particular titrating residue in the charged state. The fractional charge assigned to the residue is then given by ZQ_+ and ZQ_- where Z is the formal charge of the ion. When $\mathrm{pH} = \mathrm{p}K_\mathrm{a}$ for a specific residue, it becomes 50% ionized. Although not all titrating residues closely follow the Henderson-Hasselbalch model, it is believed to be sufficient for low resolution charge statistics.

Electrostatic interactions among charged sites cause $\mathrm{pK_a}$ shifts of polar groups from their intrinsic (bulk aqueous solution amino acid) values, which can be estimated by calculation and compared to measured values. Among the current methods used to compute the $\mathrm{p}K_\mathrm{a}$s of residues, the continuum dielectric approach has proven to give a reasonable balance between speed and accuracy [132, 133, 134]. The model relates the $\mathrm{p}K_\mathrm{a}$ of a chemical group M in a proton-exchange reaction with water

$$HM + H_2O \longleftrightarrow M^- + H_3O^+ \tag{5.13}$$

to the standard Gibbs free energy $\triangle G$ for proton exchange through a Boltzmann distribution of the reaction species,

$$\mathrm{pK_a^{calc}} = \mathrm{pK_a^m} + \frac{1}{2.303 k_b T}\left[\triangle G\left(M_p H^+ \longrightarrow M_p\right) - \triangle G\left(M_s H^+ \longrightarrow M_s\right)\right] \tag{5.14}$$

where M_p and M_s are the chemical group M in the protein and solution environments, $\mathrm{pK_a^{calc}}$ is the shifted $\mathrm{p}K_\mathrm{a}$, and $\mathrm{pK_a^m}$ is the model $\mathrm{p}K_\mathrm{a}$ of the chemical group M. For a continuous charge distribution, electrostatic contributions to the free energy are solved from the Poisson-Boltzmann equation:

$$\nabla \varepsilon\left(\mathbf{r}\right) \nabla \phi\left(\mathbf{r}\right) - \varepsilon\left(\mathbf{r}\right) \kappa\left(\mathbf{r}\right)^2 \sinh\left[\phi\left(\mathbf{r}\right)\right] + 4\pi\rho\left(\mathbf{r}\right)/k_b T = 0 \tag{5.15}$$

where $\phi\left(\mathbf{r}\right)$ is the electrostatic potential and $\kappa\left(\mathbf{r}\right)^2$ is related to the Debye length, accounting for the ionic strength of the bulk solution [135].

The permittivity $\varepsilon(\mathbf{r})$ is defined as a constant ε_{low} inside the protein and a constant $\varepsilon_{\text{high}}$ outside the protein. The assumption of a uniform permittivity throughout the protein clearly is an approximation [136]. The actual permittivity is likely to be a function of several factors, including density of packed residues, molecular cavities, and solvent penetration. The ε_{low} approximation, however, has the advantage of being simple, reducing the computational expense [132]. A pK_a shift is calculated from the difference in electrostatic energy of an amino acid residue in its charged and neutral state.

To aid in this computation, several web servers are now available. The H++ server, used primarily for estimating p$K_\text{a}s$ and adding missing hydrogen atoms to molecules from the PDB, utilizes the continuum dielectric model to predict protonation states of ionizable residues in macromolecules [19]. The H++ calculation is based on the "Macromolecular Electrostatics with Atomic Detail" (MEAD) software package [137]. The server sets up the finite-difference and Poisson-Boltzmann equations, defining the boundary between the internal and external dielectric regions as the surface 1.4 Å outside the protein, and taking as user input, ε_{low}, $\varepsilon_{\text{high}}$, and the salt concentration which influences the calculation through the $\kappa(\mathbf{r})^2$ term.

5.2.3 Dipole Moment from Partial Charges

pK_a values were calculated by the H++ server for each titratable group of the protein and Equations 5.11 and 5.12 were used to assign a partial charge value ranging from -1e to +1e to each residue. The locations of partial charges were approximated using the nitrogen atom location for amine groups and the oxygen atom location for carboxyl groups.

The total dipole moment is the vector sum of the core and side chain dipole moments. Figure 5.4 shows the point charges and the total dipole moment for β-Lg at pH 4.9 with pK_as calculated with H++ ($\varepsilon_{\text{low}} = 6$).

5.3 Calculated and Measured Dipole Moments of β-lactoglobulin

The dipole moment calculation method was first used on β-Lg for comparison to the measurements made in the previous chapter. Calculated dipole moments of β-Lg comparing Poisson-Boltzmann predictions based on 1beb and measurements based on Equation 4.9 are

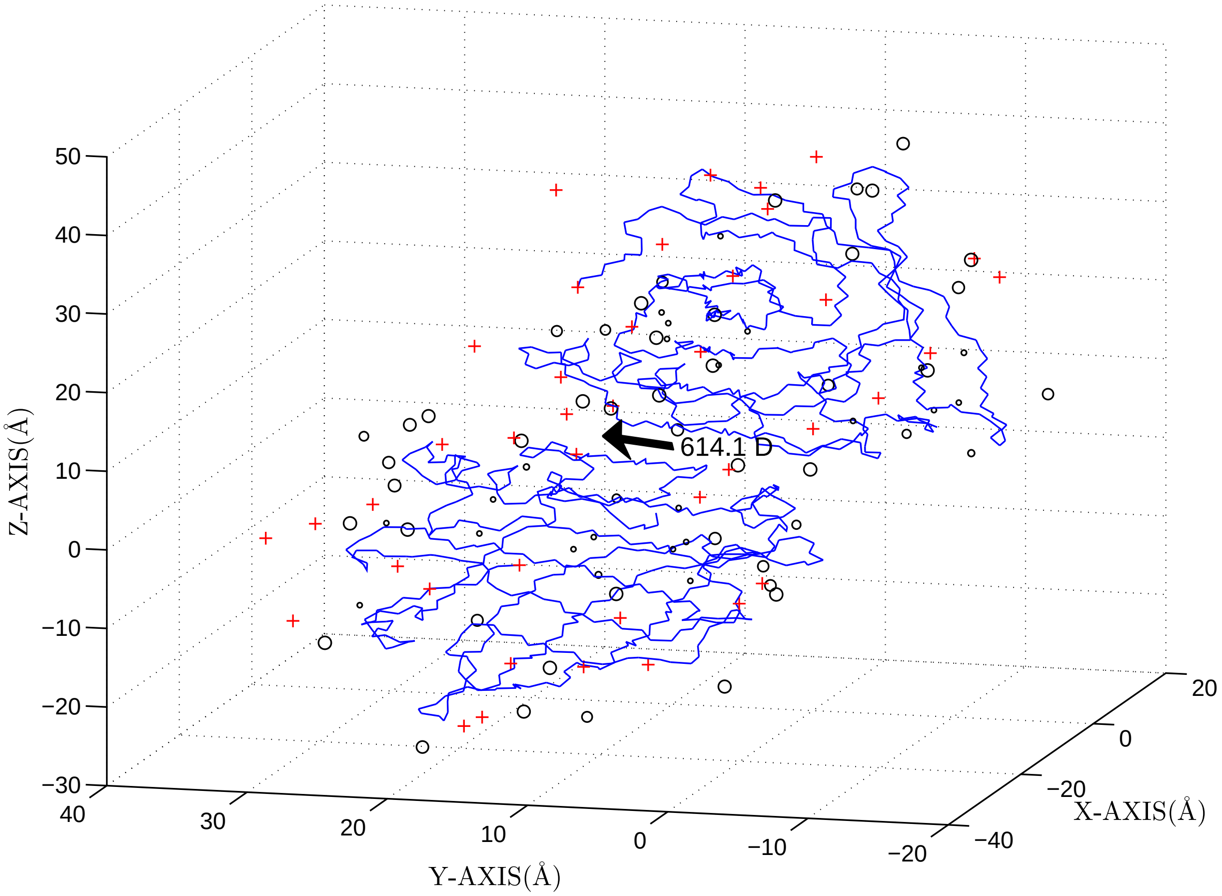


Figure 5.4: β-Lg (PDB 1beb) partial charges at pH 4.9, near the isoelectric point of the protein, pH 5.1. The vector dipole of 614.1 is shown. Positive charges are displayed as + and negative charges are displayed as circles, where size is proportional to strength. pK_as were calculated with the Poisson-Boltzmann equation using the H++ server.

shown in Figure 5.5. Predictions vary dramatically depending on ε_{low} as shown by the dashed curves. The agreement is quite good at nearly all pH for the highest values of ε_{low} considered. Only data between pH 4 to 10 should be analyzed as β-Lg dissociates to a monomer from pH 4 to 3 [9].

The accuracy to which the pH-varying dipole moment of β-Lg was predicted in Figure 5.5 appears to be a significant step in the literature. To our knowledge, such agreement is not encountered elsewhere, for any molecule. South and Grant made predictions for the pH-varying dipole moments of horse and whale myoglobin from pH 5 to 8 [138]. In that work, the error between measurements and calculations could likely have been decreased

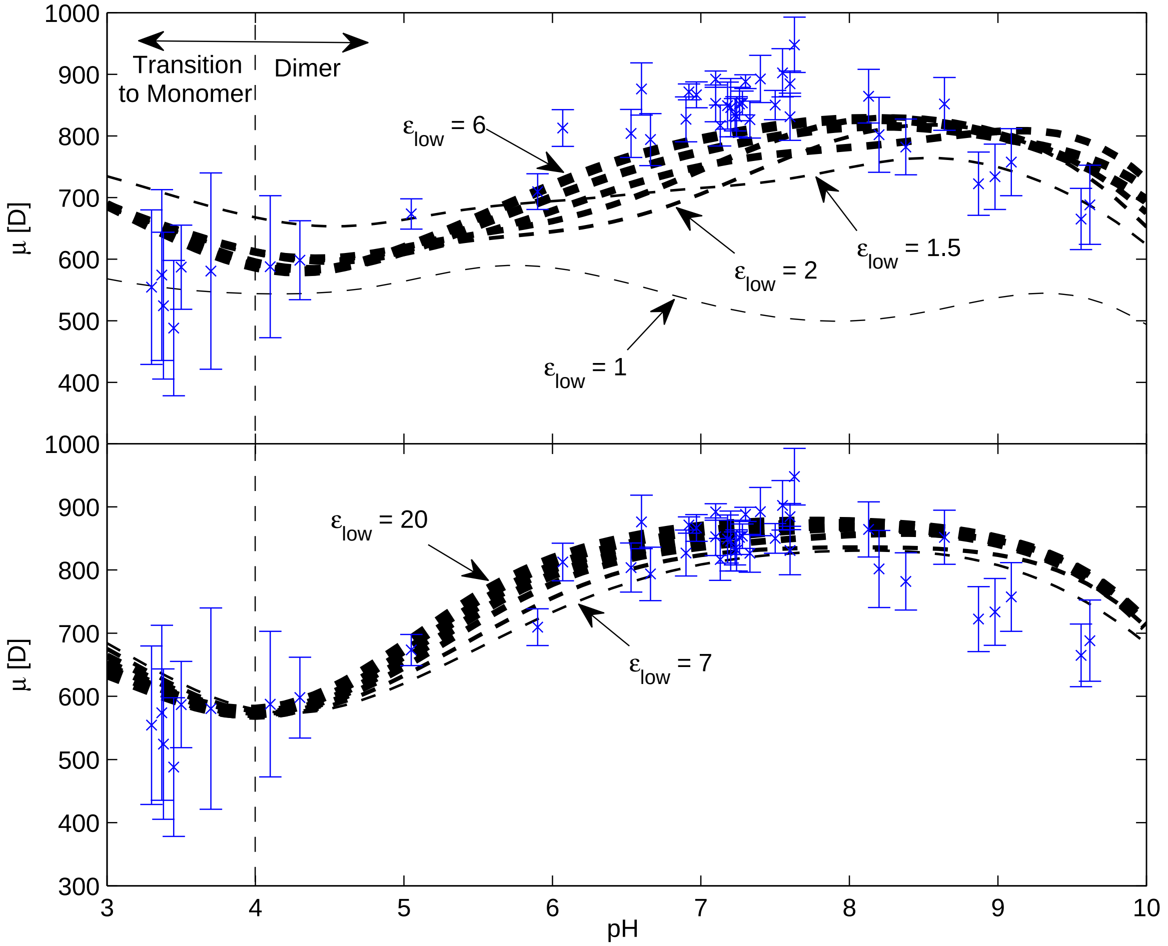


Figure 5.5: Measured and calculated dimer dipole moments μ of β-Lg at various pH for integral values of ε_{low} from 1 to 20 (and 1.5) using 1beb. Calculated μ are split into two plots (a) and (b) for easier viewing. Error bars represent the minimum and maximum values of several different least-squares fittings of the measured data. Between pH 3 and 4, β-Lg dissociates to a monomer [9]. Calculations appear to converge to the measurements as ε_{low} approaches 6 from below.

if pK_{a}s were calculated and the dielectric increment δ was fitted to measurements using a wider range of protein concentrations.

The apparent convergence of calculations and measurements in Figure 5.5 as ε_{low} approaches 6 both helps to validate the electrostatic model used and elucidates the dielectric properties of the protein interior. The curves representing $\varepsilon_{\text{low}} < 4$ are well separated from measured values, and indicate that the dielectric constant of the region must be higher.

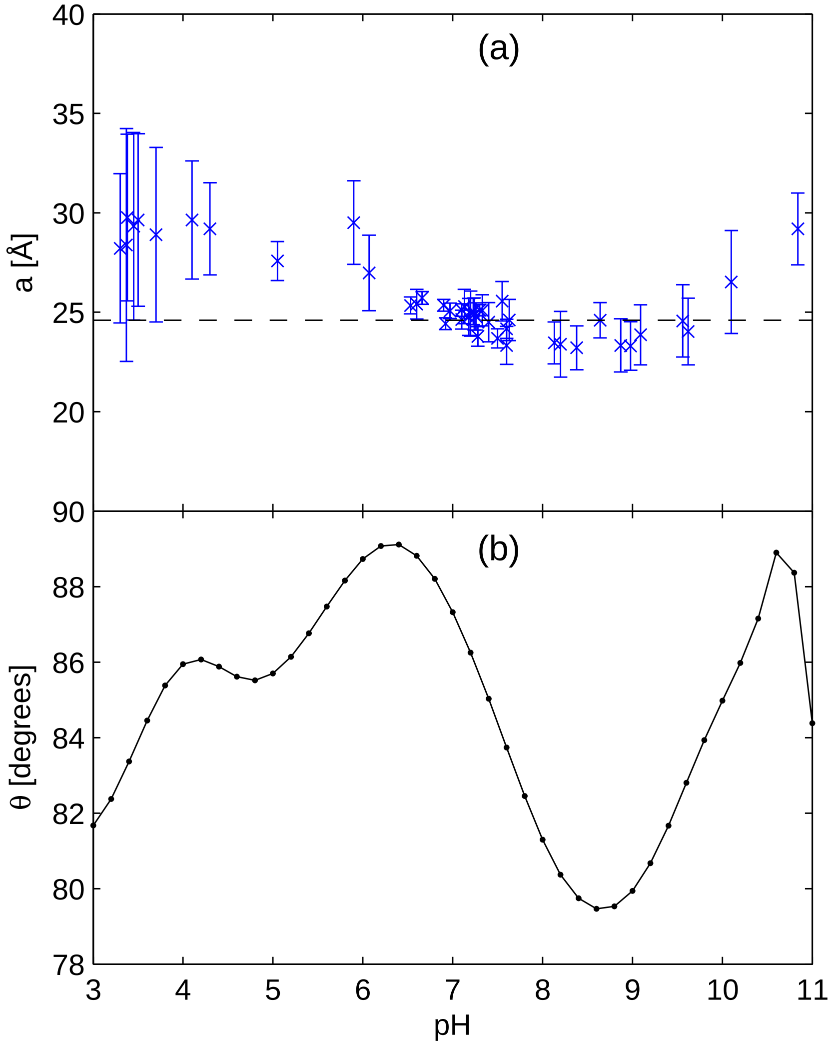


Figure 5.6: (a) Hydrodynamic radii *a* (Equation 4.5) of β-Lg at various pH. The error bars represent the minimum and maximum values of several different least-squares fits of the measured data. The horizontal dashed line represents the predicted radius of ∼24.6 Å using the equivalent volume sphere method described in Appendix B. (b) Calculated angle θ formed by the dipole moment vector and the long axis of β-Lg. pK_as were calculated from the H++ server. The dipole moment vector remains relatively aligned with the shorter axis at all pHs considered.

Using the equivalent volume sphere method described in Appendix B, the predicted radius of β-Lg from 1beb is ~24.6 Å, in good agreement with radii calculated from measurements. The β-Lg dimer is a prolate ellipsoid with an axial ratio of approximately 2. Thus two different relaxation times with a ratio of 1.3 between them, as predicted by Perrin's Equations, should exist [35]. The apparent radius of the molecule depends on the alignment of the dipole moment with each axis of rotation; e.g., if the dipole moment vector is parallel to the longer axis, the apparent radius would be larger than the apparent radius if it were aligned with the shorter axis. To determine if the observed fluctuations in radius were a result of the dipole moment vector changing alignment, the angle θ formed by the dipole moment vector and the long axis of β-Lg was calculated (Figure 5.6b). At all pHs, the dipole moment remains relatively aligned with the short axis (within 10°), and the changes in θ are unable to account for changes in a greater than 1 Å. Thus, the observed fluctuations in hydrodynamic radii in Figure 5.6a must additionally be influenced by a more complex process than changes in dipole alignment.

5.4 Estimation of the Internal Permittivity of Proteins

The electrical properties of proteins are dependent on the permittivity of the protein interior and the surrounding medium. High-permittivity environments reduce Coulombic forces. Low-permittivity environments where Coulombic forces are greater, such as the hydrophobic core of a protein, are critical to structure and function. The simplest continuum dielectric model approximates the permittivity of the protein, $\varepsilon(\mathbf{r})$, by treating the molecule as a low-permittivity medium with relative permittivity, ε_{low}, surrounded by a high-permittivity solvent with relative permittivity, $\varepsilon_{\text{high}}$, as described in the previous section. This model has been used to improve estimates of energies of interactions within [139, 140] and between proteins [141], electrons and protons [142, 143, 144], charged solutes [145], and ions moving through proteins [146, 147, 148, 149]. Mostly, the continuum dielectric model has been used to estimate titratable amino acid pK_{a} shifts based on local electric fields in proteins [132, 150], although comparison of these estimates with direct measurements with NMR have revealed examples where other factors are also involved [151, 152].

Unlike $\varepsilon_{\text{high}}$, which can and has been measured directly [3], ε_{low} has been found to be the most critical and least agreed-upon parameter in calculations of electrostatic effects in proteins [153, 124]. Though the majority of theoretical calculations and experimental determinations have placed the value of ε_{low} between 1 and 6 [154, 155, 156, 157, 158, 159, 160, 161, 162, 163], a number of contradictory studies have emerged. A theoretical study by Nakamura et al. using electronic polarization of atoms and the polarizations of local dipoles predicted internal dielectric constants from 1 to 20 [136]. Pennock and Schwan proposed a lower bound on ε_{low} of 2.5 based on series and parallel combinations of the dielectrics of water and protein powders [164, 165]. Antosiewicz et al., in calculating pK_{a} values for ionizable residues in proteins, were only able to predict satisfactory values using $\varepsilon_{\text{low}} \geq 20$ [166]. Demchuk and Wade additionally improved pK_{a} predictions using a hybrid technique where an ε_{low} close to aqueous solvent (~ 80) was used for solvent exposed residues and an ε_{low} in the range $10 - 20$ for buried residues [134]. García-Moreno et al. measured ε_{low} in terms of the permittivity that needed to be used in the Born equation to reproduce pK_{a} shifts measured by difference potentiometry [167]. Measured values were in the range of $10 - 12$. Similarly, Dwyer et al. measured the ΔpK_{a} of a single buried glutamic acid which reflected a dielectric constant of 12, interpreting this value as a small amount of solvent penetrating into the protein, which could contribute significantly to a higher value of ε_{low} [168]. Other interpretations of high dielectric constants include effects of conformational reorganization [169], fluctuations of charged side chains [170], and the reaction field of bulk water [171].

An estimate of ε_{low} can likewise be produced with dielectric spectroscopy measurements using the data from Figure 5.5. To quantitate the goodness of fit of the measurements to the Poisson-Boltzmann model, the RMS residual is used,

$$R_{\text{rms}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\mu}_{\text{calc}_i} - \boldsymbol{\mu}_{\text{meas}_i}\right)^2} \tag{5.16}$$

where N is the number of comparison points. The vector $\boldsymbol{\mu}_{\text{meas}}$ is obtained using a polynomial interpolation fitting of the measured data. R_{rms} is minimized to identify the value of ε_{low} used in the dielectric continuum model that best approximates the actual dielectric properties of the protein interior.

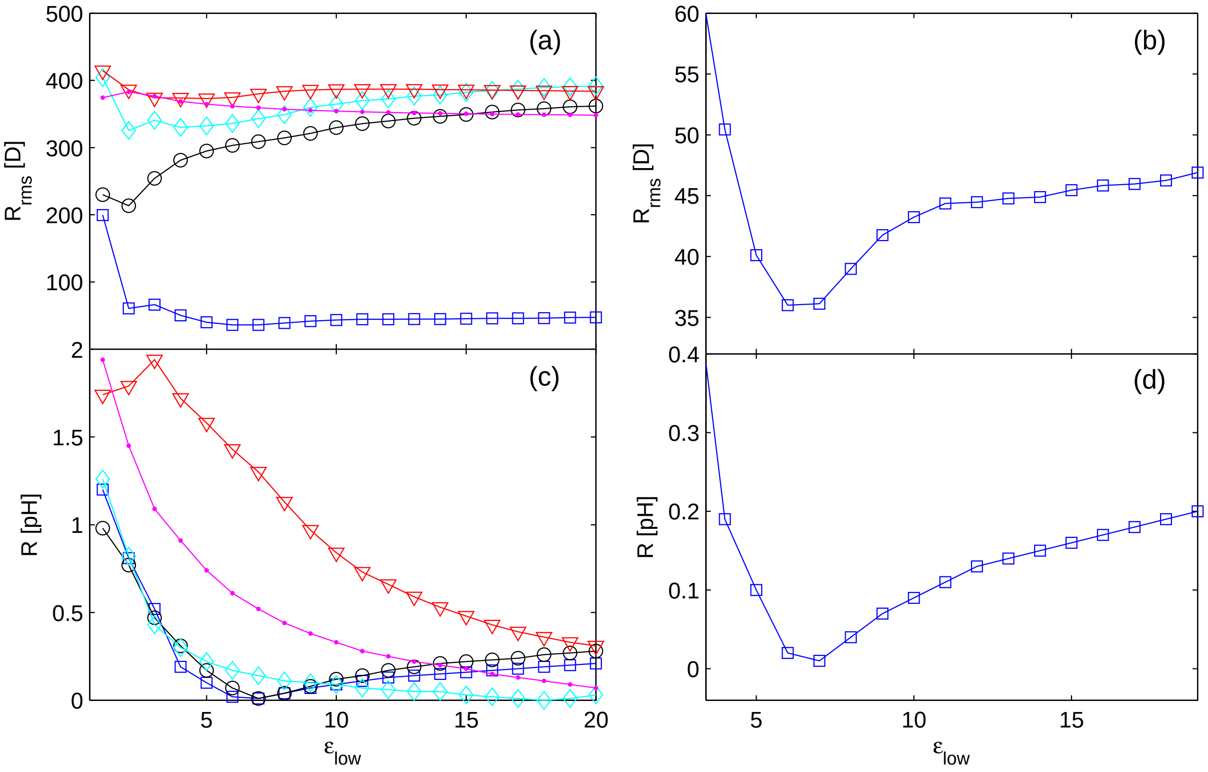


Figure 5.7: Dipole moment (a,b) and isoelectric point (c,d) residuals for 5 PDB models of β-Lg. PDB structures used are dimers(□) 1beb [5], (◯) 2akq [10], (◊) 2q39 [11], and monomers(▽) 1cj5 [7], and (·)1dv9 [9]. Data in (a) shows good compatibility of the measurements with 1beb only, evident by the small R_{rms} for that structure alone. (b) and (d) are expanded views of 1beb showing good agreement between minimum function values at approximately $\varepsilon_{\text{low}} = 6$ to 7 for both R_{rms} and R.

In addition to 1beb, two PDB models of dimeric β-Lg, 2akq and 2q39, and two monomeric structures, 1cj5 and 1dv9, are used for comparison. 2akq is the U' lattice grown from β-Lg crystals at very low ionic strength. This model exhibits the slight structural changes that occur in the almost complete absence of salt. 2q39 is a β-Lg model grown from crystals at low humidity and exhibits an asymmetric dimer. None of these four models were expected to resemble the protein's conformation in the experimental conditions, but were included to display the range of possible differences in charge moments of structural variations.

The RMS residuals (Equation 5.16) for Poisson-Boltzmann predictions for values of ε_{low} from 2 to 20 using the 5 PDB models of β-Lg in relation to the measured data are shown in Figures 5.7a and for 1beb alone in 5.7b. The agreement for 1beb with the measurements is quite remarkable for all $\varepsilon_{\text{low}} > 4$. An R_{rms} value of 35 D corresponds to an average error of 5%. Using the absolute minimum of the curve in Figure 5.7b, the best estimate of ε_{low} for β-Lg based on 1beb is between 6 and 7.

A similar estimation of ε_{low} is made by comparing the computed and measured isoelectric point of the protein, which was previously measured to be pH 5.1 [172]. The calculated value is the pH at which the overall charge Q from Equation 5.3 is 0. The residual value used here, R, is the difference between the measured and calculated pH values for the isoelectric point,

$$R = \left| pI_{\text{calc}} - pI_{\text{meas}} \right| . \tag{5.17}$$

Values of R for the 5 PDB models of β-Lg are plotted in Figures 5.7c and for 1beb alone in 5.7d. Unlike the dipole moment function, from which a confident rejection and selection of PDB models could be made, the isoelectric point residual, R, is more useful as a confirmation of the estimate for ε_{low}. This is because the total charge (at least for the dimeric structure) is not as unique as an array of pH-dependent dipole moments; many proteins share the same isoelectric point. Figures 5.7b and 5.7d show good agreement between the two methods for estimating ε_{low} for β-Lg.

5.4.1 Solvent Penetration as a Source of ε_{low}

Soon after the discovery of internal water molecules in solved crystal structures, it was suggested that water penetration could be a mechanism by which the effective dielectric constant of the protein could increase [173]. If indeed this is the case, the dielectric constant would be sensitive to breathing motions of the molecule, and the variations in hydrodynamic radius estimates with pH (Figure 5.6a) would argue against the assumption of a constant ε_{low} at all pH. Calculations using ε_{low} estimates would thus be restricted to a pH range where stable structural conditions are found. This may explain the greater deviations in Figure 5.5 at pH 8-10.

If ε_{low} were a function of solvent penetration, as was previously suggested [168], we could expect to see some correlation between a and ε_{low} for a given molecule. From

Figure 5.5, ε_{low} is best approximated by low values ($\varepsilon_{\text{low}} < 6$) at high pH ($\text{pH} > 8$) and high values ($\varepsilon_{\text{low}} > 6$) at physiological pH ($6 < \text{pH} < 8$), opposite to the expectation from the hydrodynamic radius a, which is large at high pH and small at physiological pH. Thus, the data appears to contradict the solvent penetration hypothesis. However, the radius a itself may not be strongly correlated with solvent penetration. Another factor such as the hydrophobicity of penetrable cavities may have a larger impact.

5.4.2 Implications for the Mechanism of the β-Dispersion

The β-dispersion in the dielectric spectrum of proteins in solution is the dispersion characterized by change in solution permittivity, $\Delta\varepsilon$, and the relaxation time, τ, typically occurring in the high kHz to MHz frequency range. Researchers have been measuring this dispersion for over 70 years. In the 1930s and 1940s, Oncley made dielectric measurements on several proteins, including β-Lg, and interpreted these results on the basis of the rotation of molecules having a permanent dipole moment [67, 69]. Since then, other theories have been proposed to account for this dispersion, such as proton fluctuation [174, 175], Maxwell-Wagner mechanism [176, 177], and the ion-mobility model [178]. The two theories that are still given attention today are permanent dipole rotation and proton fluctuation.

The proton fluctuation theory holds that when the pH of the solution is close to the $\text{p}K_{\text{a}}$ of any group, protons continually bind and disassociate, causing the dipole moment to fluctuate about an equilibrium. Because the change in solution permittivity $\Delta\varepsilon$ is proportional to μ^2, it would be possible to obtain a dispersion even when the average dipole moment is zero. The theory predicts that the dipole moment is a monotonically decreasing function of pH, dropping off from a maximum at pH 2 to effectively zero at pH 10 [174]. Figure 5.5 is in obvious contradiction to this. The proton fluctuation model was not accounted for explicitly in this work, neither did we consider Maxwell-Wagner and ion-mobility models. Nonetheless, the agreement in Figure 5.5 was achieved solely with the permanent dipole theory and Equation 4.1. This indicates that to account for $\Delta\varepsilon$ of β-Lg with 5% error, the permanent dipole model is sufficient.

5.5 Influence of pKa Shifts on the Calculated Dipole Moments of Proteins

Theoretical approaches to calculate charge moments numerically from PDB structures rely on a set of amino-acid partial charges. The proposed method in this chapter made this calculation by using calculated pK_as to replace intrinsic or model pK_as. Calculated pK_as come with computational costs [179], and the importance of pK_a shifts has been debated for dipole moment calculations [180]. More importantly, there has not been convincing agreement between experimental data and theory. Thus, the decision to include shifted pK_as in protein charge distribution evaluation has not been unanimous [181], and the effect of doing so is largely unknown.

The pK_a value is a measure of a titratable group's ability to enter a protonated state. Titratable groups can be partitioned into basic groups, which are positively charged in their protonated state, and acidic groups, which are neutral in their protonated state. Assuming the titration curve is Henderson-Hassalbalch-like [182], the pK_a of the titratable group completely describes its protonation state at a given pH. It follows that the set of pK_as that comprise the titratable groups of a protein determine its protonation state at a given pH and, in conjunction with the three-dimensional structure, its spatial charge distribution. Thus, a large amount of research has been invested by biophysicists into protein pK_a prediction, with an emphasis on quick and accurate structure-based predictions [132, 134].

Most approaches compute pK_a shifts using the linearized Poisson-Boltzmann equation [132, 183, 184, 185], as described earlier in this chapter. In this method, the permittivity of the protein environment is approximated in some way, and a pK_a shift is calculated from the difference in electrostatic energy of an amino acid residue in its charged and neutral state. This approach is regarded as accurate but slow, and has been criticized for ignoring the flexibility of the protein structure and simplifying permittivity environments [186].

Empirical approaches for protein pK_a prediction, such as the program PROPKA [187] and the technique by He et al. [188], have gained momentum recently due to very fast computation speeds. Generally, these methods predict pK_as by relating multiple physical characteristics of the titratable group's local environment to the characteristics of groups with experimentally measured pK_as. Remarkably, a recent evaluation of PROPKA alongside

Table 5.1: Model pK_{a} values for ionizable groups in proteins from the literature

Group	Nozaki and Tanford [194]	Thurkill et al. [1]	Cohn and Edsall [195]	Creighton [196]
Carboxyl	3.8	3.67	3.0-3.2	3.5-4.3
Asp	4.0	3.67	3.0-4.7	3.9-4.0
Glu	4.4	4.25	4.4	4.3-4.5
His	6.3	6.54	5.6-7.0	6.0-7.0
Amino	7.5	8.00	7.6-8.4	6.8-8.0
Cys	9.5	8.55	9.1-10.8	9.0-9.5
Tyr	9.6	9.84	9.8-10.4	10.0-10.3
Lys	10.4	10.40	9.4-10.6	10.4-11.1
Arg	12.0	—	11.6-12.6	12.0

three Poisson-Boltzmann approaches (MEAD [132], UHBD [189], MCCE [150]) reported that PROPKA outperformed the other methods in terms of accuracy and speed for all residues except Glu and His for a dataset of 100 proteins with experimentally determined pK_{a}s [190]. In view of this result and other developments, it would be worthwhile to compute dipole moments using empirically- and Poisson-Boltzmann-calculated pK_{a}s, alongside model pK_{a}s, and to report observable differences.

This study on the influence of pK_{a} shifts is motivated by the fact that various servers are now available for dipole moment computation of proteins in different solutions [191, 192, 181, 193]. The methods of these servers typically differ by the set of pK_{a}s used, yet there is little in the literature to validate which pK_{a}s yield dipole moments close to those measured experimentally. Clearly, a calculated pK_{a} which takes into account the titratable group's environment in the protein through computation should provide a better estimate of the actual pK_{a}. However, the extent to which dipole moment calculations are sensitive to these shifts is generally unknown. If the inclusion of shifted pK_{a}s is unnecessary, the computational expense of obtaining multiple pK_{a}s representing various solution conditions would be largely eliminated. The principle aim of this study is to determine how the variation between model and calculated pK_{a}s influence the dipole moment, and under what conditions, if any, can the simpler approach of using model pK_{a}s be used for the calculation of protein dipole moments.

5.5.1 Model pK_as

pK_a^ms of free amino acids are the simplest approximations of protein pK_as [195, 194, 196, 1]. They are defined as the pK_as of titratable groups in model compounds, which simulate an isolated, undisturbed condition in aqueous solution. These values represent amino acids in an unperturbed state, and are the basis for pK_a calculations in a perturbed state. A calculated pK_a, p$K_\mathrm{a}^\mathrm{calc}$, is the sum of a model pK_a and a calculated shift,

$$\mathrm{p}K_\mathrm{a}^\mathrm{calc} = \mathrm{p}K_\mathrm{a}^\mathrm{m} + \Delta\mathrm{p}K_\mathrm{a}^\mathrm{calc}. \tag{5.18}$$

Typically, pK_a^ms are determined using small molecules or peptides where electrical self-interactions are small and the solvent is fully accessible. Titration curves of the molecules are then fitted to Eqs. 5.11 and 5.12, where pK_a is the parameter being fitted. The sets of pK_a^ms used here are taken from the literature and shown in Table 5.1. Values from Table 5.1 that are expressed as ranges were assigned the midpoint of the range to extract a single value.

5.5.2 Empirically-Calculated pK_as

The increasingly popular program PROPKA, developed by Jensen and co-workers, predicts pK_as by empirically relating desolvation effects and intra-protein interactions to an ionizable group's position [187]. This method operates under the assumption that ionizable residues with similar surrounding features have similar pK_as. Recent benchmarks have reported that PROPKA performs very competitively with other methods, especially for Lys residues. Its weaknesses are predicting pK_as for Glu and His residues, as well as most buried residues [190]. PROPKA is freely available to the scientific community through an easy-to-use online server, and most proteins are processed in less than 5 seconds. The version used here was PROPKA 3.0.

5.5.3 Influence of pK_a Shifts on β-lactoglobulin

Measurements of the pH-dependent dipole moment of β-Lg are shown in Figure 5.8 alongside calculated dipole moments using various sets of pK_as. For β-Lg, predictions using

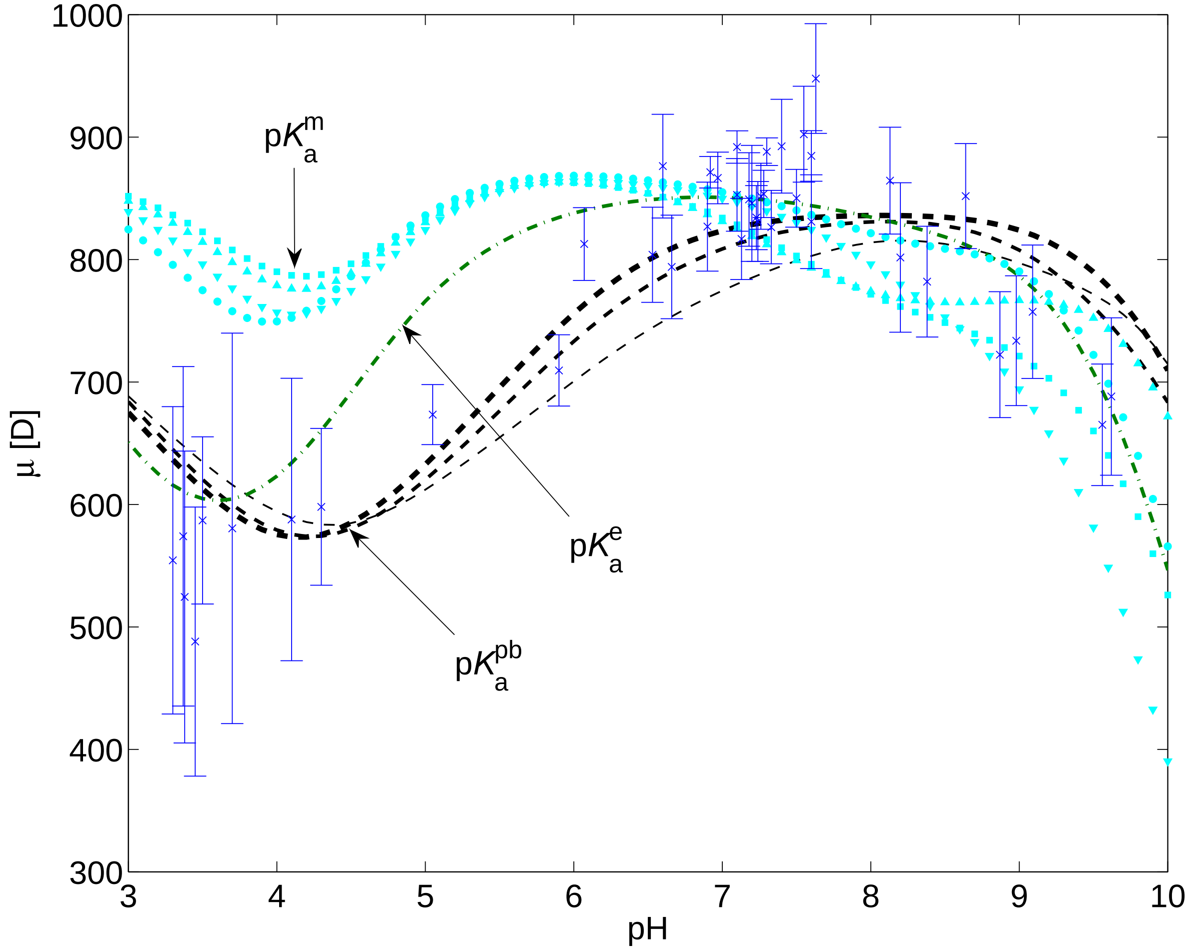


Figure 5.8: Measured and calculated dipole moments of β-Lg at various pH using $\mathrm{p}K_\mathrm{a}^\mathrm{m}$s, $\mathrm{p}K_\mathrm{a}^\mathrm{e}$s, and $\mathrm{p}K_\mathrm{a}^\mathrm{pb}$s and PDB model 1beb. Measurements are taken from the previous chapter and are represented with error bars. $\mathrm{p}K_\mathrm{a}^\mathrm{m}$s are denoted with filled cyan symbols △ (Nozaki and Tanford), □ (Thurkill), ◯ (Cohn and Edsall), and ▽ (Creighton). Black dashed lines represent $\mathrm{p}K_\mathrm{a}^\mathrm{pb}$s for $\varepsilon_{low} = 4$, 6, and 8, with increasing line thickness corresponding to increasing ε_low. The green dot-dashed line represents $\mathrm{p}K_\mathrm{a}^\mathrm{e}$.

$\mathrm{p}K_\mathrm{a}^\mathrm{m}$s overestimate μ at low pH, yet are consistent with measurements from pH 7 to 9. The $\mathrm{p}K_\mathrm{a}^\mathrm{m}$s from Creighton et al. have very large deviations at pH 10 caused by only slightly higher published values for Tyr and Lys residues. Predictions using $\mathrm{p}K_\mathrm{a}^\mathrm{e}$s reduce the error at low pH from $\mathrm{p}K_\mathrm{a}^\mathrm{m}$s, but still fail from pH 4.5 to 7. Predictions using $\mathrm{p}K_\mathrm{a}^\mathrm{pb}$s are within experimental errors at nearly all pH considered, and clearly best predict the pH-dependent dipole moment of β-Lg. The largest discrepancies between the various predictions occur near the isoelectric point (pI) of β-Lg, at pH 5.1 [172]. Here overall positive and negative

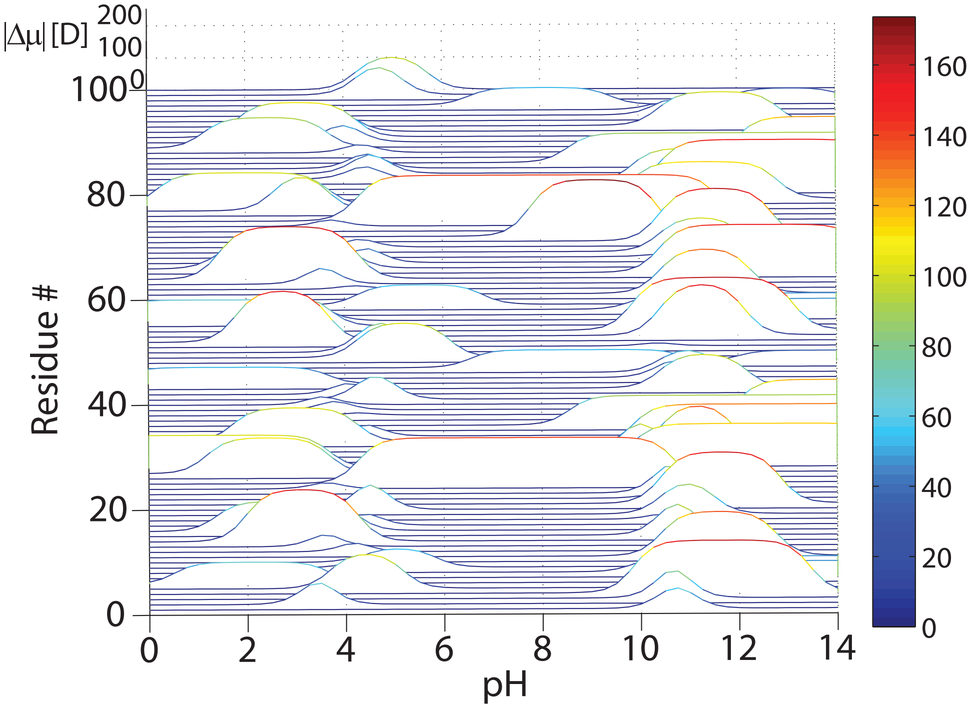


Figure 5.9: Magnitude difference in calculated dipole moments using pK_a^ms and pK_a^{pb}s, $|\Delta\mu|$, for ionizable residues in β-Lg. pK_a^ms used were from Thurkill et al. and pK_a^{pb}s were calculated with $\varepsilon_{\text{low}} = 6$. For each trace, the relative height corresponds to the $|\Delta\mu|$ as calibrated by the scale at the top of the plot, while the color reports the same information as calibrated by the color bar to the right. Residue # here corresponds to the sequence number of consecutive ionizable residues, 50 from each subunit and 100 total. Peak widths represent the range where there is disagreement between uncorrected model predictions and Poisson-Boltzmann corrected predictions for a given residue. Peak heights indicate the impact of that titratable group (and hence, the associated ambiguity it causes) on the total dipole moment. For most residues ($\sim$ 92%), $|\Delta\mu|$ is nonzero only outside the range pH 7 to 9.

charges are equal in magnitude, which may induce greater interactions among charged sites; yet, oddly, μ is near its minimum value.

The convergence at pH 7 of the predictions and measurements is interesting and merits discussion. In the case where a dipole moment at pH 7 were being computed, any set of pK_as would predict equally well. This raises the following question: Although calculations using pK_a^ms do not accurately predict measured dipole moments over all pH, is there a smaller range of pH where they do? At a given pH, the protonation state of a protein is only affected by pK_a shifts that occur within ~1.5 pK_a units of the pH. In Table 5.1, most model pK_as are outside the range of 7-9, with the exception of His and amino-terminus groups which are limited to one-per-protein monomer. Therefore, we expect different computed dipole moments to be very similar at pH 7-9 if the pK_a shifts are small or if Δp$K_\mathrm{a}^\mathrm{calc}$s are in the direction opposite those pHs.

To illustrate this, we define the difference in dipole moment between an ionizable residue with a shifted and non-shifted pK_a as

$$\Delta\mu = \mu'_\mathrm{m} - \mu'_\mathrm{pb} \tag{5.19}$$

where μ'_m and μ'_pb are dipole moments of a single ionizable residue calculated with pK_a^ms and pK_a^pbs, respectively. $|\Delta\mu|$ for 100 ionizable residues of β-Lg are shown in Figure 5.9. For most residues, the pH where $|\Delta\mu|$ is nonzero is confined to outside the range 7 to 9, suggesting that differences in calculated dipole moments will be minimized within that range as confirmed in Figure 5.8. Generally, Δp$K_\mathrm{a}^\mathrm{calc}$s for Tyr (p$K_\mathrm{a}^\mathrm{m}$ = 9.84) and Lys (pK_a^m = 10.40) were positive and Δp$K_\mathrm{a}^\mathrm{calc}$s for Glu (p$K_\mathrm{a}^\mathrm{m}$ = 4.25) were negative, thus, the penalty for not using pK_a^pbs is larger at low and high pH. Also noteworthy are the large contributions of individual residues to the total dipole moment, many of which were greater than 120 D. Because the total dipole moment is ~800 D, a large amount of vector cancellation must occur due to oppositely-situated residues in the molecule.

5.5.4 Influence of pK_a Shifts on Protein Dataset

To address the pK_a^m approximation question raised in the previous section with a larger sample size, dipole moment calculations using the three types of pK_as were made on a protein dataset consisting of 66 protein models from the PDB (available in Appendix C). The proteins selected had molecular weights in the range of 5 to 50 kDa and roughly globular shape. The mean and standard deviation of the weights were 16.5 kDa and 9.7 kDa, respectively. PDB files were imported into MATLAB where all computations were performed. Six proteins with calculated charges and dipole moments with are shown in Figure 5.10.

Because the present objective is similar to the problem of estimating one parameter from another, two statistical measures from estimation theory were used in the analysis. The

Table 5.2: *B* (top, italicized) and *D* (bottom) between calculated dipole moments of proteins using pK_a^ms and pK_a^es relative to pK_a^{pb}s

		pH 4-10		
		$\varepsilon_{\text{low}} = 4$	$\varepsilon_{\text{low}} = 6$	$\varepsilon_{\text{low}} = 8$
pK_a^es		*13.9*	*11.1*	*10.0*
		18.1%	15.3%	13.9%
pK_a^ms	Nozaki and Tanford	*14.2*	*11.3*	*10.2*
		23.2%	21.6%	21.2%
	Thurkill et al.	*18.2*	*15.3*	*14.2*
		23.8%	21.6%	20.8%
	Cohn and Edsall	*11.8*	*8.9*	*7.9*
		21.7%	19.5%	18.8%
	Creighton	*15.5*	*12.6*	*11.5*
		24.0%	22.4%	22.0%
		Neutral pH (pH 7)		
		$\varepsilon_{\text{low}} = 4$	$\varepsilon_{\text{low}} = 6$	$\varepsilon_{\text{low}} = 8$
pK_a^es		*8.9*	*6.5*	*5.5*
		12.7%	11.0%	10.0%
pK_a^ms	Nozaki and Tanford	*10.8*	*8.5*	*7.5*
		13.5%	12.9%	11.9%
	Thurkill et al.	*11.2*	*8.9*	*7.8*
		12.9%	10.7%	9.5%
	Cohn and Edsall	*12.5*	*10.2*	*9.2*
		12.7%	10.6%	9.4%
	Creighton	*9.3*	*7.0*	*6.0*
		14.0%	13.6%	12.6%
		Isoelectric point (pI)		
		$\varepsilon_{\text{low}} = 4$	$\varepsilon_{\text{low}} = 6$	$\varepsilon_{\text{low}} = 8$
pK_a^es		*15.9*	*14.6*	*16.1*
		24.0%	21.4%	21.1%
pK_a^ms	Nozaki and Tanford	*32.9*	*31.7*	*33.2*
		30.3%	29.4%	32.6%
	Thurkill et al.	*36.7*	*35.5*	*37.0*
		32.9%	32.1%	32.6%
	Cohn and Edsall	*28.0*	*26.8*	*28.3*
		29.4%	29.0%	29.4%
	Creighton	*36.1*	*34.9*	*36.4*
		32.4%	31.5%	31.9%

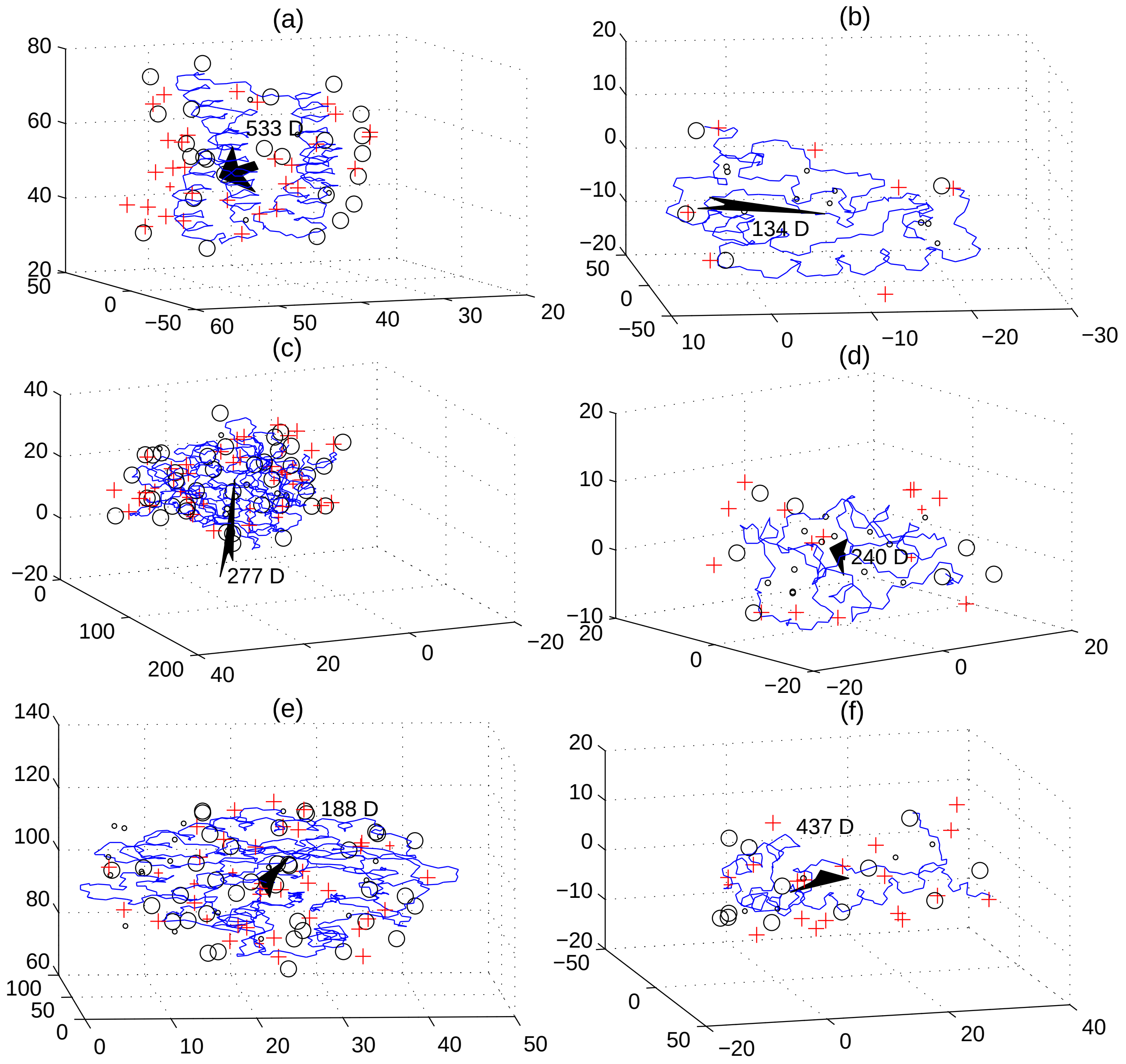


Figure 5.10: Calculated partial charges and dipole moment for six proteins from the 66 protein dataset: (a) 1gs9 [12], (b) 1beo [13], (c) 1cb4 [14], (d) 1gdc [15], (e) 1gym [16], and (f) 1hme [17]. Dimensions are in Å.

first measure is the estimation bias, B, defined as

$$B = \mathrm{E}\left[\mu_{\mathrm{m}} - \mu_{\mathrm{pb}}\right] \tag{5.20}$$

where μ_{m} is the dipole moment calculated using p$K_{\mathrm{a}}^{\mathrm{m}}$s, μ_{pb} is that calculated using p$K_{\mathrm{a}}^{\mathrm{pb}}$s, and the operator $\mathrm{E}[\cdot]$ is the expected value. The second measure is the average percentage difference, D, given by

$$D = \mathrm{E}\left[\frac{|\mu_{\mathrm{m}} - \mu_{\mathrm{pb}}|}{\mu_{\mathrm{pb}}}\right]. \tag{5.21}$$

Thus B is expressed in Debye and indicates whether using μ_{m} as an estimator for μ_{pb} is biased on average, and D is expressed as a percentage and indicates the average percentage error of using μ_{m} to estimate μ_{pb}.

The assumption we rely on by using parameters B and D is that dipole moments calculated with p$K_{\mathrm{a}}^{\mathrm{pb}}$s are much more representative of actual dipole moments than those calculated using p$K_{\mathrm{a}}^{\mathrm{m}}$s. This assumption is primarily based on the substantial volume of work regarding Poisson-Boltzmann pK_{a} prediction in the literature, as well as the additional experimental result for β-Lg discussed in the previous sections. The ideal scenario would be to use experimentally measured dipole moments as the reference; however, very few published measurements of dipole moments are available, especially at different pH.

Calculated values of B and D are shown in Table 5.2 for pH 4-10, pH 7, and the isoelectric point pI, and for $\varepsilon_{\mathrm{low}} = 4$, 6, and 8. Calculations over pH 4-10 were performed in 0.2 pH increments and then averaged over all 31 data points. Isoelectric points were simulated using the method described here [113]. $\varepsilon_{\mathrm{low}}$ was varied because the actual permittivites are unknown, however the range 4 to 8 covers several estimates. Note that values of B and D were also calculated between p$K_{\mathrm{a}}^{\mathrm{e}}$ and p$K_{\mathrm{a}}^{\mathrm{pb}}$ predicted dipole moments.

Values of the estimation bias B were positive in all cases, indicating that p$K_{\mathrm{a}}^{\mathrm{m}}$s consistently predict dipole moments that are too large. This lends itself to a simple correctional procedure when using p$K_{\mathrm{a}}^{\mathrm{m}}$s for molecules between 5-50 kDa: scale dipole moments calculated at pH 7 by 97% and scale dipole moments calculated at pI by 87%. Obviously, this will have a greater effect at pI, where differences between μ_{m}s and μ_{pb}s are much greater but, nonetheless, will yield smaller values of D (prediction error) in all cases when this correction is applied.

Values of the average percentage difference D were smaller at higher ε_{low} and were notably different in each pH range. The correlation with ε_{low} is understood in the context of electrostatic shielding. As ε_{low} becomes large, titratable groups become more electrically isolated and less prone to system perturbations (Coulombic forces between charges decrease with $1/\varepsilon$). At this limit, $\text{p}K_{\text{a}}^{\text{pb}}$s should converge to $\text{p}K_{\text{a}}^{\text{m}}$s, which represent the titratable groups in an unperturbed state. Thus, the behavior of D with ε_{low} is expected, and implies that proteins with larger values of ε_{low}, i.e., smaller molecular weights and/or larger surface area to molecular weight ratios (due to solvent penetration [168]), will be better approximated by $\text{p}K_{\text{a}}^{\text{m}}$s. Indeed, for proteins with weights below 30 kDa, D had a small positive correlation with molecular weight (not shown). The disparity of D with pH in the protein dataset parallels the trend for β-Lg: D smallest at pH 7 and largest at pI. Values of D were remarkably high at pI, and many of the largest proteins considered had values close to 40%. For pI even dipole moments computed using $\text{p}K_{\text{a}}^{\text{e}}$s showed large differences from $\text{p}K_{\text{a}}^{\text{pb}}$s. Nevertheless the improvement when using $\text{p}K_{\text{a}}^{\text{e}}$s in place of $\text{p}K_{\text{a}}^{\text{m}}$s is noticeable at pH 4-10 and pI, even though at pH 7 the difference is negligible.

Although the data in Table 5.2 opposes the practice of using $\text{p}K_{\text{a}}^{\text{m}}$s for high precision dipole moment calculations, it does indicate situations where the penalty for not using $\text{p}K_{\text{a}}^{\text{calc}}$s is less. Under the condition of neutral pH, calculated dipole moments using $\text{p}K_{\text{a}}^{\text{m}}$s are within 15% of those calculated with the more complex Poisson-Boltzmann approach. The dipole moment server by Felder et al. [181], which assumes a protonation state corresponding to pH 7, probably falls in this error range for most of their calculations. This is probably acceptable given the purpose of their server, which is to find proteins with unusually large dipole moments that interact with nucleic acids under physiological conditions.

5.5.5 Sensitivity of Calculated Dipole Moment to Internal Permittivity

Poisson-Boltzmann $\text{p}K_{\text{a}}$ calculations are historically plagued by the difficulty in selecting ε_{low}. Because it has been previously demonstrated that protein electrostatics in general are highly susceptible to changes in $\text{p}K_{\text{a}}$ values [197], we probed the sensitivity of protein dipole moments to $\text{p}K_{\text{a}}$ shifts by varying the internal permittivity. Calculations of dipole moments using $\text{p}K_{\text{a}}^{\text{pb}}$s were performed as ε_{low} was varied over a reasonable range from 1 to 6

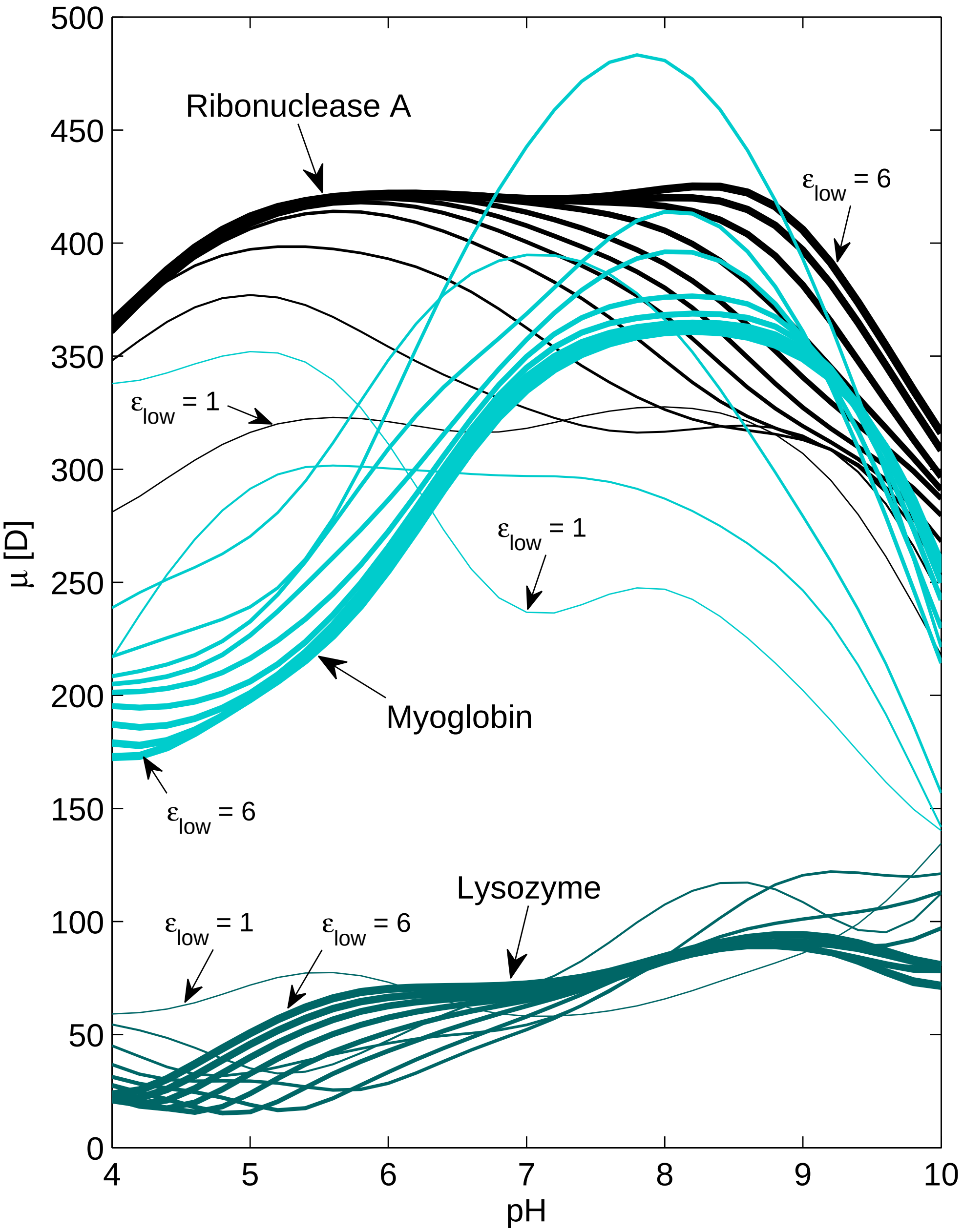


Figure 5.11: Dipole moments of ribonuclease-A (5rsa), myoglobin (1myf), and lysozyme (1lyz) over variable pH and internal permittivity. The three proteins are represented by three colors with line thickness indicating the internal dielectric used in the calculation ranging from 1 to 6 at 1/2 increments (thinner lines corresponding to lower permittivities). Sensitivity to internal permittivity appears to be greater at lower values, particularly $\varepsilon_{low} < 3$ for all three proteins.

Table 5.3: Shifted pK_as and dipole moment contributions for ionizable sites of Lys and Tyr in RNase A at pH 9.6.

			$\varepsilon_{low} = 2$		$\varepsilon_{low} = 4$		$\varepsilon_{low} = 6$	
site	pK_a^m	μ	ΔpK_a^{calc}s	$\Delta\mu$	ΔpK_a^{calc}	$\Delta\mu$	ΔpK_a^{calc}	$\Delta\mu$
Lys1	10.4	77.8	0.16	3.39	0.18	3.90	0.19	3.92
Lys7	10.4	46.2	-2.17	-44.03	-1.47	-36.77	-1.23	-31.63
Tyr25	9.8	16.3	> 12.0	-16.30	> 12.0	-16.30	> 12.0	-16.30
Lys31	10.4	78.5	-2.36	-76.05	-1.48	-62.90	-1.09	-47.52
Lys37	10.4	73.1	-3.31	-72.84	-1.41	-56.36	-0.85	-33.36
Lys41	10.4	26.1	-5.07	-26.10	-2.76	-25.77	-0.17	-1.65
Lys61	10.4	85.1	-1.76	-75.32	-0.66	-27.78	-0.10	-3.03
Lys66	10.4	48.6	2.86	7.68	1.99	7.60	1.75	7.53
Tyr73	9.8	32.4	9.54	-32.40	5.34	-32.40	4.10	-32.40
Tyr76	9.8	37.1	2.83	-37.01	1.45	-34.86	0.84	-28.53
Lys91	10.4	112.6	0.07	2.44	0.07	2.22	0.07	2.16
Tyr92	9.8	36.6	7.81	-36.60	4.35	-36.60	3.17	-36.56
Tyr97	9.8	22.1	> 12.0	-22.10	> 12.0	-22.10	9.78	-22.10
Lys98	10.4	78.2	-0.41	-13.71	-0.31	-9.65	-0.27	-8.15
Lys104	10.4	56.8	-0.11	-2.18	0.51	5.95	0.64	6.83
Tyr115	9.8	31.2	> 12.0	-31.20	7.62	-31.20	5.14	-31.20

at 1/2 unit increments for the commonly-studied proteins ribonuclease-A (RNase A) (5rsa [198]), myoglobin (1myf [199]), and lysozyme (1lyz [200]). These calculations are shown in Figure 5.11. At lower permittivities, the dipole moments for the 3 proteins appear to be highly erratic, especially for curves corresponding to $\varepsilon_{\text{low}} = 1$. The behavior of myoglobin at $\varepsilon_{\text{low}} = 2.5$ is peculiar in comparison to other values of ε_{low}. At that permittivity, the charge configuration arranges in such a way to predict a dipole moment well beyond 400 D at physiological pH, a value much larger than experimental values (around 200 D [138]). As the dielectric constant approaches 6, the changes in dipole moment across all pH decrease and the curves appear to converge.

The dipole moment of RNase A at its isoelectric point is further decomposed in Table 5.3 where pK_a shifts and dipole moment contributions of all Lys and Tyr residues are shown for three permittivities of the protein interior. At $\varepsilon_{\text{low}} = 6$, 9 out of 16 residues had changes in dipole moment greater than 20 D from μ calculated using pK_a^ms, while six of those had changes greater than 30 D. The results for Lys31 and Lys37 provide a marked example of dipole moment sensitivity to small pK_a shifts. For Lys31, the pK_a shifted down from 10.4 to 9.31, a drop of 1.09, which yielded a decrease in dipole moment of almost 50 D for that residue.

The sensitivity of the dipole moment to Lys residues in RNase A is attributed to two factors. First, the isoelectric point, the steep part of the protein titration curve, is near the pK_as of Lys residues, so one expects large changes in Lys titration state for small changes in pH in that region. The second factor is that Lys residues tend to be near the surface [198]. Charged surface residues, due to their large distance from the center of mass of the protein, have a larger effect on the overall dipole moment than charged buried residues for an equivalent charge. Quantitatively, the dipole moment of a charge increases linearly with the distance $\mathbf{r}_i - \mathbf{r}_r$ (Equation 5.4).

Individual dipole moment contributions for Tyr residues also exhibited large changes between pK_a^pb and pK_a^m, whose 6 residues were on average 27.8 D lower when using pK_a^pbs. Like Lys, the pK_a^m of Tyr is close to the isoelectric point of RNase A, and due to electrical interactions in the folded protein, the pK_as of Tyr residues were all raised and all but Tyr76 became uncharged. As a result, the pK_a^m values grossly overestimated the actual charge accrued on Tyr residues and the overall dipole moment of RNase A was too large.

It is evident that the calculated protein dipole moment has a strong dependence on the choice of ε_low, particularly at values close to 1. Interestingly, a few investigators have suggested an ε_low of 1 be used in the Poisson-Boltzmann approach [156]. Low values of ε_low assume strong Coulombic forces in the interior of the molecule, and the danger of an erroneously low ε_low as observed in Figure 5.11 is that the error could surpass the error using pK_a^ms. Even at pH 7, lower ε_low values yield strongly varying dipole moments. It will become increasingly important to develop accurate permittivity models of the protein to obtain exact charge distributions and their related pH-dependent dipole moments.

Chapter 6

The Role of Molecular Dynamics Simulations in Dielectric Studies

Computer simulations of liquids provide a powerful tool for analyzing complex systems consisting of large numbers of particles. They provide a direct connection between microscopic details and experimentally obtainable quantities. The assumption underlying such simulations is that the molecular behavior and interactions in a system can be described sufficiently by empirical force-fields derived from crystalographic structure analysis and spectroscopic analysis of bond vibration frequencies, as well as the fundamental laws of electrostatics, quantum mechanics, and statistical mechanics. The computer in essence becomes a virtual laboratory where the conditions of a measurement are reproduced and an experimental protocol is carried out. When a simulation model consistently agrees with experiment results, an understanding of the physical underpinnings of the measured phenomenon can be achieved.

The usefulness and applicability of molecular dynamics simulations has been richly demonstrated in numerous books [201, 202, 203, 204, 205, 206, 207] and reviews [208, 209, 210]. Most simulations are directed toward one of two goals: (1) to reproduce an experimental result that can be observed directly, either to reveal its physical origins or to validate the result itself or (2) to discover a molecular phenomenon that cannot be observed directly or is too costly to carry out experimentally [211]. Molecular dynamics studies are ideally suited for applications where analytical solutions are not possible or more detail is required, which generally is true when any nontrivial molecular interaction takes place.

The use of molecular dynamics simulations to study the dielectric properties of protein solutions is not novel [212, 213]. The typical approach involves the calculation of the dielectric susceptibility through Fourier transforms of the fluctuating dipole moment in the absence of an applied field [214]. Although there has been some experimental agreement using this approach [215], it is unsettling that in several cases simulations give qualitatively wrong results for static measured permittivities [171, 216, 217]. In this chapter, an alternate approach to calculate permittivity will be introduced where an applied electric field is placed across a protein in a solvent box.

This chapter primarily focuses on the prediction of protein dielectric relaxation using molecular simulations and the utility of refined PDB models for dielectric studies. The simulation program CHARMM (Chemistry at HARvard Macromolecular Mechanics), a standard molecular dynamics program, will be used for its popularity and availability on the BYU supercomputer. First, an exposition on the CHARMM methodology will be given. This is followed by an analysis of protein permittivity using the applied-field fluctuating dipole moment approach. A new method to calculate permittivity through simulations with applied electric fields is then introduced. Finally, the effects of refining PDB models for use in the calculation methods of Chapter 5 is discussed.

6.1 CHARMM Force Fields on the BYU Supercomputer

CHARMM is a widely-used simulation program for minimization and dynamics of macromolecules [218]. It was originally developed at Harvard University as a tool for investigating complex macromolecular systems with importance in biology [219]. It has since been used successfully to model molecular dynamics simulations of solvated proteins, protein-DNA complexes and lipid systems, addressing a variety of issues including the thermodynamics of ligand binding and the folding of small proteins.

The Fulton Supercomputing Lab (FSL) at Brigham Young University provides computational resources for students and faculty. FSL currently houses 933 servers (nodes) comprising 9,880 processor cores. Most of the processors are supported with 24 GB of RAM. FSL also houses 200 terabytes of high performance storage. Because molecular simulations can require very large amounts of computational power, CHARMM was written with the capability to run on more than one processor at a time. All simulations that were performed in this dissertation were run on 32 processors using CHARMM version 35.

6.1.1 Overview of the CHARMM Program

The CHARMM program is based on sets of force fields for each atom in a system [218]. This means that fixed point charges are assigned to each atom location. These charges are due to (1) unequal sharing of electrons in covalent bonds and (2) titratable amino acid groups. At each time step, a potential energy function is computed based on energies from bonds,

$$U_b = \sum_{\text{bonds}} K_b \left(b - b_0\right)^2 \tag{6.1}$$

angles,

$$U_\theta = \sum_{\text{angles}} K_\theta \left(\theta - \theta_0\right)^2 \tag{6.2}$$

dihedrals,

$$U_\phi = \sum_{\text{dihedrals}} K_\phi \left(1 + \cos\left(n\phi_0 - \phi\right)\right) \tag{6.3}$$

improper dihedrals,

$$U_\omega = \sum_{\text{impropers}} K_\omega \left(\omega - \omega_0\right)^2 \tag{6.4}$$

Van der Waals forces,

$$U_{VW} = \sum_{\text{non-bonded pairs}} \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] \tag{6.5}$$

and electrostatic forces

$$U_q = \sum_{\text{non-bonded pairs}} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}. \tag{6.6}$$

The total potential energy function, $U(\mathbf{R})$, is the sum of each of these and is given by

$$\begin{aligned} U(\mathbf{R}) &= U_b + U_\theta + U_\phi + U_\omega + U_{VW} + U_q & (6.7) \\ &= \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi (1 + \cos(n\phi - \delta)) + \\ & \sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2 + \sum_{\text{non-bonded pairs}} \varepsilon_{ij}^{\text{min}} \left[\left(\frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^{6} \right] + \\ & \sum_{\text{non-bonded pairs}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}. & (6.8) \end{aligned}$$

The parameters b_0, θ_0, ϕ_0, and ω_0 are equilibrium values and K_b, K_θ, K_ϕ, and K_ω are force constants. For the dihedral energy, U_ϕ, n is the periodicity of the dihedral angle and ϕ is the phase shift. Coulombic interactions are determined between all point charges q_i and q_j where permittivity ε_r is equal to 1 when explicit solvent is used. Implicit solvent models, where ε_r is approximated, can be cost-efficient and sufficiently accurate for a variety of problems.

During a simulation run, the system is generally set by user input to a constant temperature. The temperature is maintained by controlling the average velocity of the atoms. The simulation proceeds in small time steps from 1 to 2 fs where at each time step the potential energy function $U(\mathbf{R})$ and a corresponding force $\mathbf{F}(\mathbf{R}) = -\nabla U(\mathbf{R})$ are calculated where $\mathbf{R}$ is the three-dimensional position vector. The atoms are then moved as dictated by the force $\mathbf{F}(\mathbf{R})$, the temperature control, and other external forces set by the user. Simulation run times generally range anywhere from 1 ps to 500 ns, depending on the application and computation power.

6.1.2 CHARMM Data Structures

Data structures in CHARMM allow the program to strike a balance between user accessibility and program versatility. CHARMM uses data structures to interface the user with new methods and approaches. What follows is a brief description of six important data structures and their role in the overall program flow:

- Residue Topology File -

  The residue topology file contains the information necessary to describe bond, angle, dihedral angle and improper dihedral angle content, as well as charge distribution, hydrogen-bond donors and acceptors and internal coordinate information for polymer components and solvents. Standard residue topology files for nucleic acids, lipids, proteins, and carbohydrates can be obtained from the developer's homepage.

- Parameter File -

  The CHARMM parameter file contains the information needed to calculate energies of structures. It defines all equilibrium structural values and force constants from Equation 6.8.

- Coordinate File-

  Coordinate files contain Cartesian coordinates of all atoms in a system. They are either written out in PDB or CHARMM ASCII format.

- Protein Structure File -

  The protein structure file lists every bond, bond angle, dihedral angle, and improper dihedral angle in addition to information needed to generate the hydrogen bonds and the non-bonded list for the molecular system. The protein structure file, in conjunction with the parameter and coordinate files, are used to calculate the energy of a system.

- Non-Bonded list -

  The non-bonded list contains atoms that are not covalently coupled within one or two bonds, but which are subject to van der Waals and electrostatic interactions. Because the length of this list can approach the number of atoms squared, a number of approximation schemes can be used to reduce its size.

- Constraints -

  Constraints are used to influence the system in some manner during simulation. Some common examples are constraints to hold the center of mass in place during dynamics, to fix bond lengths during dynamics, and to apply an external force during dynamics. Multiple constraints may be applied during a single simulation run.

6.2 Permittivity Calculations through Fourier Analysis

This section gives a brief summary of the current theory to calculate the permittivity of a solute-solvent system using Fourier analysis of the fluctuating system dipole moment under no external constraints. The original papers can be referenced for a more detailed exposition on the subject [220, 221].

6.2.1 Theory

The frequency-dependent electric susceptibility, $\chi_e(\omega)$, is always related to the permittivity through the simple equation

$$\chi_e(\omega) = \varepsilon_r(\omega) - 1. \tag{6.9}$$

Using linear response theory, the electric susceptibility can be related to the time correlation function of the system dipole moment, $\Phi(t)$, by

$$\chi_e(\omega) = \frac{1}{3Vk_bT}\mathcal{F}\left\{-\frac{d\Phi(t)}{dt}\right\} \tag{6.10}$$

where $\Phi(t) = \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle$, V is the volume, and $\mathbf{M}(t)$ is the system dipole moment. $\mathcal{F}\{\cdot\}$ is the Fourier transform and is defined

$$\mathcal{F}\{f(t)\} = \int_0^\infty f(t)\,\mathrm{e}^{j\omega t}dt. \tag{6.11}$$

Because $\varepsilon_r(\omega)$ is complex, $\chi_e(\omega)$ is also complex, i.e.

$$\chi_e(\omega) = \chi_e^{'}(\omega) + j\chi_e^{''}(\omega). \tag{6.12}$$

Exploiting the following property of the Fourier transform,

$$\mathcal{F}\left\{-\frac{df(t)}{dt}\right\} = f(t) + j\omega\mathcal{F}\{f(t)\}, \tag{6.13}$$

the real and imaginary parts of $\chi_e(\omega)$ can be expressed as

$$\chi_e^{'}(\omega) = \chi_e(0) - \frac{\omega}{3Vk_bT}\mathrm{imag}\left[\mathcal{F}\{\Phi(t)\}\right] \tag{6.14}$$

and

$$\chi_e^{''}(\omega) = \frac{\omega}{3Vk_bT}\mathrm{real}\left[\mathcal{F}\{\Phi(t)\}\right]. \tag{6.15}$$

where $\chi_e(0)$ is the static electric susceptibility. For $\omega = 0$, Equation 6.10 reduces to

$$\chi_e(0) = \frac{\Phi(0)}{3Vk_bT} \tag{6.16}$$

$$= \frac{\langle \mathbf{M}(0) \cdot \mathbf{M}(0) \rangle}{3Vk_bT} \tag{6.17}$$

$$= \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3Vk_bT} \tag{6.18}$$

$$\approx \frac{\langle \mathbf{M}^2 \rangle}{3Vk_bT}. \tag{6.19}$$

Due to random movements of the system dipole moment in the absence of an applied field, for sufficiently long simulations $\langle \mathbf{M} \rangle^2$ goes to zero and $\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2 \approx \langle \mathbf{M}^2 \rangle$.

The Fourier analysis method for determining permittivity is easily tailored to isolate components of a system. For example, to determine the permittivity contribution of the solvent, $\mathbf{M}$ from Equations 6.18 and 6.19 is replaced with $\mathbf{M}_{\mathrm{W}}$, the system dipole moment of the water. Similarly, to determine the protein's contribution, $\mathbf{M}$ is replaced with $\mathbf{M}_{\mathrm{P}}$, the system dipole moment of the protein. The Fourier analysis method has also been used to determine protein-solvent interactions. Water molecules at the protein interface behave differently than those in bulk solvent. These water molecules contribute to the so-called δ-relaxation, a relaxation process composed of two or three smaller relaxations at frequencies between those of the protein and bulk water [47]. Using a 5 ns molecular dynamics run, Boresch et al. reported a bimodal δ-relaxation process consistent with experimental data [212].

6.2.2 Simulated Dielectric Susceptibility of Protein-Water System

β-Lg was solvated in a water cube with 10 nm side lengths and simulated for 6 ns at 25 °C with no external forces. Every 200 fs the dipole moment magnitude and orientation were recorded for each component in the system (total, protein, and solvent). The autocorrelation for the components are shown in Figures 6.1 and 6.2. The goal is to fit the total autocorrelation function to the curve

$$\Phi_{\mathrm{fit}}(t) = A_1 \mathrm{e}^{-t/\tau_1} + A_2 \mathrm{e}^{-t/\tau_2}, \tag{6.20}$$

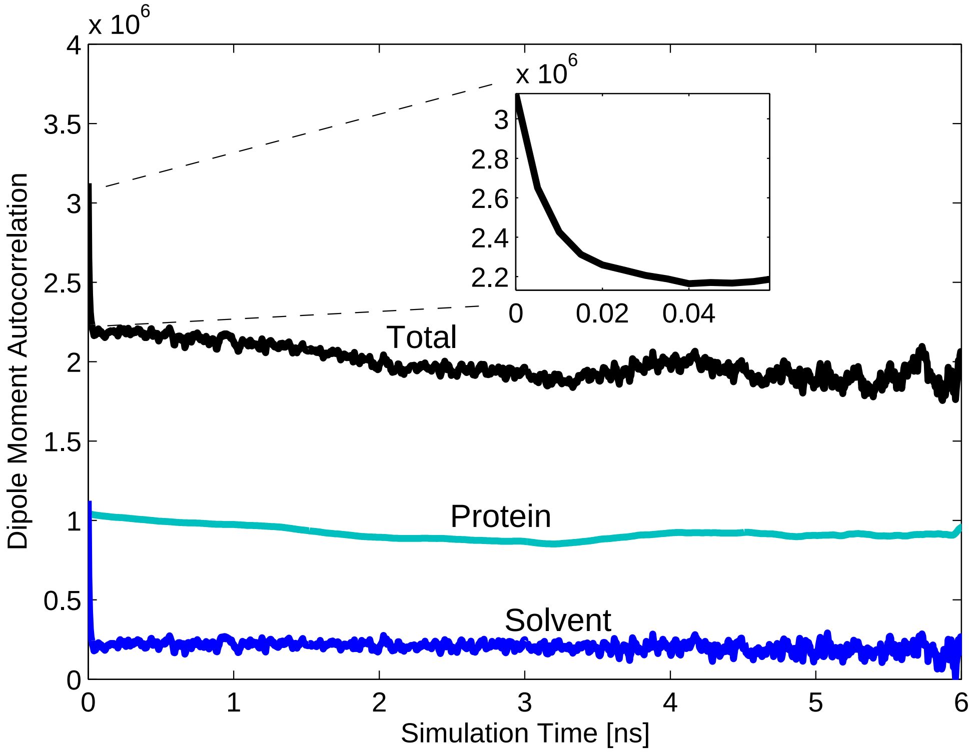


Figure 6.1: Dipole moment autocorrelation of β-Lg-water system (black), protein alone (green), and solvent alone (blue).

which, when inserted into Equations 6.14 and 6.15, would predict ε in the form of the Debye dispersion equation, where τ_1 and τ_2 are the relaxation times of the protein and water. Fitting the contribution from the solvent to Equation 6.20 gives an accurate prediction for the relaxation time of water ($\tau_2 = 8.5$ ps, measured value is 9.3 ps [117]). The fit is plotted in Figure 6.2 as a dashed line. It is obvious, however, from both figures, that to fit the protein contribution accurately the simulation length will need to be much longer to capture the protein dynamics. The measured relaxation time of β-Lg is $\sim$ 40 ns (Chapter 4), thus the simulation length should be several times longer for an accurate fitting to the autocorrelation.

The limitations of the Fourier analysis permittivity prediction method are evident in this example. First, the simulation lengths must be at least on the order of the relaxation time of the largest component in the system. At the present time, simulation lengths above 50 ns for systems of this size are very lengthy due to computation speeds. Second, the dipole moment autocorrelation is noisy and requires large ensemble averages. Consider the autocorrelation of the solvent in Figure 6.2. It is an average over 31,223 H_2O molecules and

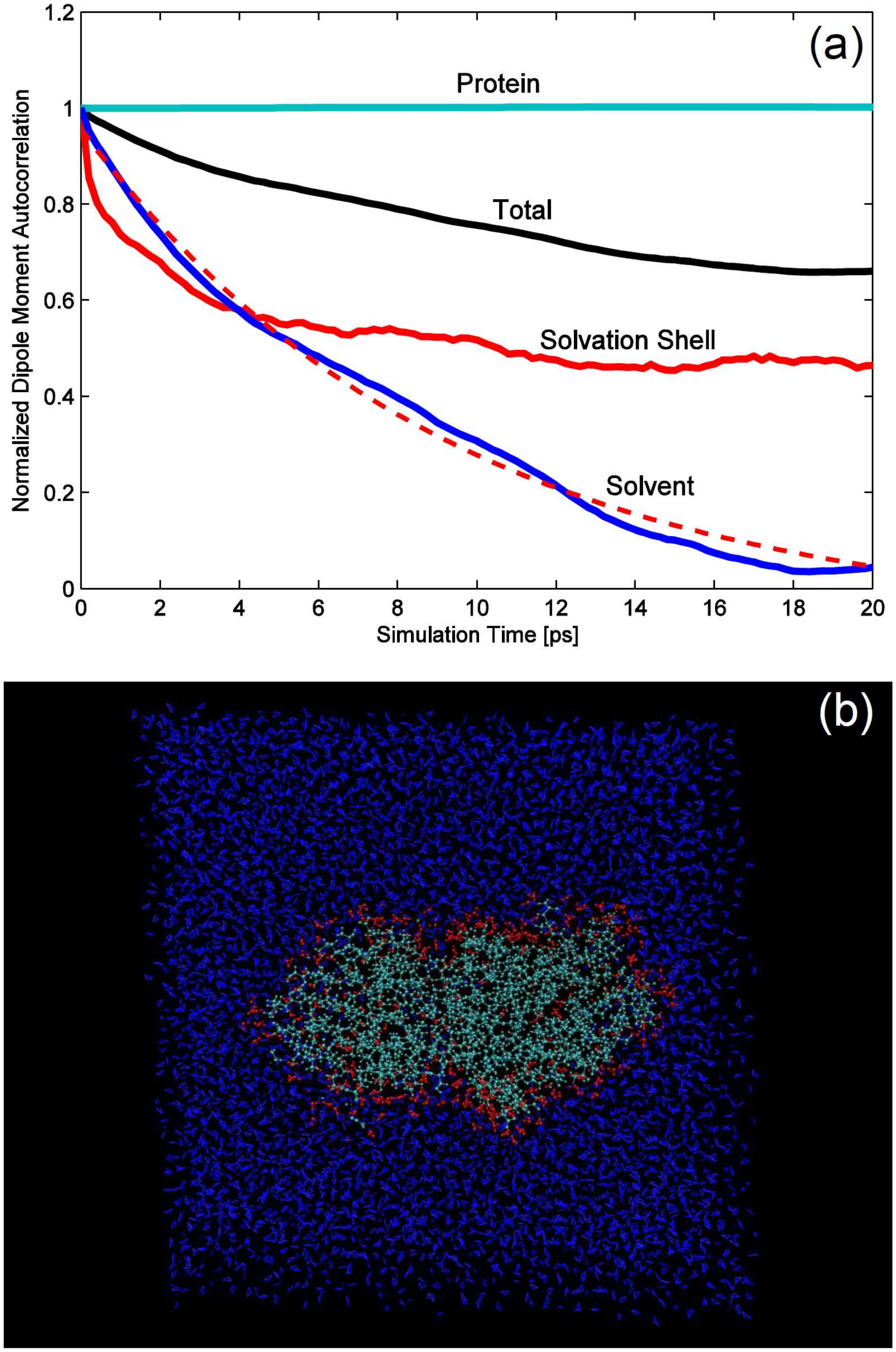


Figure 6.2: (a) Normalized dipole moment autocorrelation for β-Lg-water system (black), protein alone (green), solvent alone (blue), and solvation shell (red). The protein dipole relaxation is on a timescale much longer than that of the water and of the solvation shell. In (b), system components plotted in (a) are distinguished by color. Solvation shell is defined as all H_2O molecules within 4 Å of the protein. Dashed red line is the exponential fit using Equation 6.20.

yet is visibly noisy compared to the fit or at least visibly deviates from the two-exponential fit. Thus, for proteins and large molecules, the Fourier analysis method will likely require several greater-than-50-ns-length simulations for accurate parameterization. This calls into question the 5 ns study done by Boresch et al. [212] in which the dipole moment autocorrelation of Ubiquitin was fit to the exponential function. The relaxation time of ubiquitin is ~ 12 ns [222].

Nonetheless, a recent study by Matyushov [223] indicates that the outlook for the Fourier analysis approach could still be promising. In that study, 100-172 ns CHARMM simulations on four proteins of molecular weights ranging from 8 to 15 kDa were performed. Exponential fits to the protein autocorrelation functions of lysozyme and ubiquitin yielded relaxation times of 14.6 and 4.6 ns, respectively. These are comparable to the measured relaxation times of $\sim 10-15$ [224] ns and ~ 12 ns [222].

6.3 Rotation of Proteins in Large DC Electric Fields

An alternative method to calculate protein relaxation by applying electric fields was explored. This was motivated by the limitations of the Fourier analysis approach and the possibility of producing a faster and more accurate method to compute protein permittivity. Two different types of simulations, transient response and steady-state response, were aimed at the two parameters of interest, the relaxation time τ and the change in permittivity $\Delta\varepsilon$.

6.3.1 Transient Response

Transient response simulations are initialized with the protein dipole moment antiparallel to the applied field. As the field is turned on, the dipole vector is tracked as it rotates, eventually reaching a steady-state (i.e., an equilibrium between fully oriented and isotropic/fully disoriented) and forming an average angle $\hat{\theta}$ with the applied field. 39 different CHARMM simulations of 6 ns length were run with applied electric fields ranging from 10^5 to 10^9 V/m. The results from the simulations are displayed in Figure 6.4. At the higher field strengths the dipole moment aligns quickly with the applied field (< 0.5 ns). At lower field strengths thermal forces are more influential and impart randomization to the dipole vector.

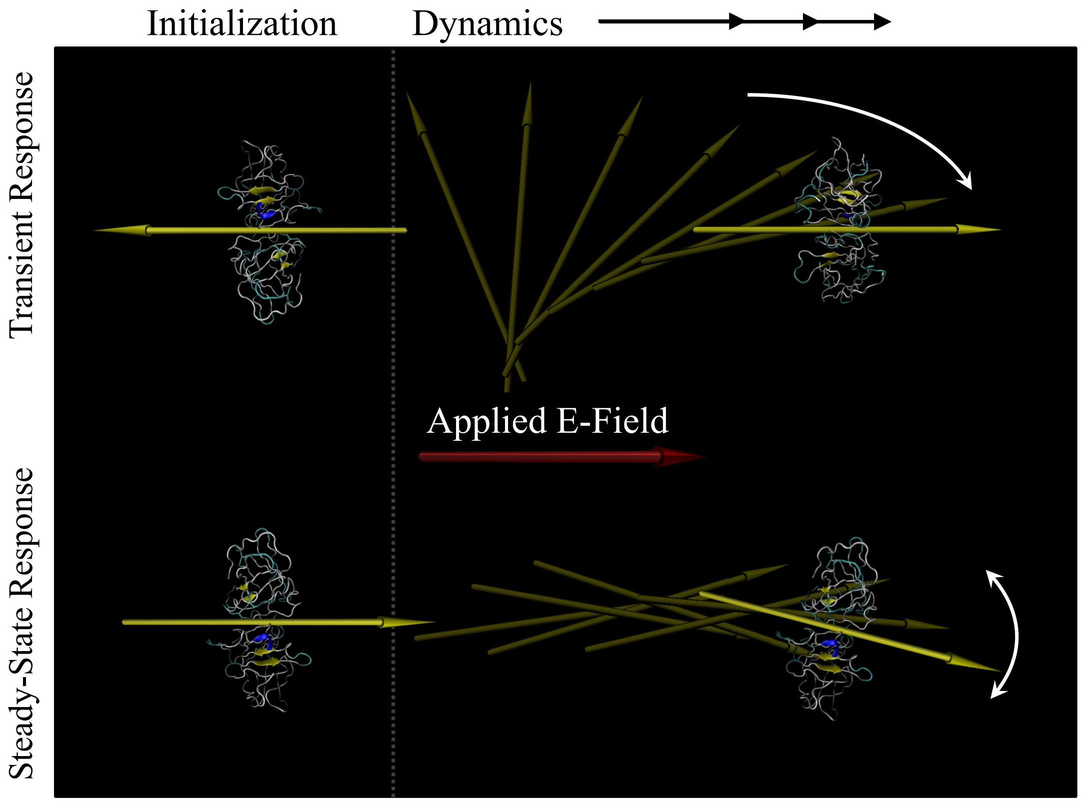


Figure 6.3: Diagram of transient and steady-state response simulations.

The dipole moment vector aligns with the applied field exponentially according to some relaxation time. This exponential behavior can be seen in the protein dipole moment alignment simulations in Figure 6.4. The relaxation time was extracted from each waveform of Figure 6.4 by fitting to an exponential function of the form,

$$\theta_{\text{fit}} = k\mathrm{e}^{-t/\tau} + \hat{\theta} \tag{6.21}$$

where τ is the fitted relaxation time and k is a constant. Results for the fitted transient-response relaxation times τ are plotted as the open circles in Figure 6.5. The objective is to relate the fitted relaxation time to the molecular relaxation time of the protein. Because the fitted relaxation times are not constant over electric field strength, we must extrapolate towards experimental conditions, i.e., field strengths less than 10^5 V/m. Two observable regimes in Figure 6.5 are designated with dashed red lines. At high field stengths the

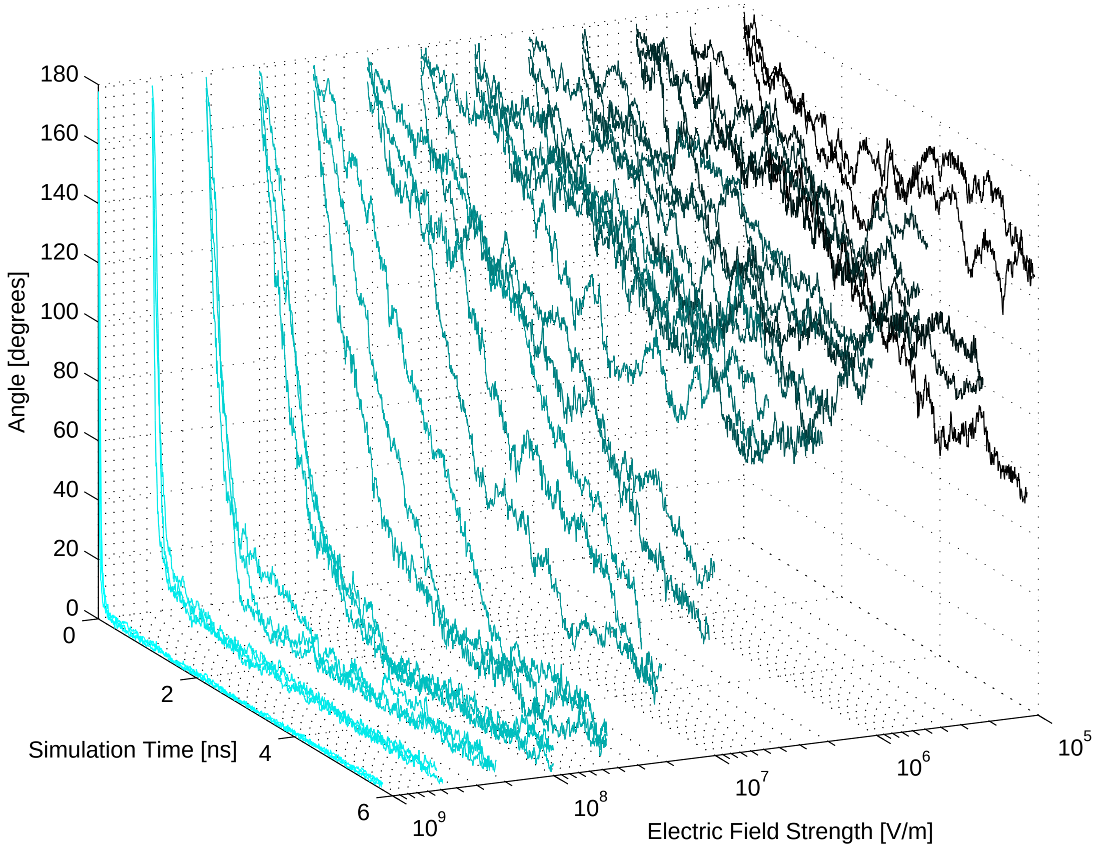


Figure 6.4: Simulated angle formed by the dipole moment of β-Lg and the applied electric field as a function of field strength and simulation time. Dipole moment is initialized antiparallel to the applied field.

molecule continues to rotate faster with field strength. As the field decreases, the dipole vector aligns slower with the field until arriving at the relaxation time of the molecule where it appears to remain constant i.e. it is no longer oriented appreciably by the applied field. Using this technique, a predicted relaxation time of 20 ns is obtained for β-Lg (measured τ for β-Lg is $\sim$ 40 ns).

It's important that at high fields the transient response fitted relaxation times match the high-field autocorrelation relaxation times (compare Figures 6.5 and 6.6). The two techniques give the same result, the transient response method being able to do so using fewer samples.

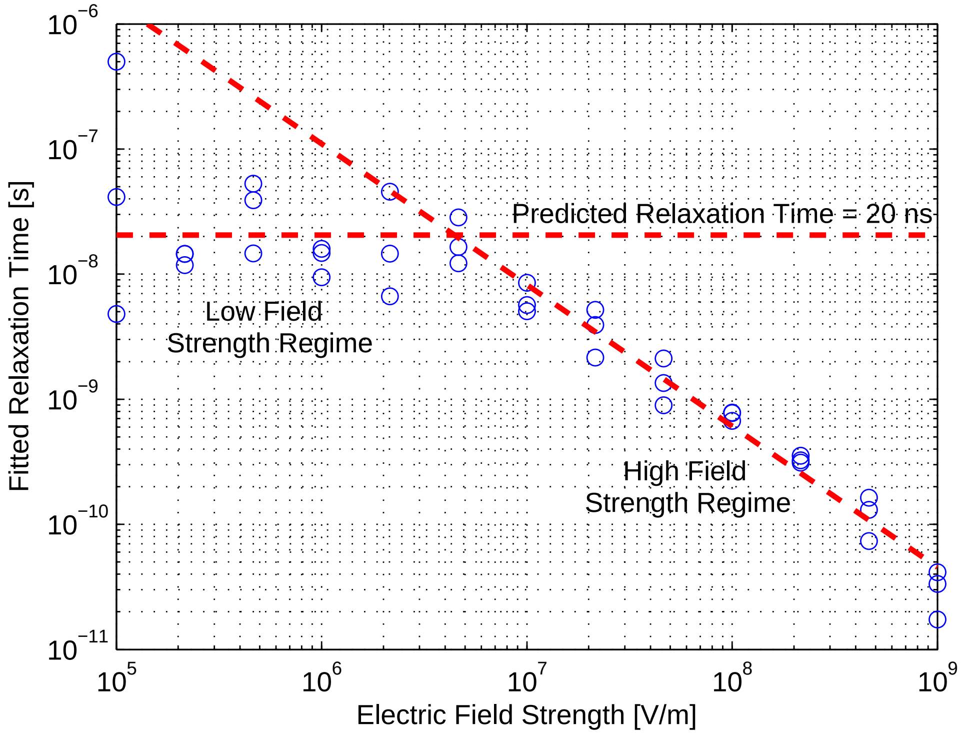


Figure 6.5: β-Lg dipole angle fitted relaxation times as a function of electric field strength. Angles from simulation (Figure 6.4) were fitted to the exponential function of Equation 6.21.

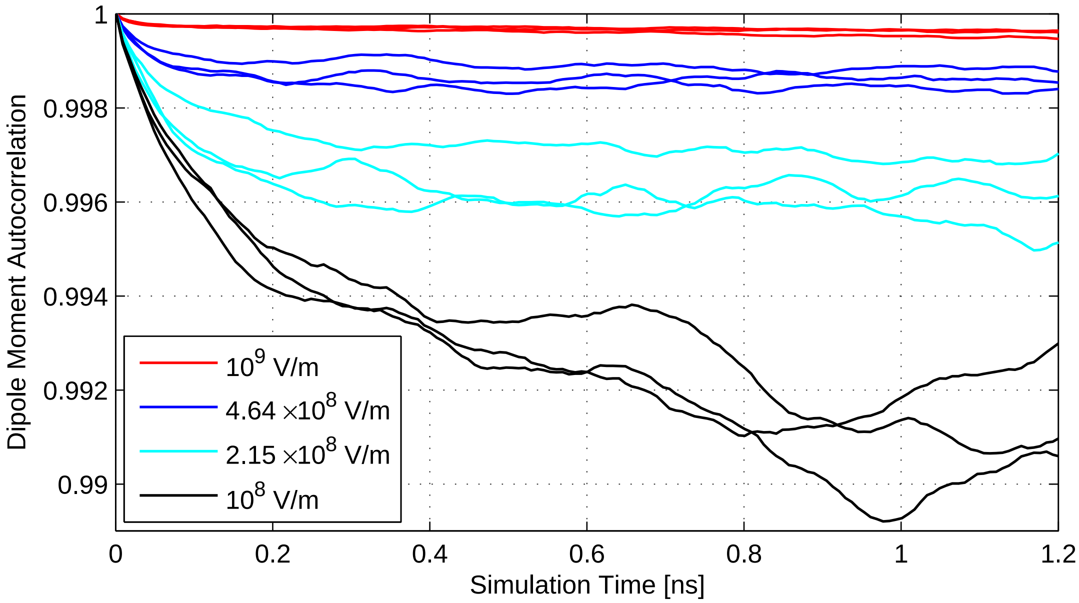


Figure 6.6: Dipole moment sample autocorrelation for β-Lg at high field strengths obtained with 6 ns CHARMM simulations.

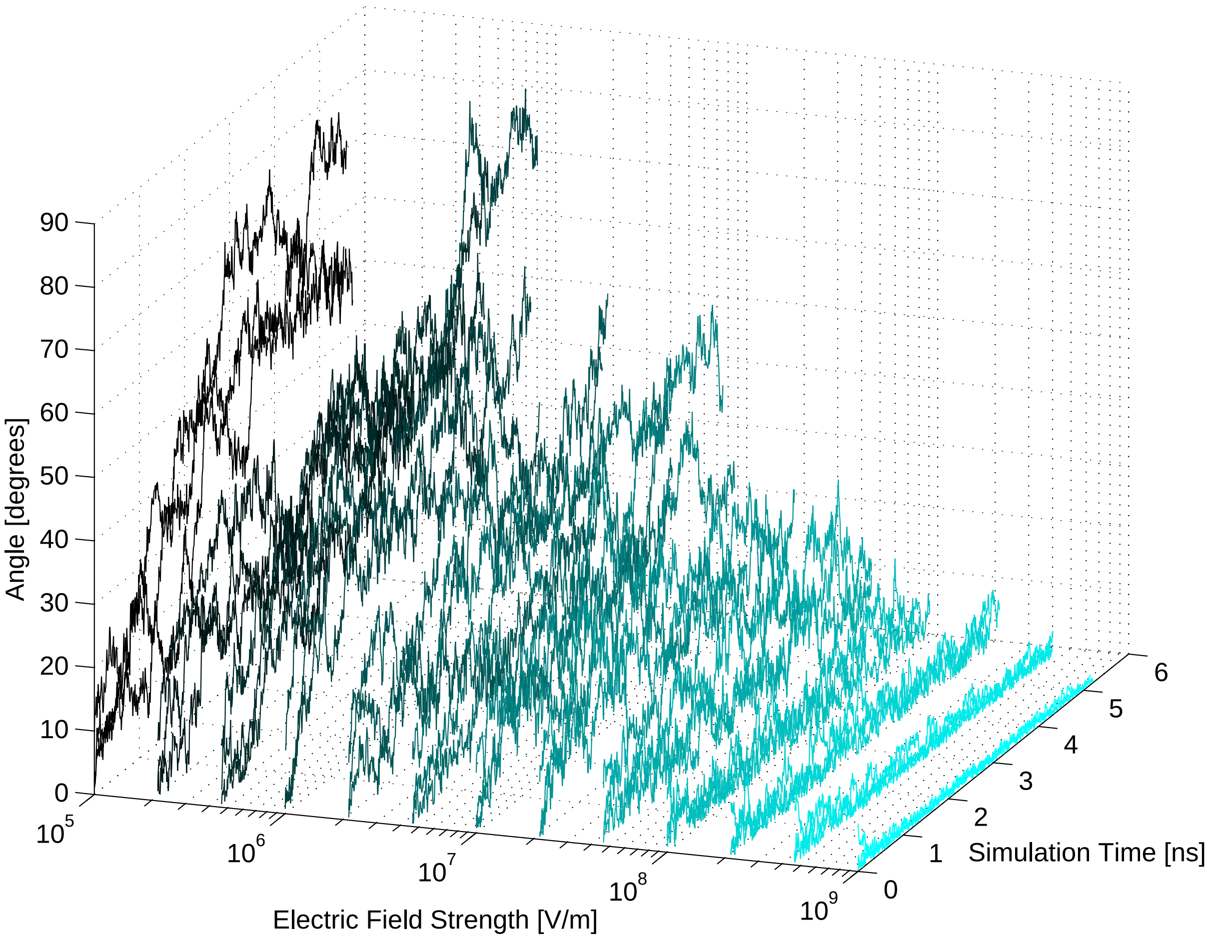


Figure 6.7: Simulated angle formed by the dipole moment of β-Lg and the applied electric field as a function of field strength and simulation time. Dipole moment is initialized parallel to the applied field.

6.3.2 Steady-State Response

Steady-state simulations are initialized with the protein dipole moment parallel to the applied field. Again, 39 CHARMM simulations of 6 ns length were run with applied electric fields ranging from 10^5 to 10^9 V/m. Results from these simulations are displayed in Figure 6.7. At lower field strengths the dipole moment vector has yet to fully explore the conformational space. Because of this, values for the average alignment with the electric field $\langle\cos(\theta)\rangle$ produced by the lower field strength simulations depart considerably from the theory (Equation 2.42) as shown in Figure 6.8. As simulation time increases, the field

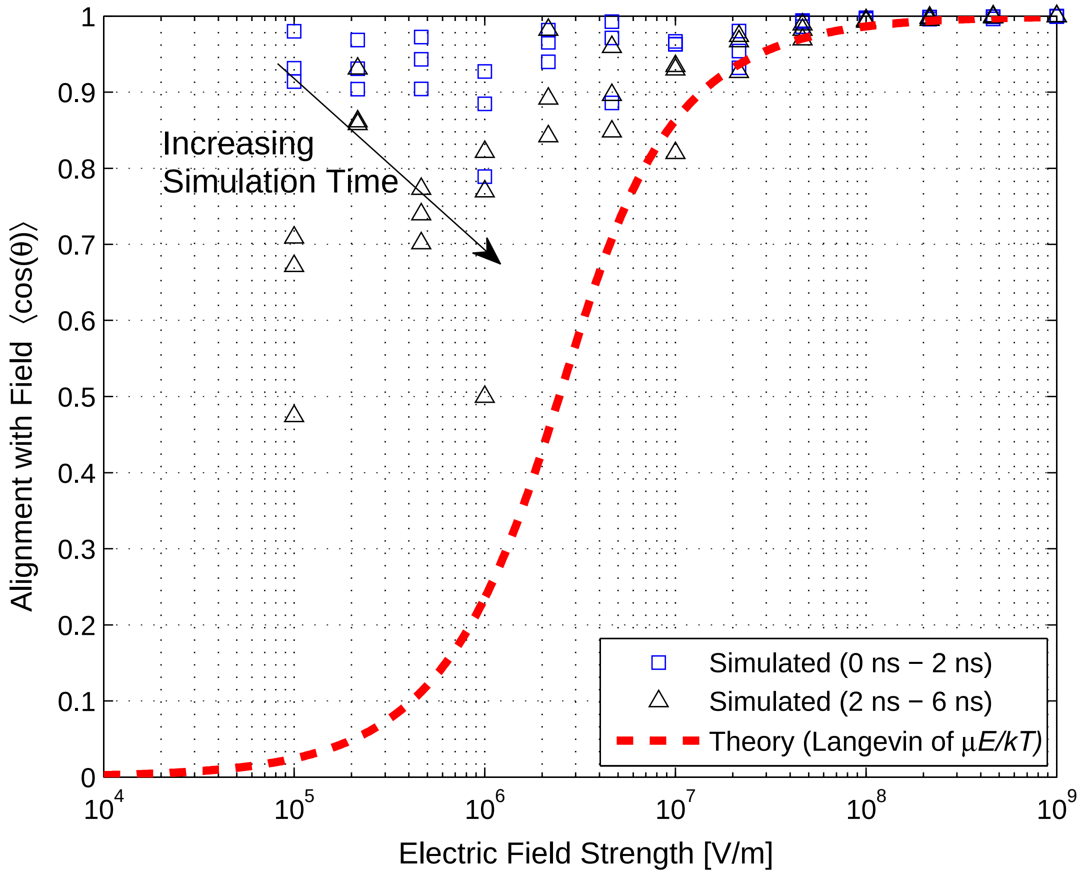


Figure 6.8: Simulated and theoretical alignment of dipole moment with applied electric field.

alignment approaches the prediction of the Langevin equation. A 6 ns simulation length appears sufficient to obtain averages for field strengths greater than $\sim 10^7$ V/m. At low field strengths, it is difficult to determine the required simulation lengths from the figure alone. Section 6.4 will address this.

6.3.3 Discussion

Although 6 ns CHARMM simulations are unable to predict the average dipole alignment, the ability to predict the protein relaxation time has clear advantages over the Fourier analysis approach. First, with the electric field method, the emphasis is placed on multiple short simulations (< 10 ns) instead of lengthy simulations (> 50 ns), and thus can be obtained more quickly (assuming the ability to run simulations in parallel). Second, the

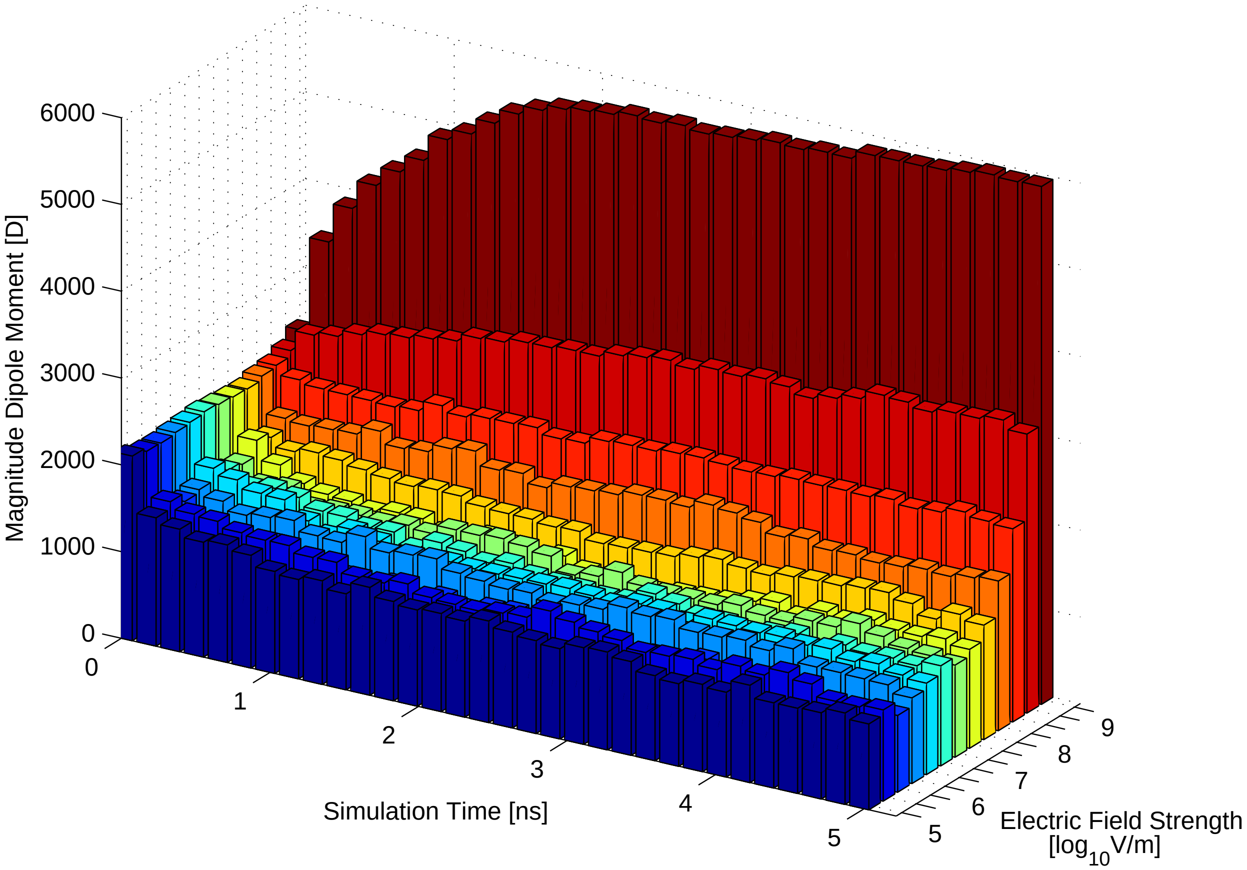


Figure 6.9: Distortion of dipole moment of β-Lg under strong electric field strengths. At very high fields, stretching the bonds takes a few ns to pull and equilibrate them. At low fields, there is an initial drop due to inadequate initial equilibration of the structure before field-application.

theory behind the electric field approach is more straightforward than the Fourier analysis approach. In essence, it is not a shortcut but a brute-force attempt to replicate an experimentally observable phenomenon by simulating the exact conditions within the dielectric cell when attached to an impedance analyzer. A disadvantage of the approach is that strong electric fields can distort the protein structure. The applied field causes positively charged amino acid groups to be pulled in the direction of the field and negatively charged groups in the opposite direction. When these forces are greater than the forces holding the protein together, the structure becomes distorted resulting in a larger magnitude dipole moment (Figure 6.9). This must be considered when using the electric field approach.

6.4 Probabilistic Analysis

A common problem in probability theory is that of estimating the mean and its corresponding error of a random process from N samples. For an independent and identically distributed (IID) random process $\{X_t;\ t = 0, 1, \ldots\}$, this problem is simple, where the sample mean, S_N, is equal to the average value of the N samples, and the variance of the sample mean, $\sigma^2_{S_N}$, is equal to σ^2_X / N. If X_t is not IID, i.e., the N samples are in some way correlated, the problem of finding $\sigma^2_{S_N}$ becomes a bit more complicated (S_N is the average value in all cases). The objective in this section is to first derive $\sigma^2_{S_N}$ for an exponentially-correlated stationary random process. In this case the normalized autocorrelation is equal to $\rho^{|i-j|}$ where $0 \leq \rho < 1$. Larger values of ρ correspond to more correlation between samples and $\rho = 0$ is the IID case. After the derivation, the result is applied to the simulated movement of the dipole moment in molecular dynamics where the random process to be estimated is the alignment of the dipole vector with the applied field.

6.4.1 Variance of the Sample Mean for Exponentially-Correlated Random Processes

Let X_t be an identically distributed, but not necessarily indepedent, stationary random process. Let m_X and σ^2_X be the common mean and variance of the process. Its autocorrelation function, $R_X(i-j)$, is defined by

$$R_X(i-j) = E[X_i X_j], \tag{6.22}$$

and its covariance function, $K_X(i-j)$, is defined by

$$\begin{aligned} K_X(i-j) &= E[(X_i - E[X_i])(X_j - E[X_j])] && (6.23) \\ &= R_X(i-j) - E[X_i]\, E[X_j]. && (6.24) \end{aligned}$$

If X_t is an exponentially-correlated process, i.e., whose normalized autocorrelation is equal to $\rho^{|i-j|}$ (or $e^{-\frac{|i-j|}{\tau}}$), its autocorrelation is given by

$$R_X(i-j) = m_X^2 + \sigma_X^2 \rho^{|i-j|}. \tag{6.25}$$

The sample mean S_N of the process is defined as

$$S_N = \frac{1}{N}\sum_{i=0}^{N-1} X_i. \tag{6.26}$$

The first and second moments are

$$E[S_N] = \frac{1}{N}\sum_{i=0}^{N-1} E[X_i] = m_X \tag{6.27}$$

and

$$\begin{aligned} E\left[S_N^2\right] &= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} E[X_i X_j] \quad (6.28)\\ &= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} R_X(i-j). \quad (6.29)\end{aligned}$$

Plugging in $R_X(i-j)$ for an exponentially-correlated process, (Equation 6.25), $E[S_N^2]$ becomes

$$\begin{aligned} E\left[S_N^2\right] &= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(m_X^2 + \sigma_X^2\rho^{|i-j|}\right) \quad (6.30)\\ &= m_X^2 + \frac{\sigma_X^2}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\rho^{|i-j|} \quad (6.31)\\ &= m_X^2 + \frac{\sigma_X^2}{N^2}\left[\sum_{i=0}^{N-1} 1 + \sum_{i=0}^{N-2}\rho^{-i}\sum_{j=i+1}^{N-1}\rho^{j} + \sum_{i=1}^{N-1}\rho^{i}\sum_{j=0}^{i-1}\rho^{-j}\right] \quad (6.32)\\ &= m_X^2 + \frac{\sigma_X^2}{N} + \frac{2\sigma_X^2\rho}{N^2(1-\rho)}\left[(N-1) - \left(\frac{\rho-\rho^N}{1-\rho}\right)\right]. \quad (6.33)\end{aligned}$$

For large N, the approximation $N-1 \approx N$ reduces the expression to

$$E\left[S_N^2\right] \approx m_X^2 + \frac{\sigma_X^2}{N}\left(\frac{1+\rho}{1-\rho}\right) - \frac{2\sigma_X^2\rho^2}{N^2(1-\rho)^2}\left[1-\rho^{N-1}\right]. \tag{6.34}$$

Again, because N is large, the term with $1/N^2$ is negligible and $E[S_N^2]$ reduces to

$$E\left[S_N^2\right] \approx m_X^2 + \frac{\sigma_X^2}{N}\left(\frac{1+\rho}{1-\rho}\right) \tag{6.35}$$

and the variance of the sample mean is thus

$$\begin{aligned}
\sigma_{S_N}^2 &= E\left[S_N^2\right] - E\left[S_N\right]^2 \qquad\qquad (6.36)\\
&\approx \frac{\sigma_X^2}{N}\left(\frac{1+\rho}{1-\rho}\right) \qquad\qquad (6.37)\\
&= \frac{\sigma_X^2}{N}\left[\frac{1+e^{\left(-\frac{1}{\tau}\right)}}{1-e^{\left(-\frac{1}{\tau}\right)}}\right] \qquad \text{for } N \gg 0. \qquad\qquad (6.38)
\end{aligned}$$

In the IID case ($\rho = 0$ and $\tau = 0$)

$$\sigma_{S_N}^2 = \frac{\sigma_X^2}{N}. \tag{6.39}$$

Another important case is when N is large, τ is large, and $N \gg \tau$ (long simulations of large molecules). Using the Taylor series approximation $e^x \approx 1 + x$ for small x, Equation 6.38 simplifies to

$$\begin{aligned}
\sigma_{S_N}^2 &\approx \frac{\sigma_X^2}{N}\left[\frac{1+\left(1-\frac{1}{\tau}\right)}{1-\left(1-\frac{1}{\tau}\right)}\right] \qquad\qquad (6.40)\\
&= \frac{\sigma_X^2}{N}\left[\frac{2-\frac{1}{\tau}}{\frac{1}{\tau}}\right] \qquad\qquad (6.41)\\
&= \frac{2\tau}{N}\sigma_X^2 \qquad \text{for } N \gg \tau \gg 0. \qquad\qquad (6.42)
\end{aligned}$$

The negative effect of the exponential correlation is observed by comparing Equations 6.39 and 6.42. Only after 2τ samples are taken is a sample considered independent from the previous. For example, a random process with $\tau = 40$ ns ($\beta-$Lg) requires a simulation length of $N = 160$ ns to reduce the variance of the sample mean to just one half of the variance of the underlying distribution. The next sections will address the implications of this.

6.4.2 Markov Chain Monte Carlo Simulations

Electric field simulations in CHARMM produce time realizations of the Markov process $\{\Theta_t;\ t = 0, 1, \ldots\}$, where Θ_t is the angle formed by the dipole vector and the applied field. To verify that this process is indeed exponentially correlated, Monte Carlo simulations were carried out in MATLAB.

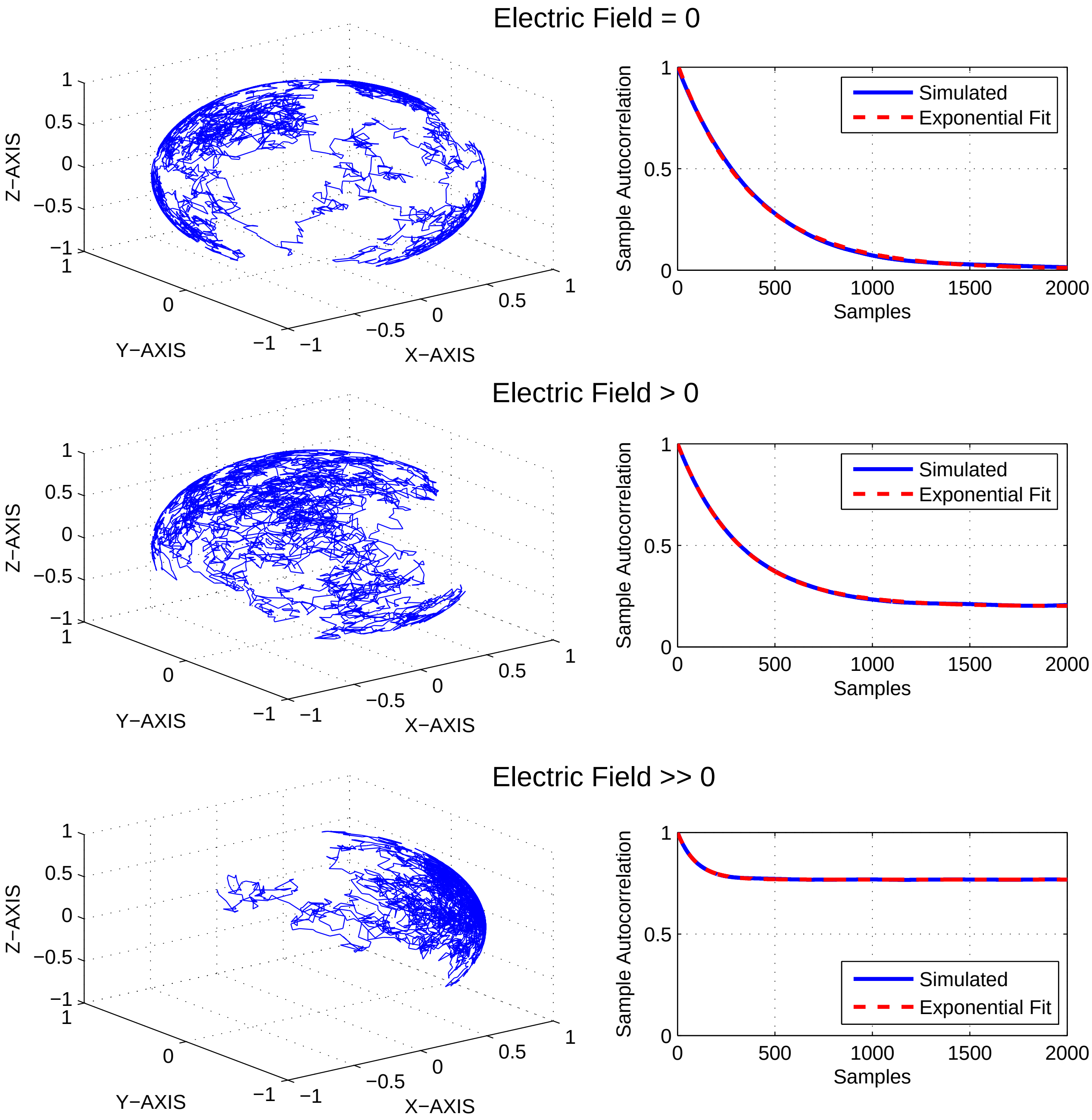
The simulations were initialized with the dipole vector in the negative x direction. To model Brownian motion, at each simulation step the dipole vector was incremented by a realization (using a random number generator) of a three-dimensional Gaussian distribution. To model the influence of an electric field, the Gaussian distribution was given a non-zero mean in the positive x direction. In the case of no electric field, the Gaussian distribution was zero mean. Figure 6.10 shows trajectories and sample autocorrelations for three different field strengths. The autocorrelations conform closely to an exponential function, which justifies the usage of Equations 6.38 and 6.42 for their analysis.

Another goal of the Monte Carlo simulations was to obtain a relationship between the mean and variance at all possible field strengths. To accomplish this, simulations of length $N = 10^7$ were run over a wide range of field strengths. A histogram was taken of the resulting trajectory projected onto the x-axis to estimate the probability density function (PDF). The PDF estimates are shown in Figure 6.11 and their means and standard deviations are plotted against each other in Figure 6.12. As an accuracy check, analytical solutions for m_Θ and σ_Θ^2 can be obtained in the zero electric field case. The PDF for Θ, $f_\Theta(\theta)$, is computed from the joint distribution, $f_{\Theta,\Phi}(\theta, \phi)$, defined as

$$f_{\Theta,\Phi}(\theta,\phi) = \frac{1}{4\pi}\sin\theta \qquad \text{for } 0 \le \theta \le \pi,\ 0 \le \phi \le 2\pi \tag{6.43}$$

using a spherical coordinate system. The marginal PDF is found by integrating the joint PDF over all possible values of Φ. This yields

$$\begin{aligned} f_\Theta(\theta) &= \int_{\phi=0}^{2\pi} f_{\Theta,\Phi}(\theta,\phi)\,d\phi && (6.44) \\ &= \frac{1}{2}\sin\theta. && (6.45) \end{aligned}$$


Electric Field = 0

Electric Field > 0

Electric Field >> 0

Figure 6.10: Simulated dipole moment trajectory and autocorrelation for three field strengths. Trajectories (left side) display the path of the dipole moment over the first 4,000 samples. Sample autocorrelations (right side) were computed for $N = 10^7$.

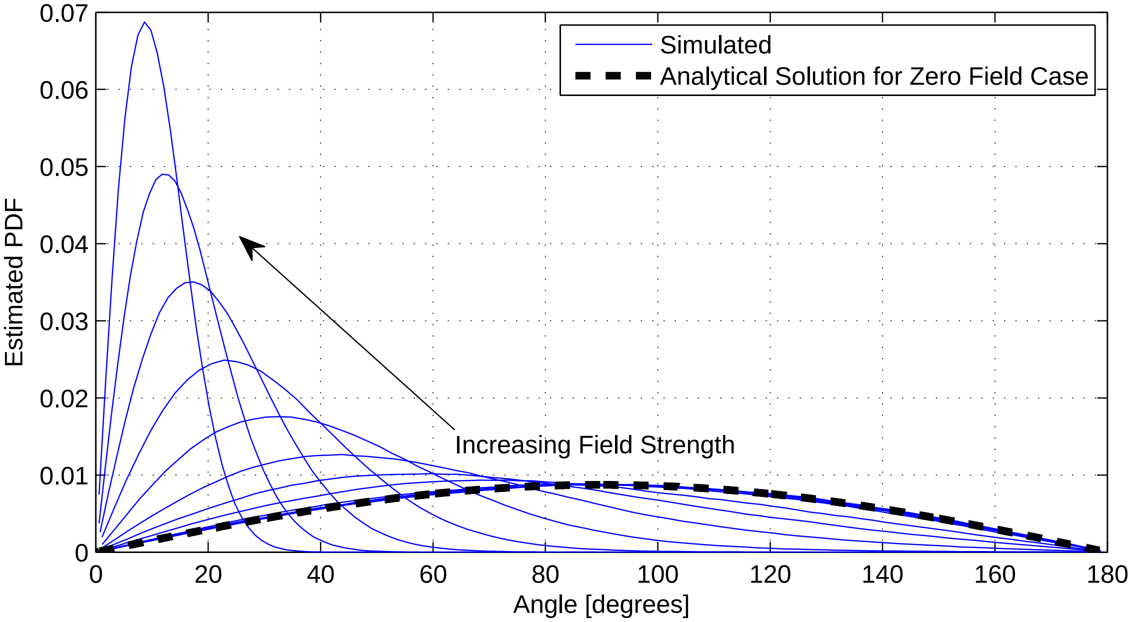


Figure 6.11: Estimated PDFs of dipole vector angle with the applied field for various field strengths ($N = 10^7$).

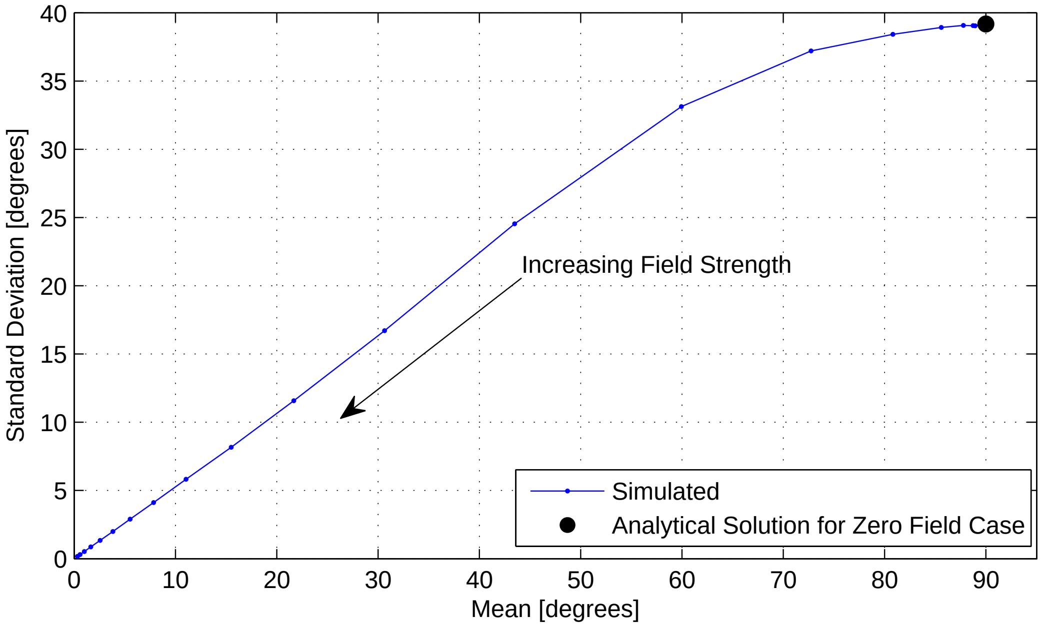


Figure 6.12: Mean and standard deviation of dipole vector angle with the applied field for various field strengths ($N = 10^7$).

Table 6.1: Simulation details for β-Lg electric field simulations in CHARMM.

Field Strength [V/m]	$\hat{\tau}$ [ns]	$\hat{\sigma_\Theta}$ [degrees]	Realizations of S_N ($N = 6$ ns) [degrees]	σ_{S_N} ($N = 6$ ns) [degrees]	N for $\sigma_{S_N} = 0.1$ [ns]
10^9	0.04	0.5	0.96 / 1.22 / 1.00	0.06	2
$4.64 \cdot 10^8$	0.08	1.5	2.98 / 1.76 / 1.69	0.25	36
$2.15 \cdot 10^8$	0.2	2	5.13 / 3.35 / 3.70	0.52	160
10^8	0.7	3	5.91 / 5.47 / 5.36	1.45	1,260
10^7	10	20	19.9 / 17.9 / 32.5	36.5	800,000
10^6	35	30	39.0 / 34.2 / 58.7	102.4	6,300,000
10^5	40	39	47.4 / 43.1 / 61.1	142.4	12,000,000

The mean and standard deviation of Θ are found through the equations

$$\begin{aligned} E\left[\Theta\right] &= \int_{\theta=0}^{\pi} \frac{1}{2} \sin\theta \left(\theta\right) d\theta && (6.46) \\ &= \frac{\pi}{2} \qquad \left(90^\circ\right) && (6.47) \end{aligned}$$

and

$$\begin{aligned} \operatorname{std}\left(\Theta\right) &= \sqrt{\operatorname{var}\left(\Theta\right)} && (6.48) \\ &= \sqrt{\int_{\theta=0}^{\pi} \frac{1}{2} \sin\theta \left(\theta - E\left[\Theta\right]\right)^2 d\theta} && (6.49) \\ &= \sqrt{\frac{\pi^2}{4} - 2} \qquad \left(39.2^\circ\right). && (6.50) \end{aligned}$$

The simulated solutions converge to the analytical solution in the zero electric field case as plotted in Figures 6.11 and 6.12.

6.4.3 Accuracy of CHARMM Electric Field Simulations

Due to the exponentially-correlated nature of the random process Θ_t, it is possible to determine the accuracy of sample means computed on its distribution. Table 6.1 lists σ_{S_N} values for the 6 ns CHARMM simulations (penultimate column). Also, the required values of N to achieve $\sigma_{S_N} = 0.1$ degrees are listed in the far right column. In the zero field case up to fields of 10^5 V/m, the required simulation length to achieve $\sigma_{S_N} = 0.1$ accuracy is 12 ms. This number is astronomical in actual time (about 9,000 years on marylou6.byu.edu!).

As field strength increases, the required simulation length decreases. For a field of 10^9 V/m, the required simulation length to achieve $\sigma_{S_N} = 0.1$ accuracy is 2 ns. This can be performed on marylou6.byu.edu in less than a week of actual time, which is much more practical.

In summary, electric field approaches reduce dipole moment correlation and allow high-field relaxation parameters to be computed in short simulation times. By themselves, the high-field parameters are not directly useful for comparison to experimental data (unless of course one is performing high-field dielectric spectroscopy [74]). Their usefulness is for extrapolating toward low-field conditions where relaxation parameters correspond to those measured experimentally and yet are impractical to simulate directly (as shown in Figures 6.5, 6.8, and Table 6.1). Future work will extend this technique to amino acids and other small molecules where it can be further validated.

6.5 Use of Molecular Dynamics to Improve Structure-Based Predictions of Dipole Moment

Analysis of protein structures obtained through X-ray crystallography is complicated by possible structural differences between frozen protein crystal and dissolved liquid states [225]. Additionally, the influence of temperature and ionic strength on the protein structure may not be negligible. Therefore, molecular dynamics simulations in CHARMM were employed to equilibrate structures in aqueous solution.

Molecular dynamics simulations were performed by placing a water cube with side lengths of 10 nm around the molecule using default side chain charges (corresponding to the model pK_a at pH 7), deleting overlapping solvent molecules, and heating the system to 25 °C. Before heating, energy minimizations were performed to remove bad contacts between atoms which could pose problems during the heating phase. The heating phase consisted of 3 ps of linear heating from -275 °C to 25 °C, followed by an equilibration phase lasting 5 ns at constant pressure and temperature. After heating and equilibration, the resulting protein structure was extracted for charge moment calculation using pK_as calculated from the H++ server.

Figure 6.13 shows computed and measured values of μ for refined and unrefined β-Lg structures. In this case, use of the refined PDB model improved agreement with theory

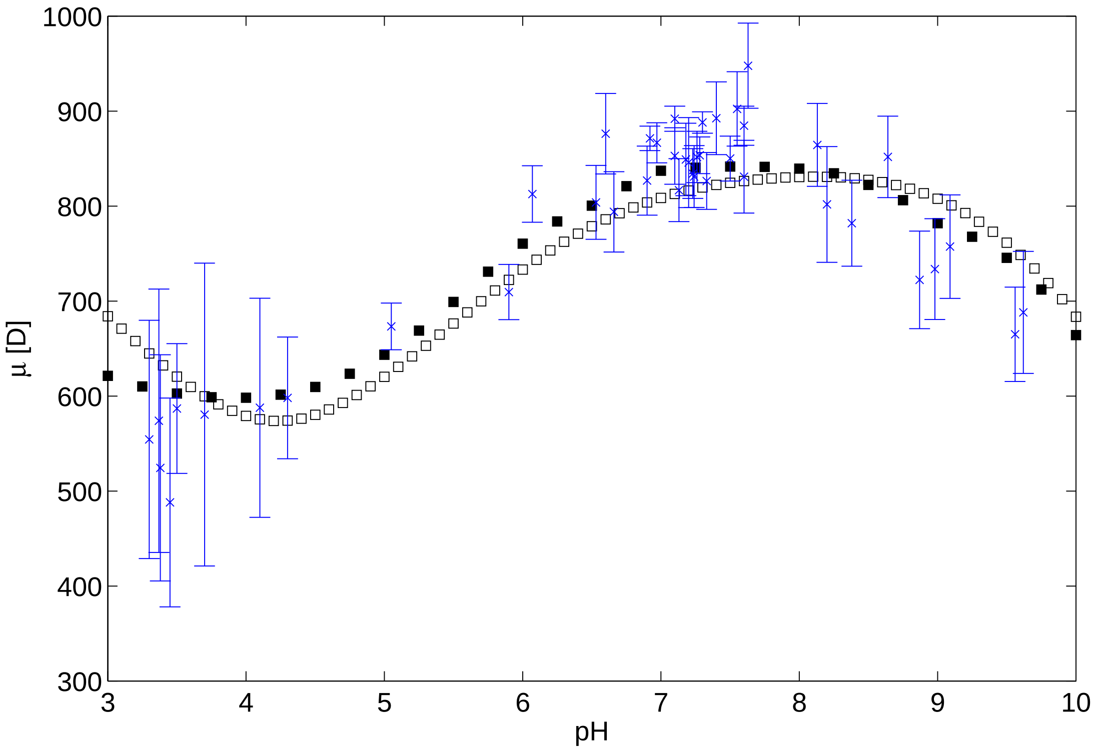


Figure 6.13: Measured and calculated dipole moments of refined (■) and unrefined (□) β-Lg structures at various pH. Both structures were calculated using an ε_{low} of 6. Use of the refined PDB model improved agreement with the measurements by $\sim$2%.

by $\sim$2%. This improvement emphasizes the important role molecular dynamics simulations could play in developing accurate electrostatic models of proteins from PDB models. Additionally, molecular dynamics of the PDB model reduced differences between theory and measurements at nearly all pH considered. Note that it is possible the improved agreement is because the refined structure achieved via molecular dynamics simply represents a better average structure, and that the improvement has nothing to do with molecular dynamics itself. Further molecular dynamics relaxations of the structure with pH-appropriate titration states could be tested for additional agreement of calculations with measurements; however, this was not done in this work.

Chapter 7

Dielectric Spectroscopy of Molecular Interactions

Molecular interactions in liquid are the driving mechanisms behind biological processes. When proteins bind to each other or another substrate, it is likely that their conformation and their electromagnetic properties will change. This is the basis of many studies using fluorescence anisotropy [226] or circular dichroism [227]. The physical changes associated with molecular interactions can be subtle and difficult to detect. For example, in drug development, to achieve high binding affinities and selectivities, drugs are conceptualized with shapes that are highly complementary to the protein receptor [228, 229]. In many cases binding is internalized in the receptor protein [230], leaving its overall shape roughly the same.

This chapter examines several well-known systems using the concepts developed in the previous six chapters. One major motivation behind sensing molecular interactions is to detect the introduction of a biological species, whether pathogenic or not. Experiments of the avidin-biotin complex demonstrate how dielectric spectroscopy can detect the introduction of a small molecule which binds internally to a much larger molecule. Next, pH-dependent binding experiments between hen lysozyme and β-Lg and hen lysozyme and bovine serum albumin are carried out. These experiments investigate the use of dielectric spectroscopy for studying protein aggregation. As some of the required proteins are quite expensive, a new low-volume dielectric cells (60 μl) is presented.

7.1 Avidin-Biotin: Protein-Ligand Interaction

The avidin-biotin binding reaction is used heavily in molecular biology because it is well understood and has a high binding constant ($K_{\mathrm{a}} \approx 10^{15}$ M^{-1} [231]). Additionally, the three-dimensional structures of the protein, ligand, and complex have been solved (Fig-

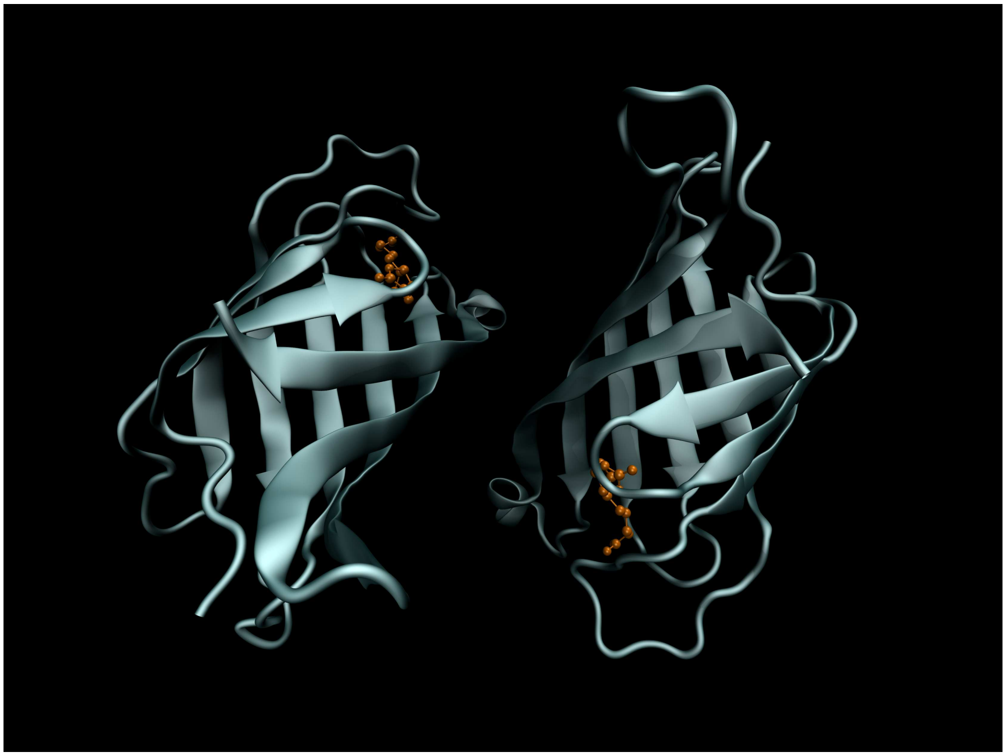


Figure 7.1: Structure of dimeric avidin-biotin complex from x-ray diffraction [18], drawn using VMD [6]. Biotin is represented by the orange ball-and-stick model.

ure 7.1). Differences in the physical characteristics of free avidin and biotin-bound avidin, although small, have been successfully detected using other biophysical techniques. Using flourescence spectroscopy peak fluorescent wavelengths were 338 nm for avidin and 328 nm for the avidin-biotin complex [232].

Here, we find that relaxation frequencies of the protein and bound complex are within the measurable bandwidth, allowing extraction of relaxation parameters before and after binding events. First, binding of avidin with biotin is demonstrated then general protein binding between avidin and biotin-labeled bovine serum albumin is performed.

7.1.1 Acrylic Low-Volume Dielectric Cell

Because the proteins being investigated are the most expensive component for the measurements, it was desired that the interrogated volume of the cell be small. The cell size was reduced considerably by machining the cavity in a single piece of acrylic with dimensions of 18 mm × 12 mm × 9 mm. A schematic and photograph of the cell attached

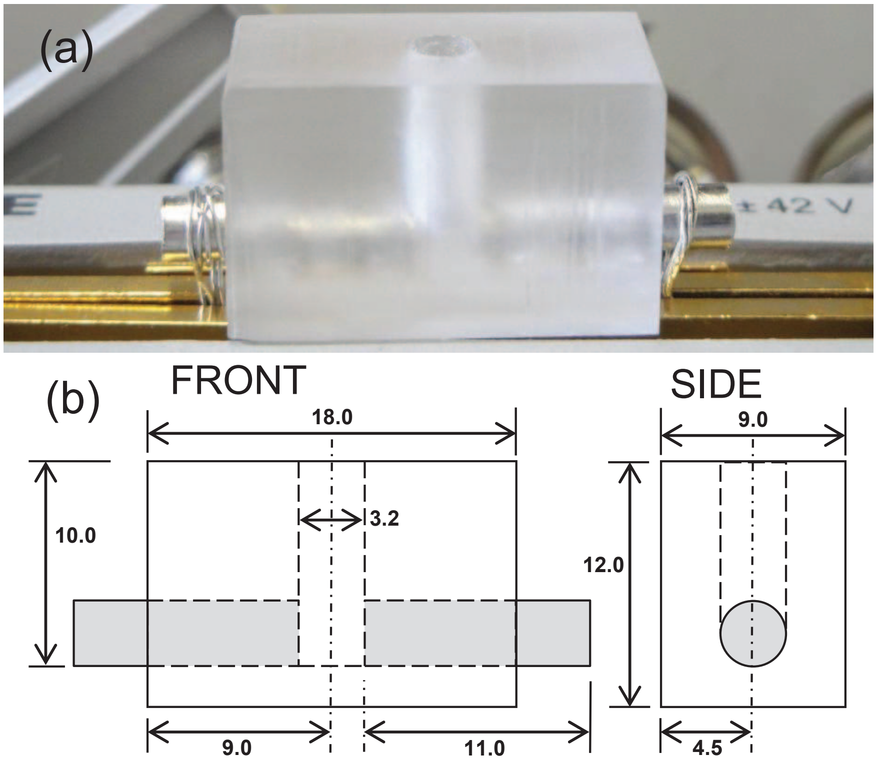


Figure 7.2: (a) Photograph and (b) schematic of dielectric cell (dimensions in mm). The interrogated volume is 60 μl. Liquids are added and removed using microcapillary tubes.

to the impedance analyzer are found in Figure 7.2. The electrodes were connected via wires to the terminals of a 16047E test fixture mounted on the Agilent 4294A Precision Impedance Analyzer.

1/8" round stainless steel (Grade 304) was used for the electrodes. Each were cut with edges slightly tapered on an end mill to ease the insertion into the acrylic. The electrode faces were ground with 600, 800, and 1200 grit silicon carbide abrasive discs in a Spectrum System 2000 (LECO Corp.). The electrodes were then press-fit into the acrylic.

7.1.2 Experimental Protocol

Avidin (A9275) and biotin (B4501) lyophilized powders were obtained from Sigma and used without further purification. The proteins were reconstituted in 0.1 mM HCl and stored in microcentrifuge tubes at concentrations of 20 mg/ml and 5 mg/ml, respectively. They were then refrigerated until used.

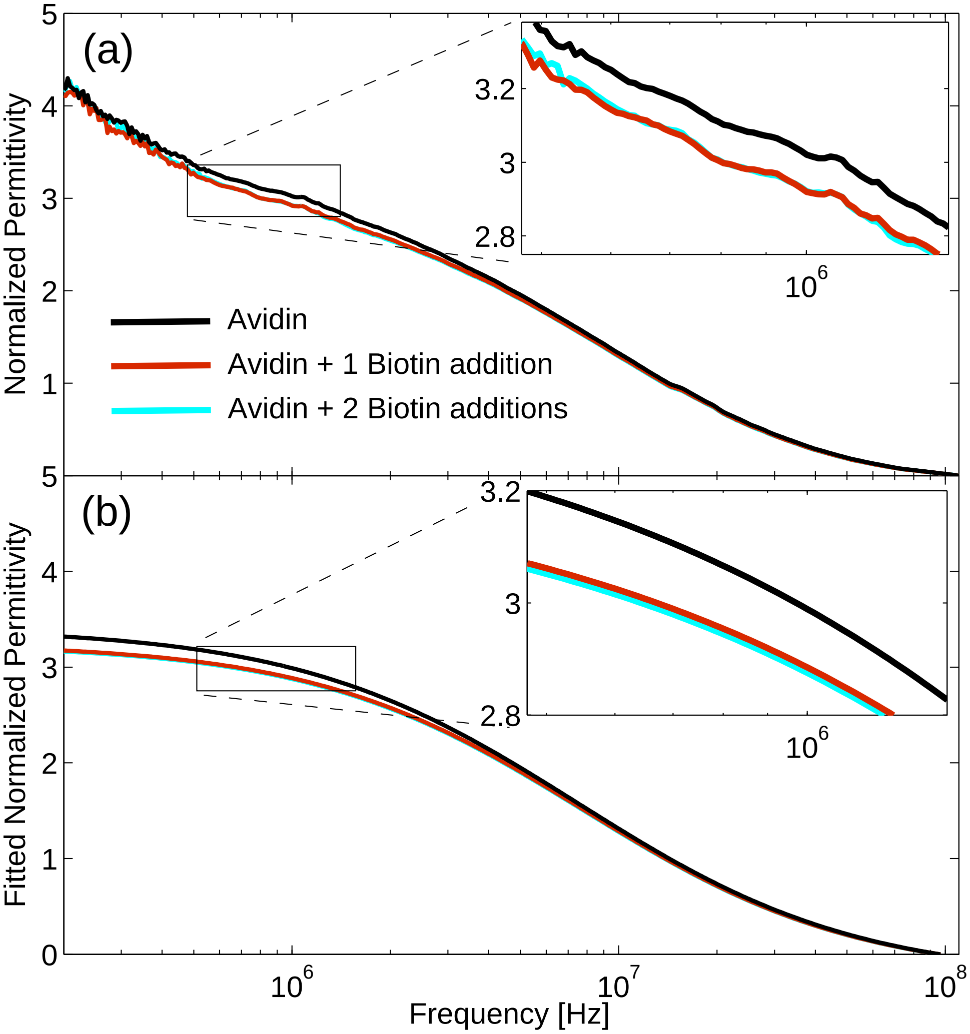


Figure 7.3: (a) Raw and (b) fitted normalized permittivity of avidin before and after binding with biotin. After the first biotin addition, the binding sites of avidin become saturated (red line). The second biotin addition does not significantly affect the observed relaxation (cyan line, barely visible behind red).

The inital concentration of avidin was raised to 3 mg/ml. While continuously recording sweeps, two additions of 2 μl of the biotin solution were made spaced 10 minutes apart. The concentration of biotin was such that, after the first addition, the number of biotin molecules exceeded the number of available avidin binding sites. The approximate molarities were 200 μM avidin monomer and 700 μM biotin.

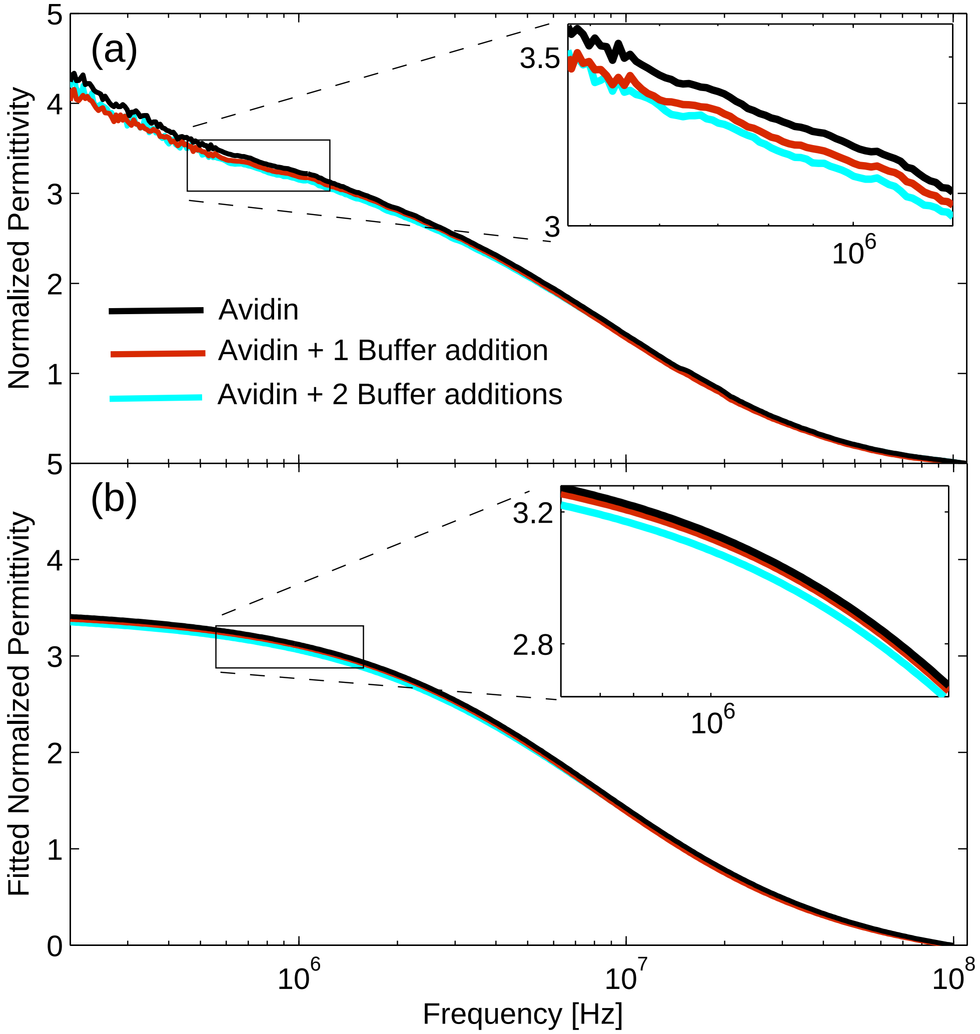


Figure 7.4: (a) Raw and (b) fitted normalized permittivity of avidin before and after two additions of buffer solution.

7.1.3 Results and Discussion

Figures 7.3 and 7.4 show the results of the avidin-biotin binding experiment. In Figure 7.3, the three curves denote the solution permittivity with (1) avidin only, (2) avidin + 1 biotin addition, and (3) avidin + 2 biotin additions. When biotin was first added, the change in permittivity $\Delta\varepsilon$ of the dominant relaxation decreased, and the relaxation time τ appeared to remain constant. This reflects the change in dipole moment before and after each avidin molecule binds two biotin. The biotin mass is small and the binding is internal (inside the avidin β-barrels), thus the hydrodynamic radius and consequently τ would not be expected to change substantially. When the second biotin addition is made, the permittivity spectrum remains relatively unchanged. This indicates the number of bound molecules remains the same; only the number of free biotin increases. The high binding affinity between the two molecules ensures that after the first biotin addition a very high percentage of binding sites are occupied. Because the relaxation of biotin occurs above 110 MHz, any permittivity contributions from free biotin would be absorbed by normalized permittivity.

In Figure 7.4 this experiment was repeated with the only difference being that the biotin solution was replaced with a buffer solution containing only 0.1 mM HCl. This was done to confirm that the perceived reaction between avidin and biotin was not simply due to an increase in solution volume. The differences between permittivity spectra when buffer solution is added are small and are apparently due only to dilution of the protein.

The observed shift in dielectric properties of bound and unbound avidin is remarkable considering the small physical changes the molecule undergoes. Avidin is dimeric at low concentrations in solution [233], and the binding of each avidin dimer ($M \approx 30$ kDa) to two biotin molecules ($M \approx 0.5$ kDa) is accompanied by an increase in mass of less than 2%. The previously unmeasured dipole moment of avidin alone was ~ 599 Debye (D). The dipole moment of biotin-bound avidin was somewhat smaller, ~ 591 D. Two possible mechanisms for this decrease are dipole vector cancellation of the added biotin molecules, and structural adjustments of the dimer that accompany binding.

The theoretical approach described in Chapter 5 was used to calculate μ and a minimum-volume enclosing ellipsoid algorithm was used to estimate axis lengths and the equivalent volume spherical radius, a_{calc} (described in Appendix B). The results of these computations are displayed in Figure 7.5 and in Table 7.1 alongside corresponding dielectric spectroscopy measurements. The measurement error ranges represent 95% confidence intervals derived from experimental and fitting errors. The agreement between measurements and calculations is quite good. The discrepancies in a may be due to the non-spherical shape of avidin. Dimeric molecules typically rotate about the shorter axis of rotation [113]; a_{calc} is more indicative of the longer axis of avidin, and a_{meas} the shorter. When the two biotin molecules were included in the calculations for μ_{calc} and a_{calc}, they remained nearly unchanged. It is plausible that avidin flexes and tightens around biotin [234], which, if true,

Table 7.1: Measured and calculated dipole moment μ and hydrodynamic radius a of avidin and biotin-bound avidin.

	Avidin	Avidin + Biotin
μ_{meas} [D]	599 ± 8	591 ± 9
μ_{calc} [D]	632^a	630^b
a_{meas} [Å]	19.4 ± 0.1	19.0 ± 0.2
a_{calc} [Å]	20.9^a	20.9^b
r_1 [Å]	15.7^a	15.7^b
r_2 [Å]	22.9^a	22.9^b
r_3 [Å]	25.3^a	25.3^b

Measurement error ranges represent 95% confidence intervals. a_{calc}, r_1, r_2, and r_3 are defined in Equations B.7-B.10. [a]Calculated using PDB 2avi with biotin removed at pH 7. [b]Calculated using PDB 2avi with biotin included at pH 7.

Table 7.2: Measured and calculated hydrodynamic radius a of avidin and biotin-bound avidin using refined PDB structures.

	Avidin	Avidin + Biotin
a_{meas} [Å]	19.4 ± 0.1	19.0 ± 0.2
a_{calc} [Å]	20.6 ± 0.4^a	19.8 ± 0.5^b
r_1 [Å]	15.6 ± 1.0^a	15.1 ± 0.8^b
r_2 [Å]	20.6 ± 2.6^a	20.1 ± 2.0^b
r_3 [Å]	27.2 ± 3.6^a	25.8 ± 3.0^b

Error ranges represent 95% confidence intervals. Values for a_{calc}, r_1, r_2, and r_3 were extracted from the last 8 ns of 10 ns molecular dynamics simulations. [a]Calculated using PDB 2avi with biotin removed. [b]Calculated using PDB 2avi with biotin included.

would explain why μ_{meas} decreases by an amount larger than the predicted amount of 2 D. This would also be consistent with the small observed decrease in measured hydrodynamic radius with the addition of biotin.

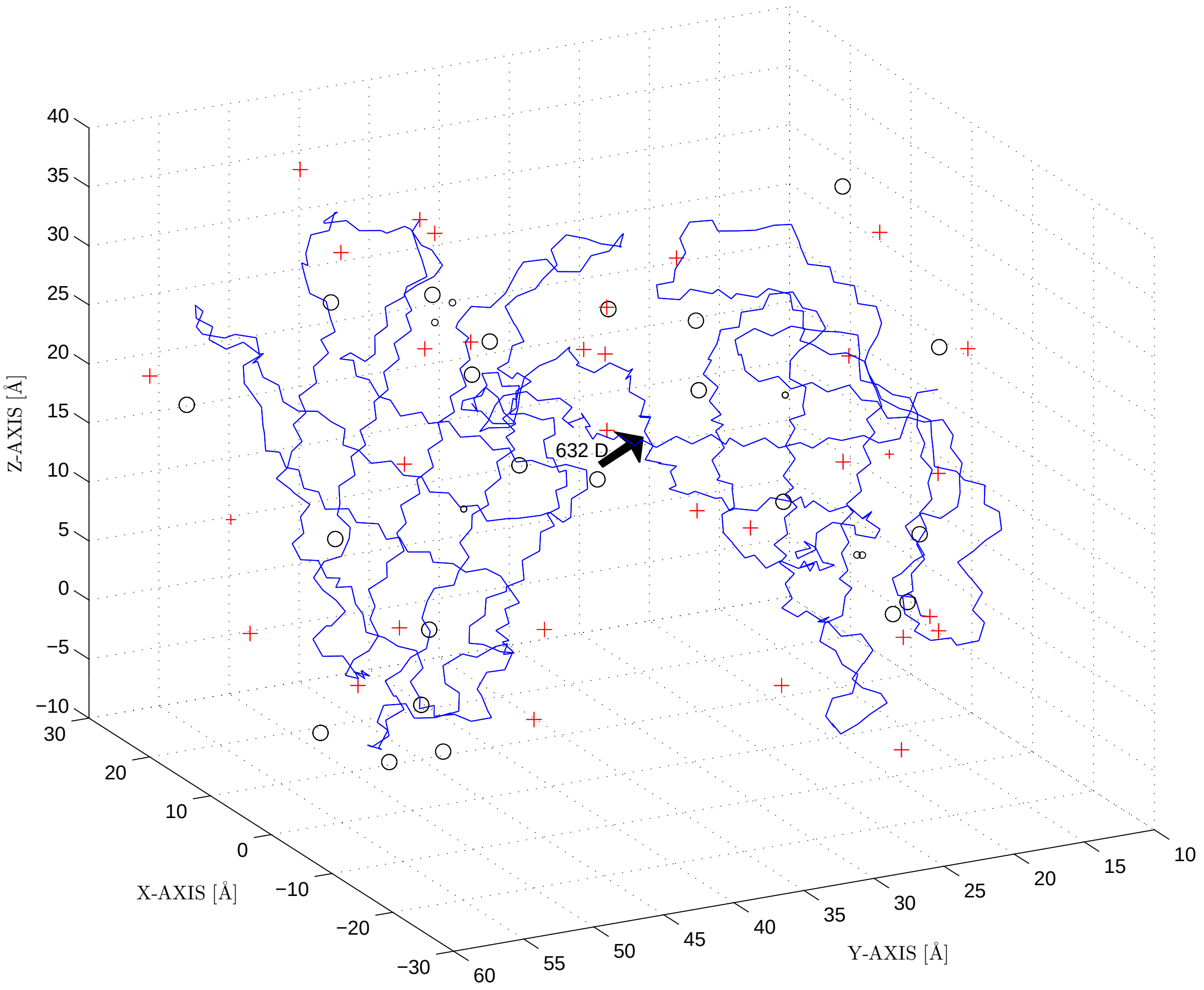


Figure 7.5: Calculated partial charges and dipole moment of avidin dimer (PDB 2avi) at pH 7. Positive charges are displayed as + and negative charges are displayed as circles, where size is proportional to strength.

7.1.4 Structure Refinement Using Molecular Dynamics Simulations

Molecular dynamics simulations were employed to examine the change in radius from the free to the biotin-bound state of avidin. Figures 7.6 and 7.7 show calculated ellipsoid radii r_1, r_2, r_3, and the equivalent volume sphere radius a_{calc} for two different 10 ns simulation runs (Appendix B). The first simulation consisted of a water cube surrounding an avidin-biotin complex. The second was a water cube surrounding free avidin (biotin molecules were deleted prior to solvation). Because both simulations used the same PDB file (2avi), their structures

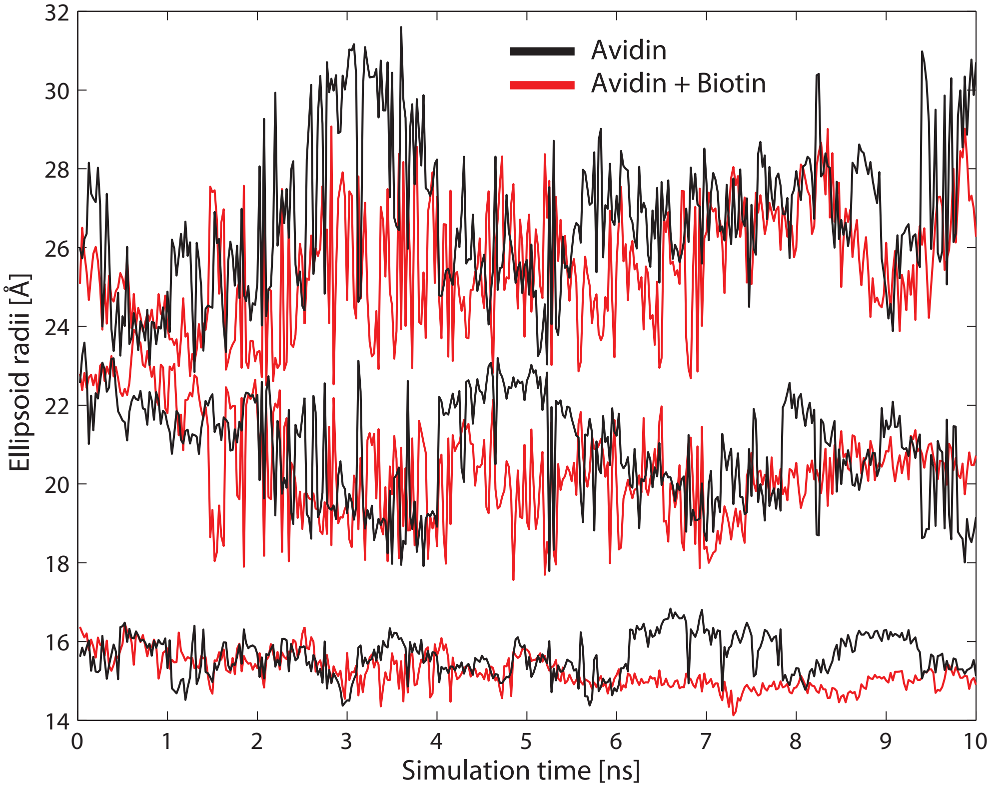


Figure 7.6: Ellipsoid radii r_1, r_2, and r_3 for a 10 ns molecular dynamics simulation of avidin and biotin-bound avidin using PDB 2avi.

are identical at the beginning of the run. The overall molecular volume was indeed smaller with the addition of biotin, confirming the measurement. This can be observed by examining a_{calc} after 2 ns of simulation time, by which time the system is roughly equilibrated. The difference in a_{calc} between the two simulations was 0.8 Å (Table 7.2), which was even larger than the measured difference of 0.4 Å. This is reconciled by simulated differences amongst r_1, r_2, and r_3, which do not decrease by proportionate amounts. In fact, the smallest radius of the ellipsoid is very similar between the two simulations. Because the dipole moment vector is partially aligned with the shortest of the three axes, we expect a change in radius smaller than the simulated 0.8 Å.

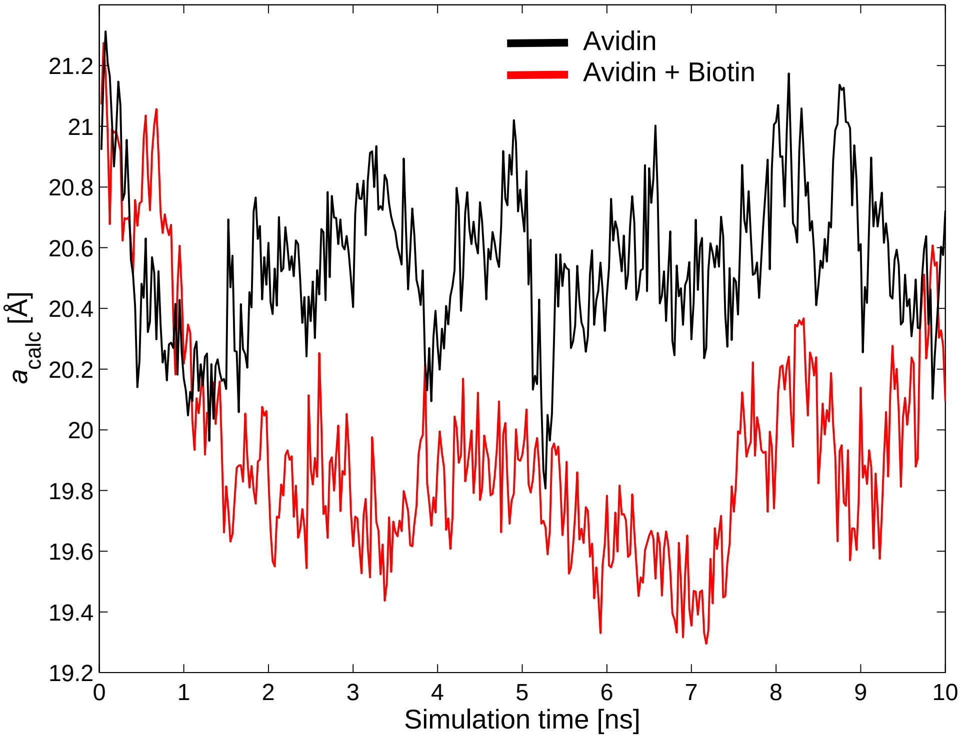


Figure 7.7: Calculated radius a_{calc} for a 10 ns molecular dynamics simulation of avidin and biotin-bound avidin using PDB 2avi.

7.2 Avidin-bBSA: Protein-Protein Interaction

Covalently attaching biotin to a protein, referred to as biotinylation, is a technique used in many areas of biotechnology. One purpose of biotinylation is to form structures that bind strongly to avidin. Dielectric spectroscopy experiments were conducted on biotin-labeled bovine serum albumin (bBSA) and avidin in the 60 μl cell to analyze the binding interaction between them.

7.2.1 Experimental Protocol

bBSA (A8549) lyophilized powder was obtained from Sigma and used without further purification. Avidin and bBSA were reconstituted in 0.1 mM HCl and stored in microcentrifuge tubes at concentrations of 20 mg/ml and 20 mg/ml, respectively. They were then refrigerated until used.

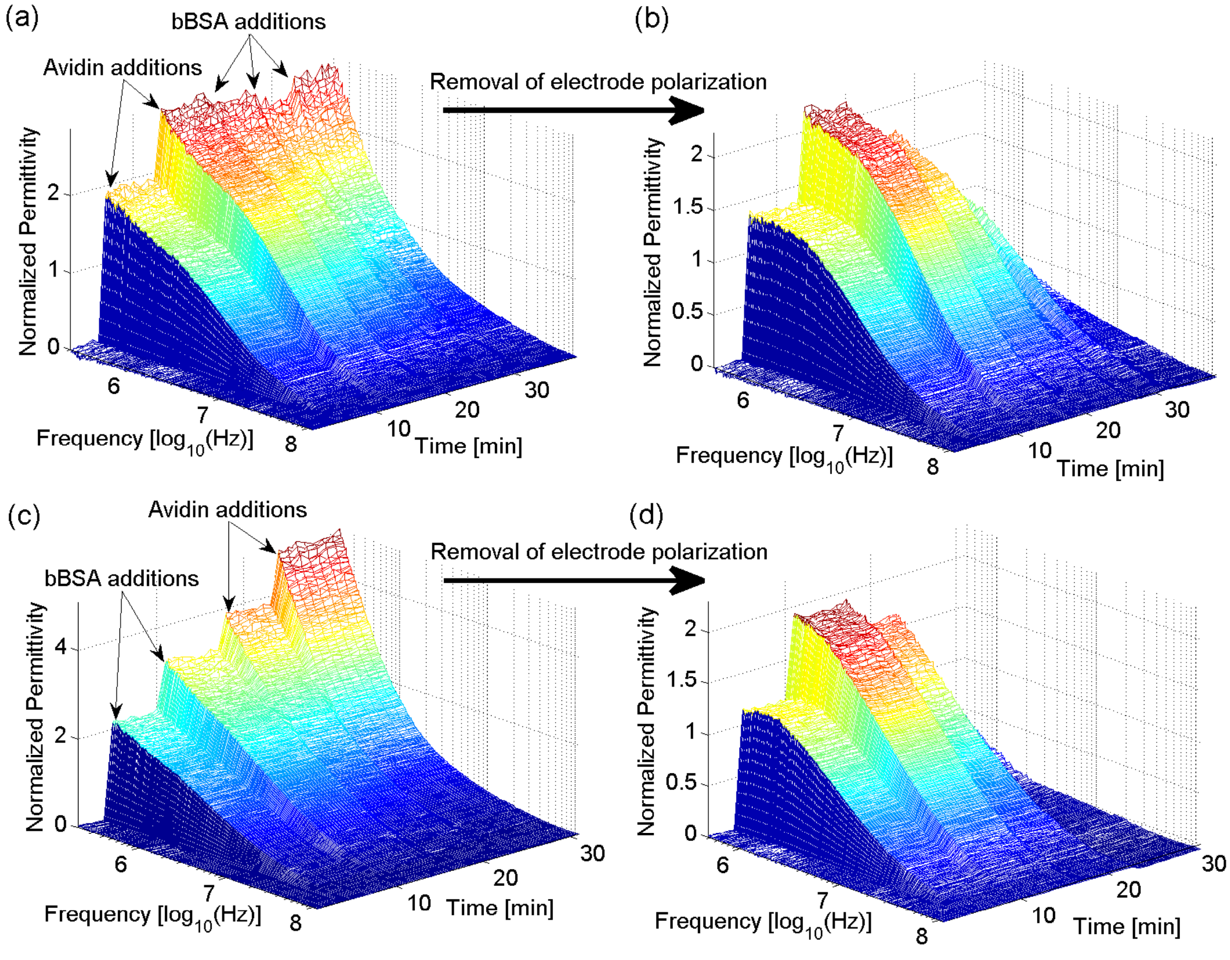


Figure 7.8: Normalized permittivity of (a,b) avidin + bBSA and (c,d) bBSA + avidin interactions. The electrode polarization contribution was subtracted from the permittivity spectra using a least-squares fitting of Equation 4.10. Plots in (b,d) clearly show that when the second protein is added, the proteins bind and the original relaxation falls off considerably.

For assessment of the reaction of avidin with bBSA, the cell was initially filled with 60 μl of 0.1 mM HCl buffer. 3 μl were withdrawn from the cell and replaced with 3 μl of the avidin solution twice. Then, bBSA was added by withdrawing 2 μl from the cell and replacing with 2 μl of the bBSA solution three times. After the cell was rinsed with ethanol and DI water and dried, the procedure was repeated starting with bBSA, followed by avidin.

7.2.2 Results and Discussion

Figure 7.8 shows the avidin-bBSA binding spectra run first with avidin, followed by bBSA (Figure 7.8a,b), then bBSA followed by avidin (Figure 7.8c,d). The aggregation of the protein is clearly visible. The first protein has a distinct relaxation that increases when the protein concentration is doubled. When the second protein is added, the original relaxation decreases in magnitude and eventually is overtaken by a larger relaxation at a lower frequency. Because the reaction occurs so quickly, it is not easily resolved in time because the diffusion of the molecules occurs on the order of minutes, even with mixing.

The data was parameterized using a least-squares fitting of Equation 4.10 for each frequency sweep. Fitted parameters $\Delta\varepsilon$ and τ are shown in Figure 7.9. Because Equation 4.10 assumes a single solute molecule, a bimodal or trimodal should be used when the complement protein is added (doing so gives a parameter fitting that lacks robustness). However we used the single relaxation model to capture the general trend of the largest relaxation. In both experiments, the relaxation time τ increased to a value between 20 and 40 ns when the first protein was added and remained relatively constant when the concentration of the first protein increased. When the second protein was added, τ increased and eventually moved outside the available frequency window of our apparatus. Thus the general size of the avidin-bBSA aggregate is inaccessible through dielectric spectroscopy, except that it must have a relaxation time of at least 150 ns and an an effective hydrodynamic radius of ≥ 38 Å.

Dynamic light scattering (90 Plus particle sizer, Brookhaven Instrument Corp.) measurements were performed at 25 °C on the resulting avidin-bBSA solution to resolve the size of the aggregate. The mean particle diameter in the solution was 2.6 kÅ and the half-height width of the distribution was 1.3 kÅ when fitting to a lognormal distribution. This reveals a very broad range of large sizes for avidin-bBSA aggregates, and confirms the interaction probed by dielectric spectroscopy. The mean particle diameter for only avidin or bBSA was below the range of the instrument (< 5 nm).

7.3 Quantitation of pH-Induced Aggregation in Binary Protein Mixtures

Protein self-assembly and aggregation are active research topics that receive significant attention [235]. Much of the attention is focused on the study of diseases involving

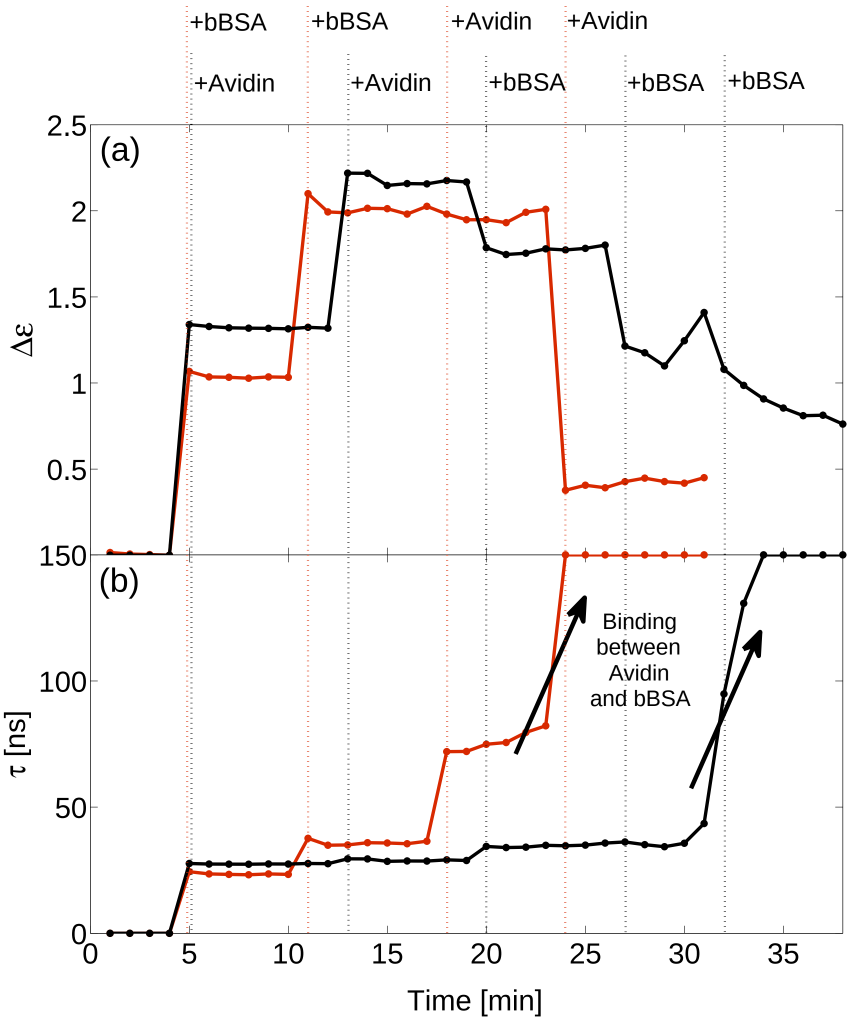


Figure 7.9: (a) Change in permittivity $\Delta\varepsilon$ and (b) relaxation time τ resulting from least-squares fittings of avidin-bBSA interaction using Equation 4.10. Red solid line corresponds to experiment where bBSA is added first followed by avidin and black solid line corresponds to experiment where avidin is added first followed by bBSA. Dotted lines indicate the time points corresponding to protein additions. Protein binding is evident in the lower plot at the time traces where the relaxation times increase beyond 150 ns, where a relaxation time of 150 ns corresponds to a molecular radius of ~ 38 Å (Equation 4.5).

amyloid formation, such as Alzheimer's and Parkinson's diseases [43, 44], in which protein aggregation is viewed as a serious, difficult-to-reverse problem. Conversely, research into benefits from protein aggregation is underway in the material sciences [236] and food industry [237]. Proteins can be treated as nanoscale building blocks which, under certain conditions, are capable of assembly into large ordered structures on the nanometer scale [238].

One of the more common approaches to detect aggregation in liquids involves an optical turbidity measurement, which describes the opacity (or fogginess) of a liquid by quantifying the attenuation of light passing through it. Turbidity is a simple, low resolution method where the signal is related to the density of aggregates. The downside of turbidity is that it is difficult to make deductions about the physical characteristics of the aggregates or the percentage of total protein molecules involved. Other biophysical techniques probe the formation of aggregates by measuring a signal related to molecular shape, size, or conformation, as determined through various microscopy [239, 240, 241], spectroscopy [242, 243, 244], or light scattering methods [245, 246]. Whether for monitoring amyloid growth or evaluating the quality of biopharmaceuticals [247], the task of detecting and measuring protein aggregation continues to be challenging, which in turn has fueled the recent development of a variety of new measurement techniques [248, 249, 250].

The purpose of this section is to demonstrate the application of dielectric spectroscopy to the problem of protein aggregation detection and characterization. Dielectric relaxations of dissolved biomolecules are strongly influenced by molecular shape and size and are thus sensitive to dramatic conformational changes like those stemming from self-assembly. Interactions between the proteins β-Lg and hen-lysozyme (HENL) and between the proteins bovine serum albumin (BSA) and HENL are examined in this section. Because of their historical use in biophysical studies, many of the basic charge properties of these proteins have previously been studied (Table 7.3). In solution, HENL reportedly aggregates with β-Lg between pH 4-11 [22] and with BSA between pH 7.5-10.5 [239], approximately. Interactions between these proteins can thus be triggered or suppressed by manipulating the solution pH as will be demonstrated here.

Table 7.3: Molecular weights (M_W), isoelectric points (pI), and measured dipole moments (μ) of the proteins HENL, β-Lg, and BSA.

		HENL	β-Lg	BSA
M_W [kDa]		14.3	18.4 (monomer) 36.8 (dimer)	66.5
pI		11.2[a]	5.1[b]	5.1[c]
μ [D]	pH 2.5	—	550[f]	—
	pH 6.5	210[d], 270[e]	800[f]	220[g]
	pH 9	—	750[f]	—

[a][251], [b][172], [c][252], [d][224], [e][98], [f][113], [g][253]

7.3.1 Protein Preparation

Lyophilized protein powders of HENL (L6876), β-Lg (L3908), and BSA (A3059) were purchased from Sigma-Aldrich and used without further purification. Each powder was dissolved in DI water at a concentration of 10 mg/ml and stored in microcentrifuge tubes and refrigerated until used in experiments. Protein concentrations were verified using a NanoDrop ND-1000 Spectrophotometer at a wavelength of 280 nm.

7.3.2 Dielectric Cell Calibration and Measurement

Measurements were made in a 60 μl parallel-plate dielectric cell connected to an Agilent 4294A impedance analyzer through an 16047E test fixture as described in the previous section. For each set of measurements, the cell was linearly calibrated using air-water capacitance measurements to establish the cell constant and parasitic capacitance in order to calculate ε_M [254]. The measurement cell was filled with DI water when not in use to avoid delamination of the polymer coating from the stainless steel [87]. Measurements were performed at room temperature.

7.3.3 Turbidity Measurements

A turbidity value, T, was calculated from an absorbance measurement using the formula $T = 2.3A_{600}/l$, where A_{600} is the absorbance measured at 600 nm and l is the path length [239]. Turbidity was measured using a Thermo Fisher Scientific Genesys 20 Spectrophotometer 15 minutes after the proteins were mixed.

7.3.4 Experimental Protocol

Two experiments were performed to demonstrate the pH-dependent binding properties of the protein mixtures. The first experiment established binding between β-Lg and HENL by adding $\sim$ 1 mg/ml of HENL into a solution already containing $\sim$ 3 mg/ml of β-Lg. This was done in DI water and repeated in a 5 mM HCl solution. The second experiment demonstrated binding between BSA and HENL by adding $\sim$ 1 mg/ml of HENL into a solution already containing $\sim$ 3 mg/ml of BSA. This was done in DI water and repeated in a 1 mM NaOH solution. The pH of each protein solution was measured using a 3 mm diameter pH electrode (Accumet MicroProbe) connected to a pH and conductivity meter (Oakton PC 510). Solutions were added and removed from the measurement cell using pipettors.

7.3.5 Interpretation of Permittivity Measurements

The increase in the permittivity spectra associated with the tumbling motion of the protein, the β-relaxation, is directly proportional to the concentration of protein molecules, c_{free}, responding to the applied alternating electric field [47]. This can be expressed mathematically by a modification of the Cole-Cole equation,

$$\varepsilon_S = \varepsilon_\infty + \frac{\delta c_{\text{free}}}{1 + (j\omega\tau)^{1-\alpha}}. \tag{7.1}$$

The Cole-Cole equation is typically expressed with $\Delta\varepsilon$ in the numerator (Equation 2.60), thus, usage of Equation 7.1 assumes c_{free} is sufficiently small so that protein-protein effects are negligible at low protein concentrations [255, 224].

The low-frequency permittivity of a low-concentration binary protein mixture has dielectric contributions from both proteins as well as the contribution from any aggregate structures. This is given by

$$\varepsilon_S = \varepsilon_\infty + \frac{\delta_1 c_{1,\text{free}}}{1 + (j\omega\tau_1)^{1-\alpha_1}} + \frac{\delta_2 c_{2,\text{free}}}{1 + (j\omega\tau_2)^{1-\alpha_2}} + \frac{\delta_A c_A}{1 + (j\omega\tau_A)^{1-\alpha_A}} \tag{7.2}$$

where subscripts 1, 2, and A correspond to the first protein, second protein, and aggregate structures, respectively. The total number of parameters can be reduced by narrowing the

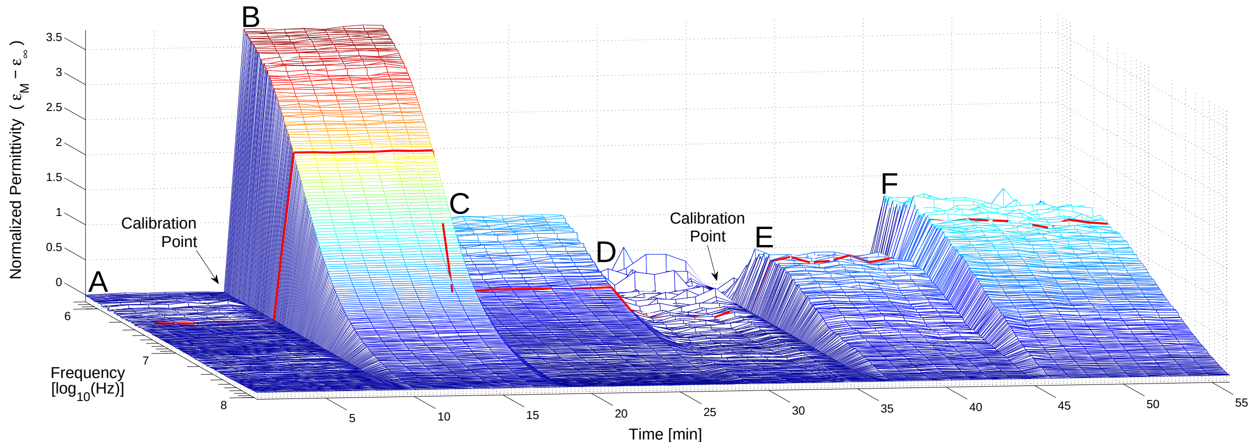


Figure 7.10: Normalized permittivity $(\varepsilon_M - \varepsilon_\infty)$ for the β-Lg/HENL experiment performed in DI water (A-C) and repeated in 5 mM HCl (D-F). (A) Cell initially filled with 60 μl of DI water. (B) 18 μl β-Lg addition. (C) 6 μl HENL addition. (D) Cell evacuated and filled with 60 μl of 5 mM HCl. (E) 18 μl β-Lg addition. (F) 6 μl HENL addition. The solid red line is the normalized permittivity at 3 MHz.

frequency range and exploiting the fact that τ_A is generally much larger than τ_1 and τ_2. For spheres of radius r, $\tau \propto r^3$ and therefore large aggregates will have significantly shifted relaxation times [62]. Thus we expect τ_A for microspheres [118] to be much larger than the individual proteins. Employing this assumption, Equation 7.2 can be expressed as a single-frequency permittivity (at ω_0) by

$$\varepsilon_S(\omega_0) \approx \varepsilon_\infty + \delta_1 c_{1,\text{free}} + \delta_2 c_{2,\text{free}} \qquad \text{for} \quad \frac{1}{\tau_A} \ll \omega_0 < \frac{1}{\tau_1}, \frac{1}{\tau_2}. \tag{7.3}$$

Because δ_1 and δ_2 are constants, single-frequency permittivity measurements can be used to approximate the unbound protein concentrations in solution. Under these conditions, when $\varepsilon_S(\omega_0) = \varepsilon_\infty$ in Equation 7.3, both $c_{1,\text{free}}$ and $c_{2,\text{free}}$ must be zero. Likewise, c_A must be zero if ε_S is equal to the sum of the dielectric contributions of the individual proteins when isolated in solution.

7.4 Results and Discussion

Normalized permittivity $(\varepsilon_M - \varepsilon_\infty)$ measurements, which fix ε_∞ to the value of ε_M at 110 MHz, for the β-Lg/HENL and BSA/HENL experiments are plotted in Figures 7.10 and

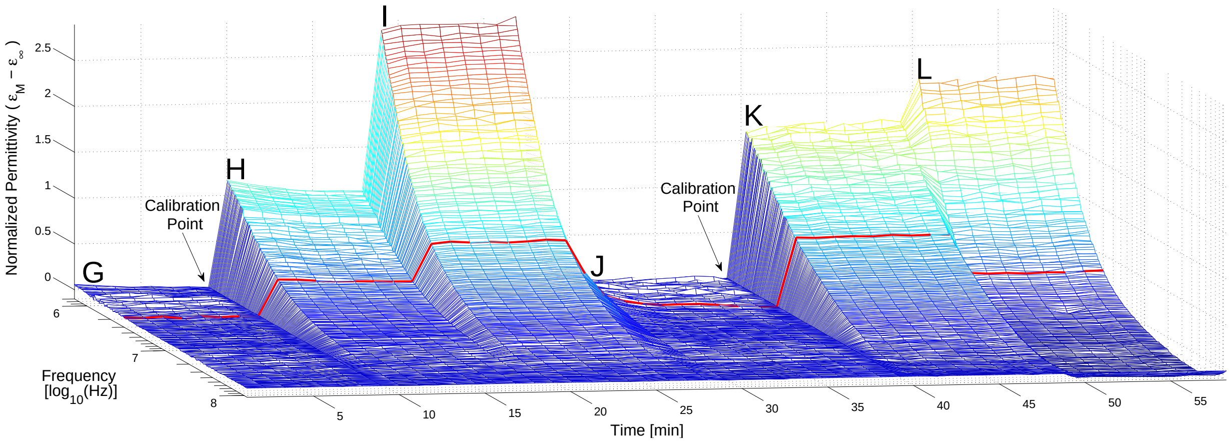


Figure 7.11: Normalized permittivity ($\varepsilon_M - \varepsilon_\infty$) for the BSA/HENL experiment performed in DI water (G-I) and repeated in 1 mM NaOH (J-L). (G) Cell initially filled with 60 μl of DI water. (H) 18 μl BSA addition. (I) 6 μl HENL addition. (J) Cell evacuated and filled with 60 μl of 1 mM NaOH. (K) 18 μl BSA addition. (L) 6 μl HENL addition. The solid red line is the normalized permittivity at 3 MHz.

7.11. Experimental steps are labeled in the figures with alphabetical letters for convenience. At step B, β-Lg is added into DI water and the protein's large β-relaxation is visible. At step C, HENL is added which causes the total dielectric increment to decrease. This procedure is repeated in 5 mM HCl at steps E and F. The β-relaxation is much larger at step B than at step E because the dipole moment of β-Lg decreases at low pH [113]. In the β-Lg/HENL experiment, it is visually evident that the proteins interact very differently depending on the solution pH (if not, steps A-C would be identical to steps D-F). In DI water the total dielectric increment decreases when HENL is added to β-Lg (step B to C) yet increases if the same proteins are in 5 mM HCl (step E to F). From this we deduce that a portion of free β-Lg molecules ($c_{1,\text{free}}$) must become aggregated β-Lg/HENL (c_A) at step C, and very little free β-Lg become aggregated at step F, i.e., β-Lg and HENL aggregate in DI water but do not in 5 mM HCl. In the BSA/HENL experiment (steps G to L) it is also apparent that the proteins interact differently at different pH. BSA, in contrast to β-Lg, does not aggregate with HENL in DI water (increase in permittivity from step H to I) but does aggregate in the 1 mM NaOH solution (decrease in permittivity from step K to L at frequencies where electrode polarization is negligible).

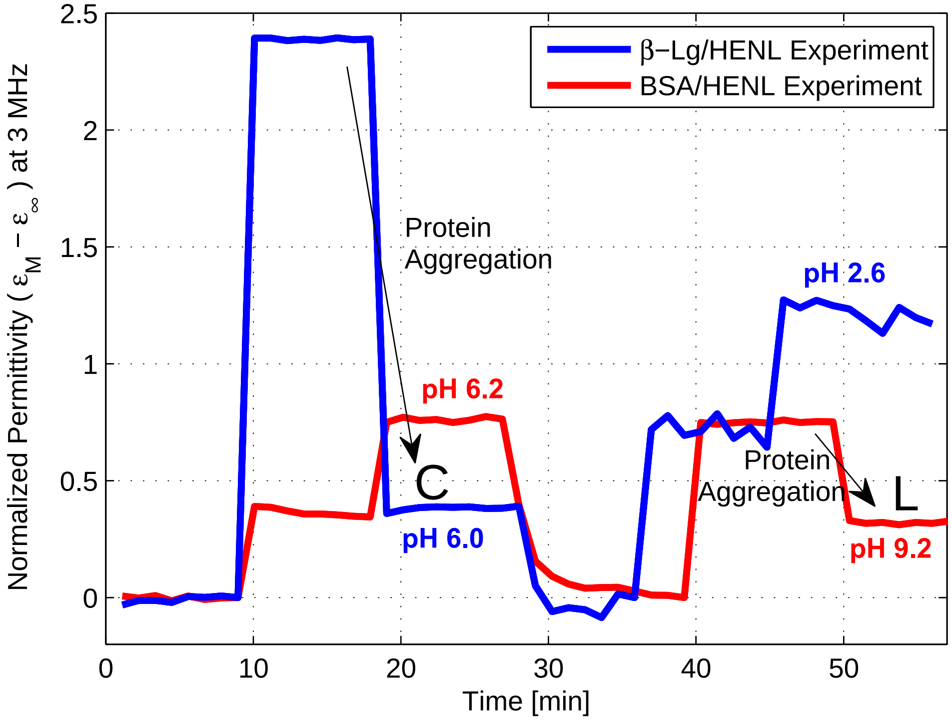


Figure 7.12: Normalized permittivity at 3 MHz for the β-Lg/HENL and BSA/HENL experiments (compare to Figures 7.10 and 7.11).

A lower bound on the percentage of $c_{1,\text{free}}$ that convert into c_A when the second protein is added can be calculated from single-frequency normalized permittivity values from the spectra in Figures 7.10 and 7.11. This is only possible under the assumptions that the frequency used is (1) much higher than the relaxation frequency of the aggregate structures, i.e., $\omega \gg 1/\tau_A$, and (2) above the onset frequency of electrode polarization effects. At 3 MHz both assumptions are appropriate, the former assumption being supported by Debye's equation which relates the relaxation time constant to the molecular size by $\tau \propto r^3$ [62]. Thus, we expect the tumbling time-constant of the aggregates to be orders of magnitude larger than for the individual molecules. Figure 7.12 shows the normalized permittivity at 3 MHz from the data in Figures 7.10 and 7.11. In the two cases where aggregation prominently occurs (denoted with arrows), corresponding drops in normalized permittivity to $\sim 16\%$ and $\sim 43\%$ of their original values were recorded for β-Lg/HENL and BSA/HENL, respectively. Therefore, by Equation 7.3, the percentage of proteins involved in aggregation must be greater than $\sim 83\%$ for β-Lg and greater than $\sim 48\%$ for BSA (values calculated accounting

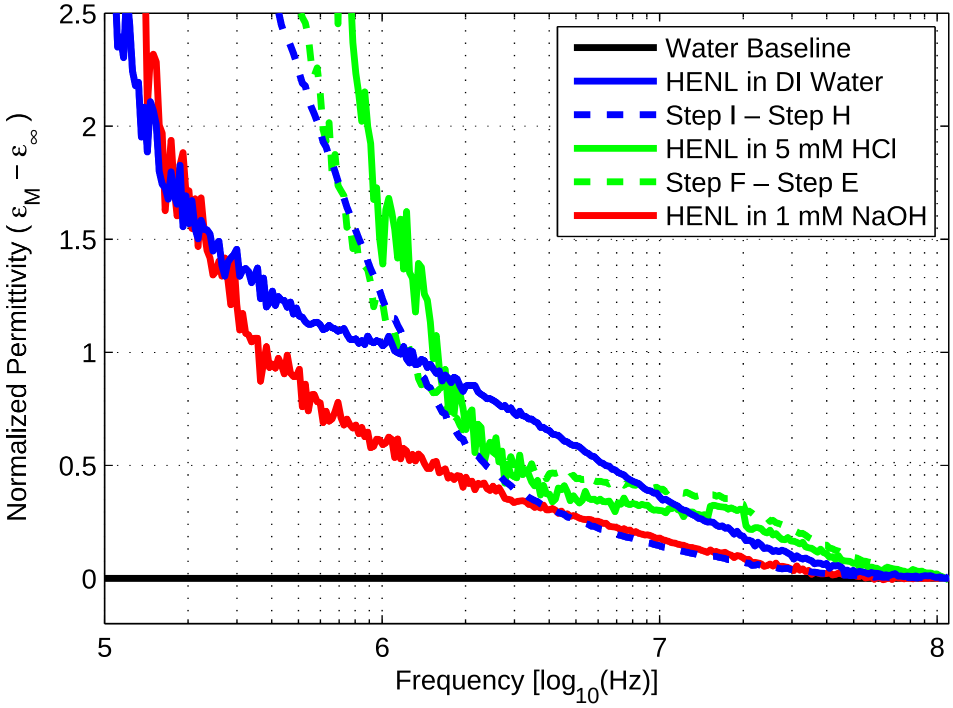


Figure 7.13: Normalized permittivity of 1 mg/ml HENL in different solutions and subtracted spectra from Figures 7.10 and 7.11.

for protein dilution). After the permittivity decreases at steps C and L, the permittivity values represent a combination of the remaining free protein contributions $c_{1,\text{free}}$ and $c_{2,\text{free}}$. In the ideal case, a least-squares fitting of the permittivity measurements to Equation 7.2 could parameterize $c_{1,\text{free}}$ and $c_{2,\text{free}}$. However, such a large number of parameters to be fitted and significant overlap of the relaxation curves makes this very difficult, particularly in the case when HENL is involved because its dielectric relaxation is notably broad (Reference [224] and Figure 7.13).

Permittivity contributions of 1 mg/ml HENL in different pH solutions are shown in Figure 7.13. The protein signal is clearest in DI water due to smaller electrode polarization effects and larger dielectric increment. A comparison of this signal (blue solid line) to the transition from step H to I (blue dashed line) indicates that the full contribution of the added HENL to the protein's β-relaxation does not occur. This suggests that a portion of the added HENL are still aggregating even though the normalized permittivity does not decrease from step H to I. That these β-Lg/HENL aggregates at pH 6.2 were not detected

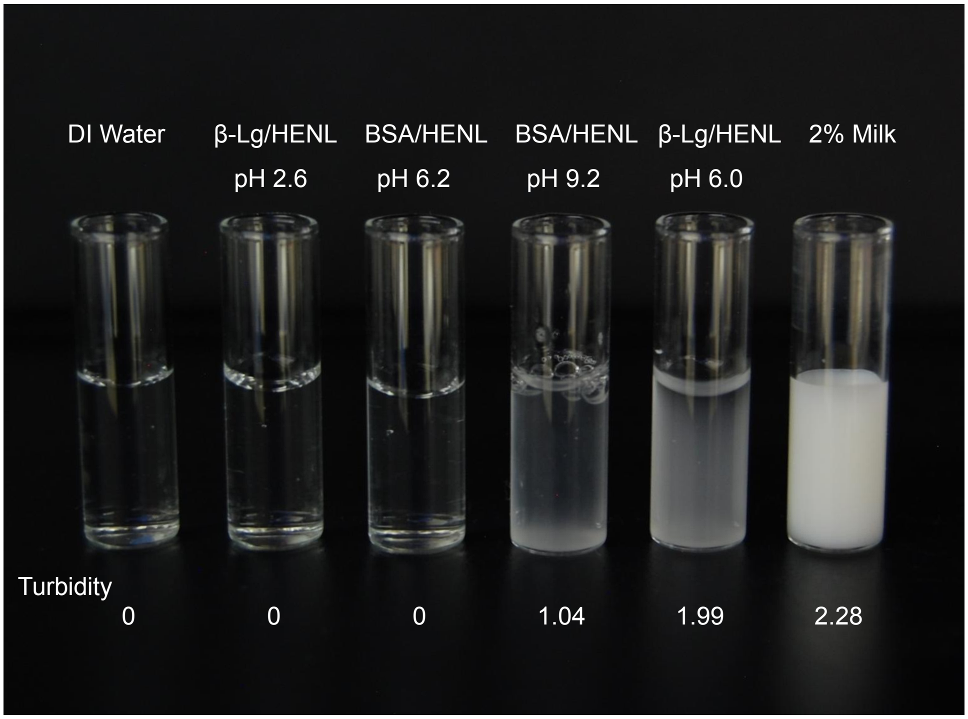


Figure 7.14: Photograph of protein solutions alongside some reference liquids. For comparison to Figures 7.10 and 7.11, the middle four solutions correspond to steps F, I, L, and C (from left to right). Turbidity was measured at 600 nm and is in units of cm^{-1}. Note that 2% milk refers to milk with approximately 2% fat content (BYU Creamery).

by turbidity (Reference [239] and Figure 7.14) nor optical microscopy [239] is probably due to their size being below the threshold of both absorbance measurements at a wavelength of 600 nm and optical microscopy measurements ($\sim$ 250 nm). A comparison of 1 mg/ml HENL in 5 mM HCl (green solid line) to the transition from step E to F (green dashed line) shows the additive nature of permittivity in the absence of interaction. Differences between these lines can be attributed to differences in electrode polarization which is significant in 5 mM HCl. By taking the 3 MHz normalized permittivity values of HENL isolated in solution and using the technique detailed in the previous paragraph, involvment of HENL must be greater than $\sim$ 42% in the β-Lg aggregation process at pH 6.0.

The deductions made using the permittivity measurements were confirmed with turbidity measurements. A photograph of the resulting protein mixures is shown in Figure 7.14 with measured pH and turbidity. Reference solutions of DI water and 2% milk (BYU

Creamery) were also included for comparison. The other four solutions correspond to steps F, I, L, and C (from left to right). Turbidity values of the L and C solutions are consistent with the two cases where decreases in permittivity measurements indicated aggregation. The turbidity of step C has a turbidity value approximately twice that of step L, which is qualitatively consistent with the larger percentage of proteins involved in aggregation for β-Lg compared to BSA.

Dielectric spectroscopy produces a measurement that is proportional to the free protein concentration and can be used to complement other approaches that are sensitive to the aggregates themselves. The downsides of using dielectric spectroscopy are the inherent limits in solution conductivity due to electrode polarization and peptide concentrations which, for average dipole moments, must be on the order of mg/ml. Dielectric spectroscopy also boasts incredible time-resolution between measurements, on the order of 10 frequency sweeps per second. While measurements such as turbidity can be easier to interpret, dielectric spectroscopy yields more information about time-resolved electrical behavior of the aggregation process and, given a larger measurement bandwidth and a proper fitting model, information about the size and shape of the aggregates.

Studying aggregation through the electrical charge properties of the molecules has important implications for understanding underlying aggregation mechanisms. Electric dipole moments and overall charge have both been linked to protein aggregation rates [256, 111, 257]. It seems that in the case of β-Lg and HENL, the mechanism for aggregation is the overall charge of the molecule. This is particularly true for the lower ionic strengths in this study which would reduce electrostatic shielding. Simulated titration curves using Poisson-Boltzmann estimates of the charges (Figure 7.15) show the proteins are oppositely charged over the range of pH 5-11 which coincides closely with experimentally measured non-zero turbidity [22]. BSA and HENL aggregation occurs in a pH range where the molecules are oppositely charged. Desfougeres et al. claim that electric neutrality of aggregates composed of 3 to 1 (HENL to BSA) molar ratios is an important condition for self assembly [239]. This explanation would resolve the question as to why the pH region of aggregation does not extend down to the isoelectric point of BSA.

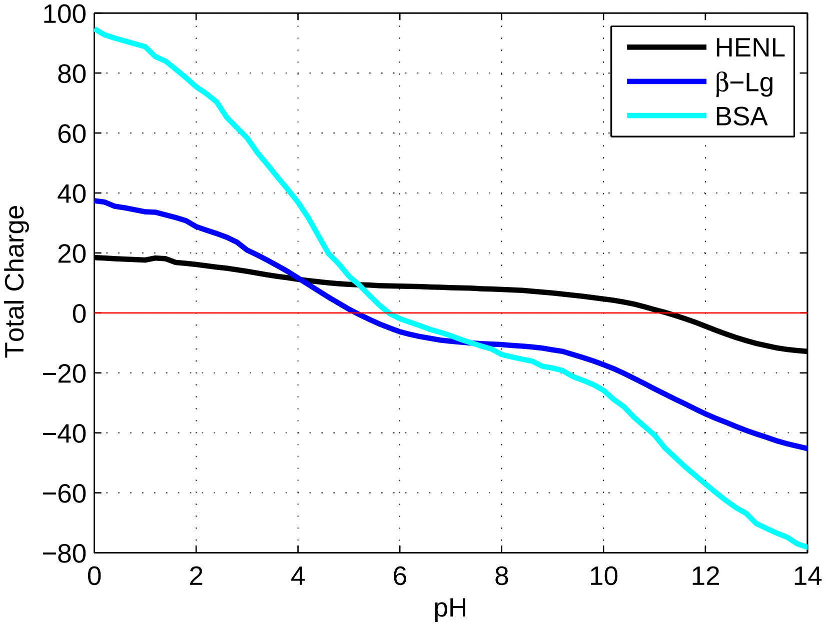


Figure 7.15: Simulated protein titration curves using H++ server [19]. Simulated isoelectric points are close to experimental isoelectric points (Table 7.3). Protein data bank files used were 1e8l (HENL), 1beb (β-Lg), and 3v03 (BSA).

Chapter 8

Conclusion

The recent explosion of new physical techniques over the last few decades is changing the way researchers approach biological problems [258, 259]. While most biophysical methods extract a signal which is reflective of molecular shape, size, or conformation (such as the various microscopy [240, 241], spectroscopy [243, 244], or light scattering methods [245, 246]), dielectric spectroscopy boasts a very different kind of measurement, one that is related to the electric dipole moment of the molecule. This unique perspective for analyzing molecular phenomena will make dielectric spectroscopy an actively pursued and attractive technique for years to come [36]. Dielectric relaxation measurements can provide a wealth of information about the structure, dynamics, and function of biomolecules and biological processes. This dissertation builds on that vision and in the process outlines some avenues in which dielectric techniques could be employed in various technological platforms.

8.1 Accomplishments and Summary

1. The research outlined in this dissertation advances the theoretical understanding and practical measurement of protein electrical properties. For decades partial charges have been known to exist, however quantitative relationships between these charges and a dipole moment have been inconsistent with measurements. The method to calculate protein dipole moments described in Chapter 5 gives a dipole moment that is consistent with that measured experimentally. This method was valuable for verification of the molecular interaction studies of Chapter 7, and will continue to be valuable for researchers interested in how charge influences other molecular properties. The accuracy to which the pH-varying dipole moment of β-Lg was predicted in Figure 5.5 is

a significant step in the literature and has already generated significant interest from other researchers in the community [260, 261, 262, 263, 264].

2. Several designs of dielectric cells were produced during this dissertation. The first cell was a general-purpose spectrometer with temperature stability (Section 2.3.3). Using this cell, the influence of temperature, pH, and concentration on the dielectric relaxation of β-Lg was measured and verified with theory. A design consisting of thermally bonded acrylic plates was used to measure the response of liquid metal electrodes (Section 3.3). Next, a cell design to accommodate larger diameter electrodes without temperature stability was used for the surface preparation experiments in Section 3.4. As the proteins being used became more expensive, it was important to redesign the cells to decrease the interrogated volume from 800 μl to 60 μl. This was the motivation behind the low-volume cell of Section 7.1.1. The electrode spacing d in the 60 μl design can be increased by a factor of four, resulting in the U-cell design (Appendix G). Each of the five cell designs are tailored specifically for different types of dielectric spectroscopy experiments.

3. The influence of electrode polarization in dielectric measurements is now more clearly understood. The origin of the $\omega^{-\text{const.}}$ (const. > 1) dependency in the real-valued permittivity spectrum was derived in Equation 3.29. This derivation enables a fitting of the electrode polarization contribution so that it can be subtracted it off. Equation 3.29 also illuminated which parameters decrease electrode polarization, i.e., smaller A, larger d, smaller κ, larger ν, and smaller σ. Using this direction one arrives at the PPy/PSS polymer as the current best electrode surface preparation technique. Electrodeposition of this polymer results in a stable, biocompatible electrode with a strong resistance to electrode polarization.

4. Molecular dynamics are the state-of-the-art simulation method for molecular systems. In Chapter 6, the relationship between dielectric studies of proteins and molecular dynamics was explored. A novel technique was introduced to estimate relaxation parameters $\Delta\varepsilon$ and τ from protein simulations in CHARMM. The technique seems to produce relaxation parameters on the order of those measured experimentally; however,

much longer simulation lengths will be needed before it can be determined whether the method actually converges to experimental values. It was demonstrated in Section 6.5 how molecular dynamics can improve the structure-based dipole moment prediction from Chapter 5. In Section 7.1, molecular dynamics were used to estimate the hydrodynamic radius of avidin in the bound and unbound state.

5. In Chapter 7, various molecular interactions were studied using the techniques introduced in the previous 6 chapters. Each of those experiments represent significant milestones, and many were the first dielectric experiments performed of their kind. In Section 7.1 the first before and after parameterization of a protein binding with a ligand demonstrated small shifts of the β-relaxation. The detection of pH-induced protein aggregation of β-Lg/HENL and BSA/HENL is also novel (Section 7.3). The techniques developed in this dissertation can be used to analyze a wide variety of reactions experimentally, theoretically, and computationally.

8.2 Future Research

During the preparation of this dissertation, several research paths were left unexplored due to time restrictions and the scope of the project. Here I list ten topics not yet discussed in the literature that could have a significant impact in the field. Many of these could by themselves be well suited as graduate or undergraduate level projects.

1. The physical dimensions of the dielectric cell are not optimized. The limitation of a large A/d ratio was quantified in terms of electrode polarization, however, the new cells were built with arbitrary larger spacings. The limiting factor on the other end (small A/d ratio) is low SNR and large parasitic capacitance C_P. To find the optimal cell dimensions the volume of the sample as well as titration ability must also be taken into consideration. I believe the most efficient way of going about this problem is to begin with simulation using one of several available programs such as Ansoft Maxwell 3D or COMSOL Multiphysics.

2. When the dielectric cell was miniaturized to reduce the required solution volume, temperature stability was eliminated. To increase the sensitivity of the smaller cells,

temperature stability must be reintroduced using either the flow-through liquid method or some other technique.

3. The partial charges used in molecular dynamics simulations are independent of the residue's environment. This is tantamount to using model pK_as in place of calculated pK_as to estimate partial charges. An important endeavor is thus bridging the two techniques, updating molecular dynamics runs with current partial charge computed with shifted pK_as. This would increase the accuracy of the simulations, particularly in cases where charged groups interact strongly.

4. The electrode surfaces could be engineered further by combining surface finishing and PPy/PSS deposition. Depending on the thickness of the applied PPy/PSS layer, the macroscopic roughness of the metal could add constructively with the microscopic roughness using the polymer. Furthermore, other techniques such as chemical etching could increase roughness better than the approaches used in this dissertation.

5. The oxide layer that is formed on the surface of Galinstan is removed at low pH. An interesting extension to the experiments performed in Section 3.3 would be to remove the oxide layer at low pH and then dilute to neutral pH and see if an atomically smooth surface is formed with the solution. The resulting onset frequency of electrode polarization f_{on} would then be compared to those of Table 3.1.

6. On the molecular biology side, aggregation effects of β-Lg with HENL and BSA with HENL need to be studied more systematically. The orientation of the molecules in these complexes is unknown and dielectric studies could aid interpretations made through simulation [118]. Protein docking programs like ZDOCK take two proteins and predicts the most likely binding configurations [265]. A combination of dipole moment measurements and calculations on these predicted complexes (as shown in Figure 8.1) could narrow the search for likely candidates.

7. Polarization phenomena caused by the introduction of ions and other molecular objects with net charge needs to be investigated. For example, when μl drops of NaOH are

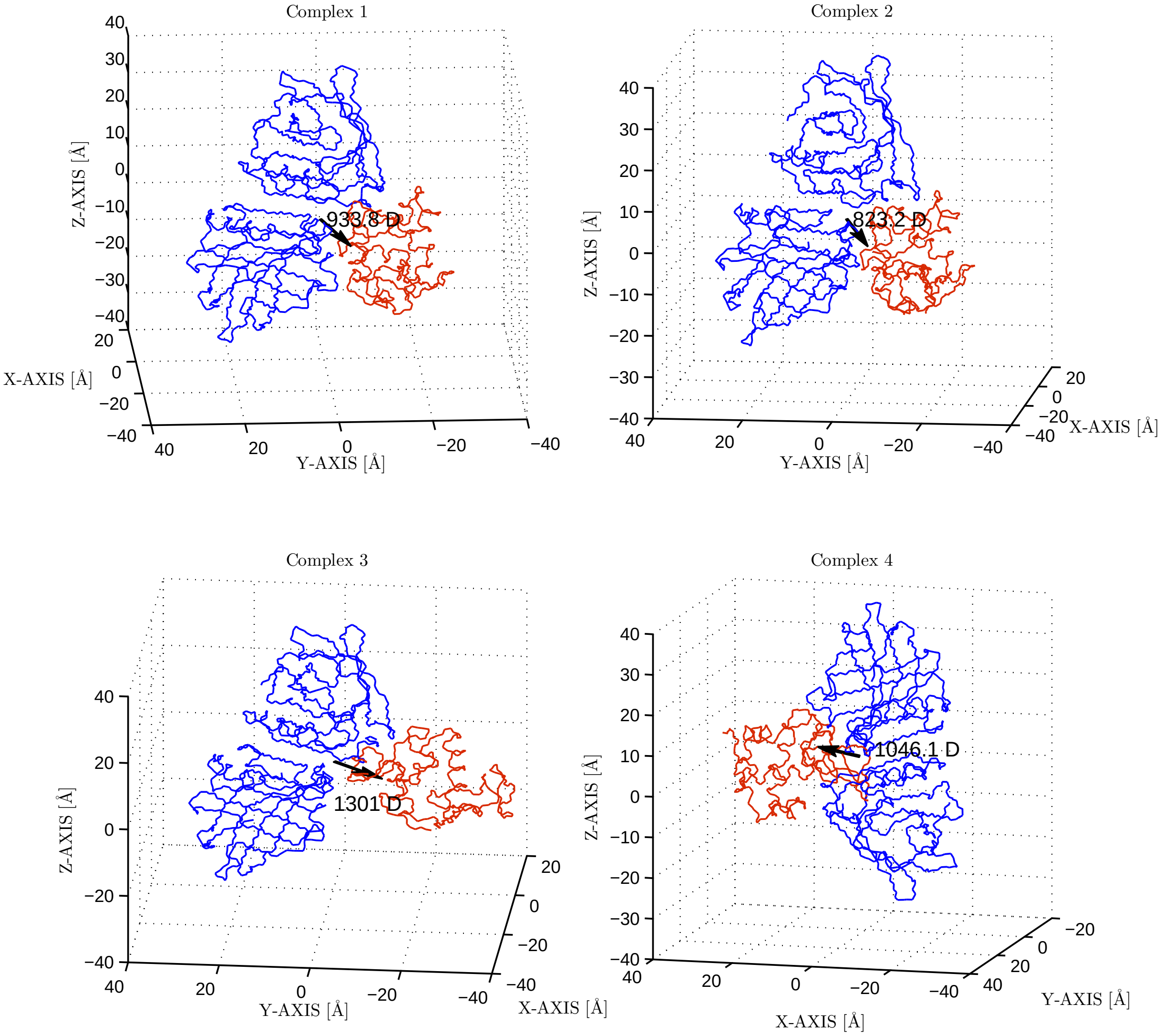


Figure 8.1: Dipole moment calculations of top four ZDOCK binding predictions for β-Lg and HENL.

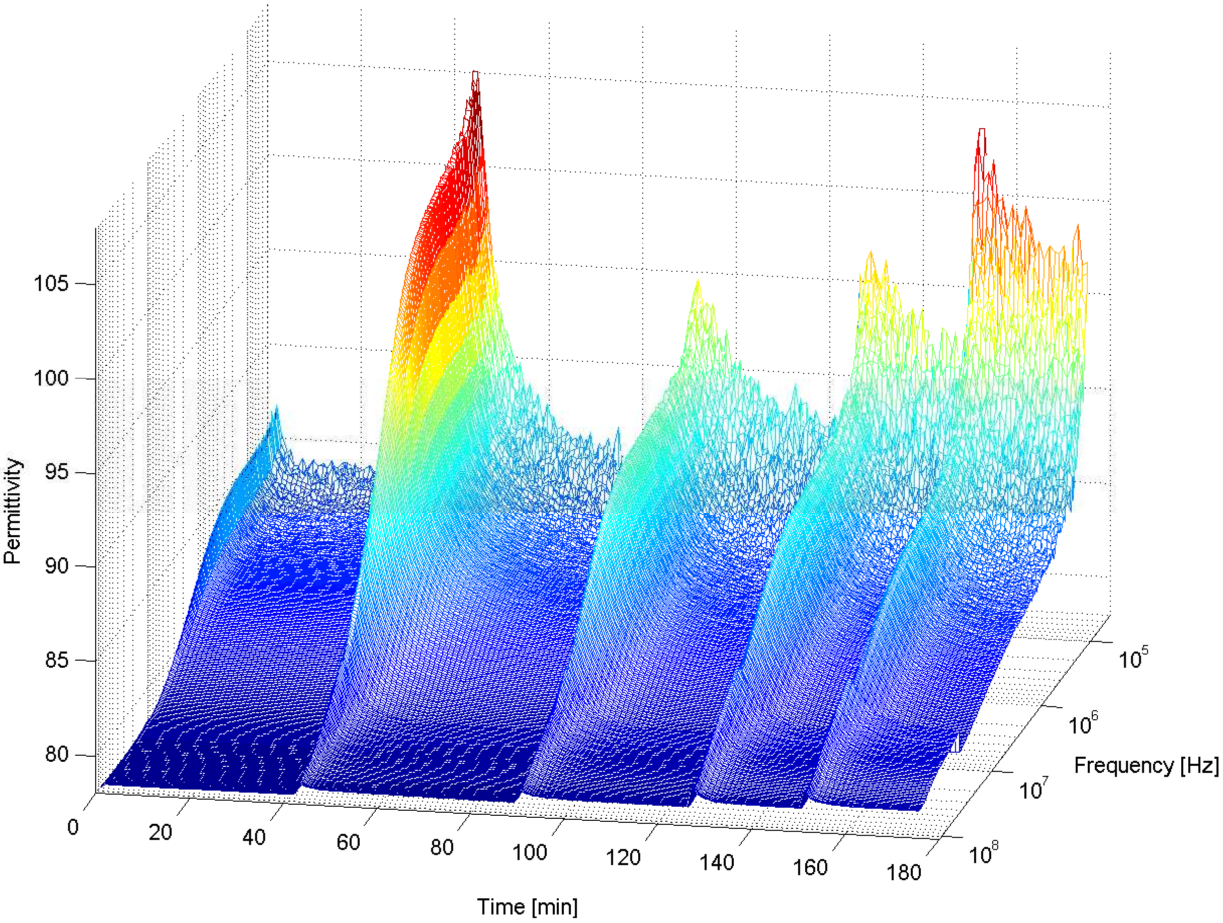


Figure 8.2: Polarization induced by 1 μl additions of 100 mM NaOH into 0.8 ml of DI water with 2 mg/ml β-Lg. A similar effect was observed with NaCl and HCl.

added to a protein solution, the permittivity can increase several times larger than the protein's $\Delta\varepsilon$ (shown in Figure 8.2).

8. The steady-state electric field simulations need to be repeated with longer simulation runs. To properly test the theory, at least 100 ns simulation times are needed. This would take several months to run on a supercomputer. Nonetheless, if the method leads to accurate predictions of $\Delta\varepsilon$, it would help validate the technique and eliminate the need for numerous costly experiments. The dielectric relaxation of any protein or molecular object could be simulated in varying conditions. This would be ideal for pre-

dicting how different protein and drug interactions influence the solution permittivity and capacitance.

9. The Kirkwood correlation parameter g (Section 4.1) was assumed to be 1 for dilute protein solutions. At higher concentrations, g could be derived both theoretically and experimentally as a function of protein concentration. The theoretical approach could calculate the average distance between molecules as a function of concentration and quantify how their electric fields influence each other. The experimental approach would consist of least-squares fits of $\Delta\varepsilon$ versus c similar to what was done in Figure 4.6.

10. Influenza (flu) is an infectious disease that results in approximately 40,000 deaths each year in the United States [266]. In the past, influenza A was treated using antiviral drugs such as amantadine, which disabled the virus by inhibiting the M2 ion channel. Over the past decade, the virus has mutated, and in 2009 the Centers for Disease Control and Prevention (CDC) announced that 100% of tested flu samples were resistant to amantadine [267]. This has caused researchers to search for alternative solutions to contain potential flu outbreaks. A drug that replaces amantadine as an M2 blocker would bind internally to the ion channel and, in the process, enhance M2 tetramerization. This assertion is based on the observation, using analytical ultracentrifugation, that amantadine enhances the tetramerization of a marginally stable peptide fragment (the transmembrane domain, residues 21-45, from the dominant pre-2005 strain) of M2 solubilized in detergent micelles [41]. The hypothesis taken here is that if dielectric spectroscopy measurements indicate that M2 tetramerizes in the presence of a certain drug, that drug will be a strong candidate in the search for an M2 blocker and amantadine replacement. The problem is well suited for dielectric spectroscopy as both the hydrodynamic radius and dipole moment should be strongly affected by tetramerization. Preliminary work is currently underway to characterize the interaction of M2 with antiviral drugs using circular dichroism (Figure 8.3). The observed increase in ellipticity ratio represents increased tetramerization and should have a parallel effect in permittivity measurements.

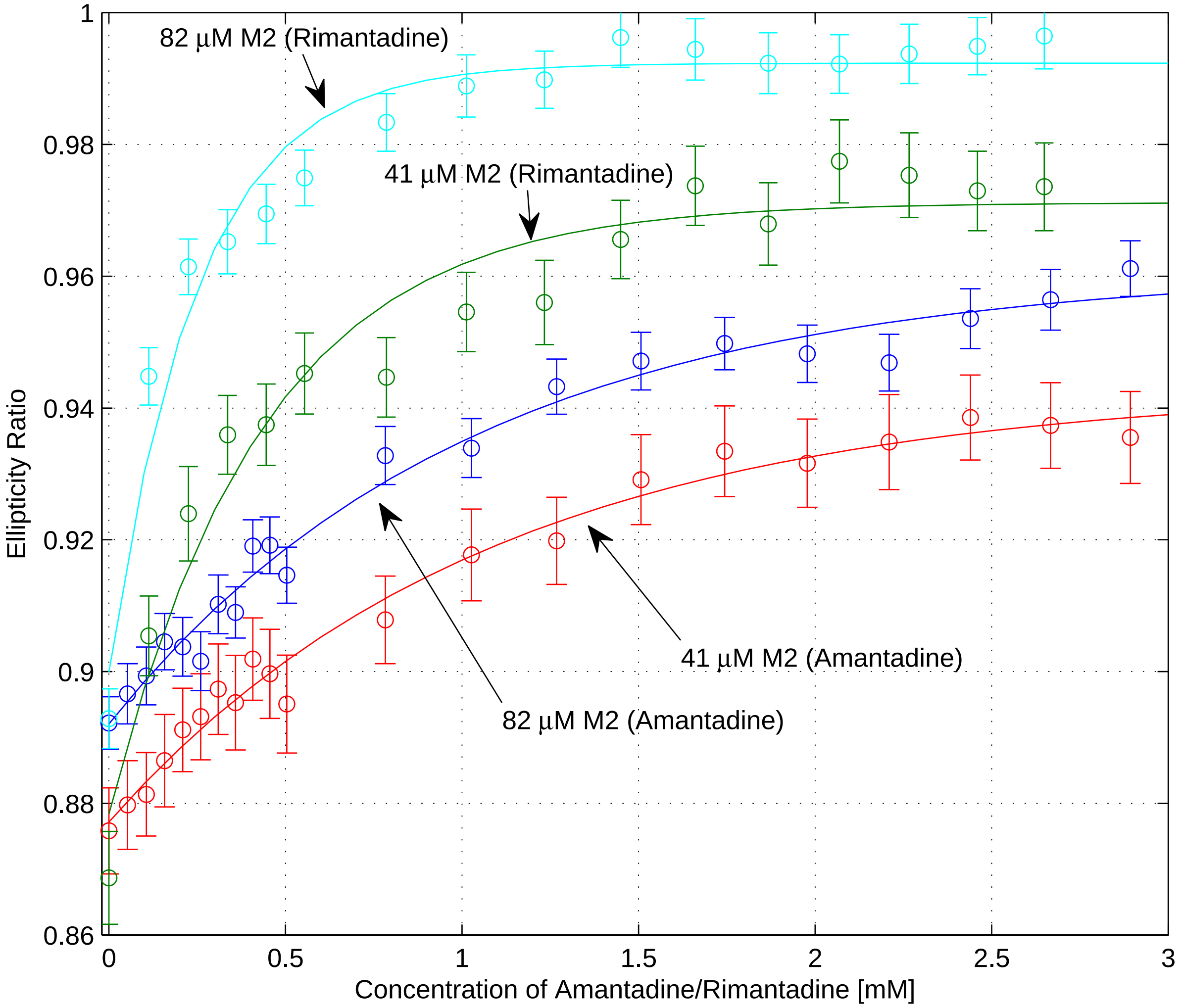


Figure 8.3: Ratio of the measured circular dichoism at 223 nm to that at 209 nm for two peptide concentrations (41 and 82 μM) and two drugs (amantadine and rimantadine). Measurements were made by Trevor Anderson.

Bibliography

[1] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace, "pK values of the ionizable groups of proteins," *Protein Science*, vol. 15, no. 5, pp. 1214–1218, May 2006. 80, 81

[2] A. Herraez, "Biomolecules in the computer: Jmol to the rescue," *Biochem. Mol. Biol. Educ.*, vol. 34, no. 4, pp. 255–261, 2006. xv, 3

[3] W. J. Ellison, K. Lamkaouchi, and J. M. Moreau, "Water: A dielectric reference," *Journal Of Molecular Liquids*, vol. 68, no. 2-3, pp. 171–279, Apr. 1996. xv, 8, 23, 24, 27, 75

[4] B. Mellor, N. Kellis, and B. Mazzeo, "Dielectric spectroscopy of molecular interactions based on the avidin-biotin complex," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 30 2011-sept. 3 2011, pp. 3660 –3663. xvi, 41

[5] S. Brownlow, J. H. M. Cabral, R. Cooper, D. R. Flower, S. J. Yewdall, I. Polikarpov, A. C. T. North, and L. Sawyer, "Bovine beta-lactoglobulin at 1.8 angstrom resolution - still an enigmatic lipocalin," *Structure*, vol. 5, no. 4, pp. 481–495, Apr. 1997. xvii, xix, 53, 76

[6] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal Of Molecular Graphics*, vol. 14, no. 1, pp. 33–&, 1996. xvii, xxi, 53, 120

[7] K. Kuwata, M. Hoshino, V. Forge, S. Era, C. A. Batt, and Y. Goto, "Solution structure and dynamics of bovine beta-lactoglobulin A," *Protein Science*, vol. 8, no. 11, pp. 2541–2545, Nov. 1999. xviii, xix, 67, 76

[8] H. Vis, M. Mariani, C. E. Vorgais, K. S. Wilson, R. Kaptein, and R. Boelens, "Solution structure of the HU protein from bacillus-stearothermophilus," *Journal Of Molecular Biology*, vol. 254, no. 4, pp. 692–703, Dec. 1995. xviii, 68

[9] S. Uhrinova, M. H. Smith, G. B. Jameson, D. Uhrin, L. Sawyer, and P. N. Barlow, "Structural changes accompanying pH-induced dissociation of the beta-lactoglobulin dimer," *Biochemistry*, vol. 39, no. 13, pp. 3565–3574, Apr. 2000. xviii, xix, 71, 72, 76

[10] J. J. Adams, B. F. Anderson, G. E. Norris, L. K. Creamer, and G. B. Jameson, "Structure of bovine beta-lactoglobulin (variant A) at very low ionic strength," *Journal Of Structural Biology*, vol. 154, no. 3, pp. 246–254, June 2006. xix, 76

[11] L. Vijayalakshmi, R. Krishna, R. Sankaranarayanan, and M. Vijayan, "An asymmetric dimer of beta-lactoglobulin in a low humidity crystal form - Structural changes that accompany partial dehydration and protein action," *Proteins-Structure Function And Bioinformatics*, vol. 71, no. 1, pp. 241–249, Apr. 2008. xix, 76

[12] B. W. Segelke, M. Forstner, M. Knapp, S. D. Trakhanov, S. Parkin, Y. M. Newhouse, H. D. Bellamy, K. H. Weisgraber, and B. Rupp, "Conformational flexibility in the apolipoprotein E amino-terminal domain structure determined from three new crystal forms: Implications for lipid binding," *Protein Science*, vol. 9, no. 5, pp. 886–897, May 2000. xix, 86

[13] G. Boissy, E. deLaFortelle, R. Kahn, J. C. Huet, G. Bricogne, J. C. Pernollet, and S. Brunie, "Crystal structure of a fungal elicitor secreted by phytophthora cryptogea, a member of a novel class of plant necrotic proteins," *Structure*, vol. 4, no. 12, pp. 1429–1439, Dec. 1996. xix, 86

[14] M. A. Hough and S. S. Hasnain, "Crystallographic structures of bovine copper-zinc superoxide dismutase reveal asymmetry in two subunits: Functionally important three and five coordinate copper sites captured in the same crystal," *Journal Of Molecular Biology*, vol. 287, no. 3, pp. 579–592, Apr. 1999. xix, 86

[15] H. Baumann, K. Paulsen, H. Kovacs, H. Berglund, A. P. H. Wright, J. A. Gustafsson, and T. Hard, "Refined solution structure of the glucocorticoid receptor DNA-binding domain," *Biochemistry*, vol. 32, no. 49, pp. 13 463–13 471, Dec. 1993. xix, 86

[16] D. W. Heinz, M. Ryan, M. P. Smith, L. H. Weaver, J. F. W. Keana, and O. H. Griffith, "Crystal structure of phosphatidylinositol-specific phospholipase C from Bacillus cereus in complex with glucosaminyl(alpha 1–6)-D-myo-inositol, an essential fragment of GPI anchors," *Biochemistry*, vol. 35, no. 29, pp. 9496–9504, July 1996. xix, 86

[17] H. M. Weir, P. J. Kraulis, C. S. Hill, A. R. C. Raine, E. D. Laue, and J. O. Thomas, "Structure of the HMG box motif in the B-domain of HMG1," *Embo Journal*, vol. 12, no. 4, pp. 1311–1319, Apr. 1993. xix, 86

[18] O. Livnah, E. A. Bayer, M. Wilchek, and J. L. Sussman, "3-dimensional structures of avidin and the avidin-biotin complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 11, pp. 5076–5080, 1993. xxi, 120

[19] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, "H++: A server for estimating pK(a)s and adding missing hydrogens to macromolecules," *Nucleic Acids Research*, vol. 33, pp. W368–W371, July 2005. xxii, 70, 141

[20] T. Simonson, "Electrostatics and dynamics of proteins," *Reports On Progress In Physics*, vol. 66, no. 5, pp. 737–787, May 2003. 1

[21] M. F. Perutz, "Electrostatic effects in proteins," *Science*, vol. 201, no. 4362, pp. 1187–1191, 1978. 1, 63

[22] N. K. Howell, N. A. Yeboah, and D. F. V. Lewis, "Studies on the electrostatic interactions of lysozyme with alpha-lactalbumin and beta-lactoglobulin," *International Journal Of Food Science And Technology*, vol. 30, no. 6, pp. 813–824, Dec. 1995. 1, 132, 140

[23] N. J. Burgoyne and R. M. Jackson, "Predicting protein interaction sites: Binding hot-spots in protein-protein and protein-ligand interfaces," *Bioinformatics*, vol. 22, no. 11, pp. 1335–1342, June 2006. 1

[24] F. Fogolari, L. Ragona, S. Licciardi, S. Romagnoli, R. Michelutti, R. Ugolini, and H. Molinari, "Electrostatic properties of bovine beta-lactoglobulin," *Proteins-Structure Function And Genetics*, vol. 39, no. 4, pp. 317–330, June 2000. 1

[25] L. A. Rosen and D. A. Saville, "Dielectric spectroscopy of colloidal dispersions: Comparisons between experiment and theory," *Langmuir*, vol. 7, no. 1, pp. 36–42, 1991. 1

[26] R. A. Pethrick and D. Hayward, "Real time dielectric relaxation studies of dynamic polymeric systems," *Progress In Polymer Science*, vol. 27, no. 9, pp. 27–28, Nov. 2002. 1

[27] C. Liu, Q. Bi, A. Leyland, and A. Matthews, "An electrochemical impedance spectroscopy study of the corrosion behaviour of PVD coated steels in 0.5 N NaCl aqueous solution: Part II. EIS interpretation of corrosion behaviour," *Corrosion Science*, vol. 45, no. 6, pp. 214–217, June 2003. 1

[28] G. Smith, A. P. Duffy, J. Shen, and C. J. Olliff, "Dielectric-relaxation spectroscopy and some applications in the pharmaceutical sciences," *Journal Of Pharmaceutical Sciences*, vol. 84, no. 9, pp. 1029–1044, Sept. 1995. 1

[29] S. Romano and P. Sorichetti, *Dielectric Spectroscopy in Biodiesel Production and Characterization*, ser. Green Energy and Technology. Springer, 2010. 1

[30] P. Metherall, D. C. Barber, R. H. Smallwood, and B. H. Brown, "Three-dimensional electrical impedance tomography," *Nature*, vol. 380, no. 6574, pp. 509–512, Apr. 1996. 1

[31] J. W. Schmidt and M. R. Moldover, "Dielectric permittivity of eight gases measured with cross capacitors," *International Journal Of Thermophysics*, vol. 24, no. 2, pp. 375–403, Mar. 2003. 2

[32] D. Pollock, *Physical properties of materials for engineers.* CRC Press, 1993. 3, 13

[33] A. P. Gregory and R. N. Clarke, "Traceable measurements of the static permittivity of dielectric reference liquids over the temperature range 5-50 degrees C," *Measurement Science & Technology*, vol. 16, no. 7, pp. 1506–1516, July 2005. 3

[34] K. Kao, *Dielectric phenomena in solids: With emphasis on physical concepts of electronic processes*, ser. Referex Engineering. Academic Press, 2004. 3

[35] E. H. Grant, R. J. Sheppard, and G. P. South, *Dielectric Behaviour of Biological Molecules in Solution.* Oxford University Press, 1978. 3, 29, 52, 74

[36] U. Kaatze, "Perspectives in dielectric measurement techniques for liquids," *Measurement Science & Technology*, vol. 19, no. 11, p. 112001, Nov. 2008. 4, 143

[37] H. Lodish, *Molecular cell biology*, 6th ed. W.H. Freeman, 2008. 4

[38] T. A. Khan, F. W. Sellke, and R. J. Laham, "Therapeutic angiogenesis: Protein-based therapy for coronary artery disease," *Expert Opinion on Pharmacotherapy*, vol. 4, no. 2, pp. 219–226, 2003. 4

[39] Jack and Jacobs, "Combating cardiovascular disease with angiogenic therapy," *Drug Discovery Today*, vol. 12, no. 23-24, pp. 1040 – 1045, 2007. 4

[40] J. Fischer, A. Kolk, S. Wolfart, C. Pautke, P. H. Warnke, C. Plank, and R. Smeets, "Future of local bone regeneration - protein versus gene therapy," *Journal of Craniomaxillofacial Surgery*, vol. 39, no. 1, pp. 54–64, Jan. 2011. 4

[41] D. Salom, B. R. Hill, J. D. Lear, and W. F. DeGrado, "pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus," *Biochemistry*, vol. 39, no. 46, pp. 14 160–14 170, 2000. 4, 149

[42] M. Hashimoto, E. Rockenstein, L. Crews, and E. Masliah, "Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases," *Neuromolecular Medicine*, vol. 4, no. 1-2, pp. 21–35, 2003. 4

[43] W. Xie, O. W. Wan, and K. K. K. Chung, "New insights into the role of mitochondrial dysfunction and protein aggregation in parkinson's disease," *Biochimica Et Biophysica Acta-Molecular Basis Of Disease*, vol. 1802, no. 11, pp. 935–941, Nov. 2010. 4, 132

[44] C. A. Ross and M. A. Poirier, "Protein aggregation and neurodegenerative disease," *Nature Medicine*, vol. 10, no. 7, pp. S10–S17, July 2004. 4, 132

[45] S. Takashima, "Electric dipole moments of globular proteins: Measurement and calculation with NMR and X-ray databases," *Journal Of Non-Crystalline Solids*, vol. 305, no. 1-3, pp. 303–310, July 2002. 4

[46] A. Bonincontro and G. Risuleo, "Dielectric spectroscopy as a probe for the investigation of conformational properties of proteins," *Spectrochimica Acta Part A-Molecular And Biomolecular Spectroscopy*, vol. 59, no. 12, pp. 2677–2684, Oct. 2003. 4

[47] A. Oleinikova, P. Sasisanker, and H. Weingartner, "What can really be learned from dielectric spectroscopy of protein solutions? A case study of ribonuclease A," *Journal Of Physical Chemistry B*, vol. 108, no. 24, pp. 8467–8474, June 2004. 4, 7, 8, 99, 134

[48] B. A. Mazzeo and A. J. Flewitt, "Observation of protein-protein interaction by dielectric relaxation spectroscopy of protein solutions for biosensor application," *Applied Physics Letters*, vol. 90, no. 12, p. 123901, 2007. 4, 24

[49] "Global in vitro diagnostic market analysis," RNCOS, Tech. Rep., December 2011. 5

[50] A. K. Bej, M. H. Mahbubani, J. L. Dicesare, and R. M. Atlas, "Polymerase chain reaction-gene probe detection of microorganisms by using filter-concentrated samples," *Applied and Environmental Microbiology*, vol. 57, no. 12, pp. 3529–3534, Dec. 1991. 5

[51] E. Leoni and P. P. Legnani, "Comparison of selective procedures for isolation and enumeration of legionella species from hot water systems," *Journal of Applied Microbiology*, vol. 90, no. 1, pp. 27–33, Jan. 2001. 5

[52] A. Voller, A. Bartlett, and D. E. Bidwell, "Enzyme immunoassays with special reference to ELISA techniques," *Journal of Clinical Pathology*, vol. 31, no. 6, pp. 507–520, 1978. 5

[53] O. Lazcka, F. J. D. Campo, and F. X. Muoz, "Pathogen detection: A perspective of traditional methods and biosensors," *Biosensors and Bioelectronics*, vol. 22, no. 7, pp. 1205 – 1217, 2007. 5

[54] B. W. Brooks, J. Devenish, C. L. Lutze-Wallace, D. Milnes, R. H. Robertson, and G. Berlie-Surujballi, "Evaluation of a monoclonal antibody-based enzyme-linked immunosorbent assay for detection of campylobacter fetus in bovine preputial washing and vaginal mucus samples," *Veterinary Microbiology*, vol. 103, no. 1-2, pp. 77–84, Oct. 2004. 5

[55] Y. H. Kim, J. S. Park, and H. I. Jung, "An impedimetric biosensor for real-time monitoring of bacterial growth in a microbial fermentor," *Sensors and Actuators B-chemical*, vol. 138, no. 1, pp. 270–277, 2009. 5

[56] T. Sun, C. van Berkel, N. G. Green, and H. Morgan, "Digital signal processing methods for impedance microfluidic cytometry," *Microfluidics and Nanofluidics*, vol. 6, no. 2, pp. 179–187, 2009. 5

[57] J. S. Daniels and N. Pourmand, "Label-free impedance biosensors: Opportunities and challenges," *Electroanalysis*, vol. 19, no. 12, pp. 1239–1257, 2007. 6

[58] U. Kaatze and Y. Feldman, "Broadband dielectric spectrometry of liquids and biosystems," *Measurement Science & Technology*, vol. 17, no. 2, pp. R17–R35, Feb. 2006. 7

[59] J. L. Oncley, "Dielectric behavior and atomic structure of serum albumin," *Biophysical Chemistry*, vol. 100, no. 1-3, pp. 151–158, 2003. 7

[60] J. Wyman, "Measurements of the dielectric constants of conducting media," *Phys. Rev.*, vol. 35, no. 6, pp. 623–634, Mar 1930. 8

[61] M. Faraday, J. W. Hittorf, and F. W. G. Kohlrausch, *The fundamental laws of electrolytic conduction: Memoirs by Faraday, Hittorf and F Kohlrausch*, H. M. Goodwin, Ed. Harper and Brothers, 1899. 8

[62] P. Debye, *Polar Molecules.* New York: Chemical Catalog Company, Inc., 1929. 8, 53, 135, 137

[63] W. Nernst, “Method zur bestimmung von dielektrizitatskonstanten,” *Z. phys. Chem.*, vol. 14, pp. 622–663, 1894. 8

[64] J. T. Edsall, “Wyman, Jeffries - Scientist, philosopher and adventurer,” *Biophysical Chemistry*, vol. 37, no. 1-3, pp. 7–14, Aug. 1990. 8

[65] J. Wyman, “Studies on the dielectric constant of proteins. I. Zein,” *J. Biol. Chem.*, vol. 90, pp. 443–476, 1931. 8

[66] L. Onsager, “Electric moments of molecules in liquids,” *Journal of the American Chemical Society*, vol. 58, pp. 1486–1493, 1936. 8

[67] J. L. Oncley, “Electric moments and relaxation times of protein molecules.” *The Journal of Physical Chemistry*, vol. 44, no. 9, pp. 1103–1113, 1940. 8, 78

[68] ——, “Studies of the dielectric properties of protein solutions. I. Carboxyhemoglobin,” *Journal of the American Chemical Society*, vol. 60, no. 5, pp. 1115–1123, 1938. 8

[69] ——, *The electric moments and the relaxation times of proteins as measured from their influence upon the dielectric constants of solutions.* Reinhold Publishing Corporation, 1943, ch. 22, pp. 543–567. 8, 78

[70] J. D. Ferry and J. L. Oncley, “Studies of the dielectric properties of protein solutions. III. Lactoglobulin,” *Journal of the American Chemical Society*, vol. 63, pp. 272–278, 1941. 8, 54, 58

[71] R. H. Cole and P. M. Gross, “A wide range capacitance-conductance bridge,” *Review of Scientific Instruments*, vol. 20, no. 4, pp. 252 –260, apr 1949. 8

[72] B. P. Jordan and E. H. Grant, “A cell for measuring the complex permittivity of lossy liquids at very high radio frequencies,” *Journal of Physics E: Scientific Instruments*, vol. 3, no. 10, p. 764, 1970. 8, 26

[73] R. Pethig, *Dielectric and Electronic Properties of Biological Materials.* John Wiley & Sons Ltd, 1979. 14, 24, 52

[74] R. Richert, “The modulus of dielectric and conductive materials and its modification by high electric fields,” *Journal Of Non-Crystalline Solids*, vol. 305, no. 1-3, pp. 29–39, 2002. 16, 116

[75] K. S. Cole and R. H. Cole, “Dispersion and absorption in dielectrics I. Alternating current characteristics,” *Journal Of Chemical Physics*, vol. 9, pp. 341–351, 1941. 20

[76] C. G. Essex, G. P. South, R. J. Sheppard, and E. H. Grant, “Bridge technique for measuring permittivity of a biological solution between 1 and 100 MHz,” *Journal of Physics E-scientific Instruments*, vol. 8, no. 5, pp. 385–389, 1975. 24, 26

[77] W. M. Haynes, Ed., *CRC Handbook of Chemistry and Physics*, 91st ed. CRC Press, 2011. 24

[78] A. Serghei, M. Tress, J. R. Sangoro, and F. Kremer, "Electrode polarization and charge transport at solid interfaces," *Physical Review B*, vol. 80, no. 18, p. 184301, 2009. 29, 34

[79] W. Scheider, "Theory of the frequency dispersion of electrode polarization. Topology of networks with fractional power frequency dependence," *The Journal of Physical Chemistry*, vol. 79, no. 2, pp. 127–136, 1975. 29, 35

[80] I. Wolff, "A study of polarization capacity over a wide frequency band," *Phys. Rev.*, vol. 27, pp. 755–763, Jun 1926. 29

[81] H. P. Schwan, "Electrode polarization impedance and measurements in biological materials," *Annals of the New York Academy of Sciences*, vol. 148, no. 1, pp. 191–209, 1968. 29

[82] J. B. Bates, Y. T. Chu, and W. T. Stribling, "Surface-topography and impedance of metal-electrolyte interfaces," *Physical Review Letters*, vol. 60, no. 7, pp. 627–630, Feb. 1988. 29

[83] F. Kremer and A. Schnhals, *Broadband Dielectric Spectroscopy*, F. Kremer and A. Schnhals, Eds. Springer, 2003. 29, 35

[84] L. Nyikos and T. Pajkossy, "Fractal dimension and fractional power frequency-dependent impedance of blocking electrodes," *Electrochimica Acta*, vol. 30, no. 11, pp. 1533–1540, 1985. 29

[85] F. Bordi, C. Cametti, and T. Gili, "Reduction of the contribution of electrode polarization effects in the radiowave dielectric measurements of highly conductive biological cell suspensions," *Bioelectrochemistry*, vol. 54, no. 1, pp. 53–61, 2001. 29, 30

[86] B. L. Mellor, N. A. Kellis, and B. A. Mazzeo, "Electrode polarization of Galinstan electrodes for liquid impedance spectroscopy," *Review Of Scientific Instruments*, vol. 82, no. 4, p. 046110, Apr. 2011. 29

[87] D. Malleo, J. T. Nevill, A. van Ooyen, U. Schnakenberg, L. P. Lee, and H. Morgan, "Note: Characterization of electrode materials for dielectric spectroscopy," *Review Of Scientific Instruments*, vol. 81, no. 1, p. 016104, Jan. 2010. 29, 44, 133

[88] T. Pajkossy, "Impedance spectroscopy at interfaces of metals and aqueous solutions – surface roughness, CPE and related issues," *Solid State Ionics*, vol. 176, no. 25-28, pp. 1997 – 2003, 2005, international Workshop on Impedance Spectroscopy for Characterisation of Materials and Structures. 29

[89] G. Ruiz and C. J. Felice, "Non-linear response of an electrode-electrolyte interface impedance with the frequency," *Chaos Solitons & Fractals*, vol. 31, no. 2, pp. 327–335, Jan. 2007. 29, 44

[90] H. Sanabria and J. Miller, John H., “Relaxation processes due to the electrode-electrolyte interface in ionic solutions,” *Physical Review E*, vol. 74, no. 5, p. 051505, Nov. 2006. 29

[91] P. Mirtaheri, S. Grimnes, and O. G. Martinsen, “Electrode polarization impedance in weak NaCl aqueous solutions,” *Ieee Transactions On Biomedical Engineering*, vol. 52, no. 12, pp. 2093–2099, Dec. 2005. 29, 35

[92] B. A. Mazzeo, “Parasitic capacitance influence of potential-sensing electrodes on four-electrode liquid impedance measurements,” *Journal of Applied Physics*, vol. 105, p. 094106, 2009. 30

[93] R. Richert, “Insulated electrodes for eliminating conductivity in dielectric relaxation experiments,” *Eur. Phys. J. B*, vol. 68, no. 2, pp. 197–200, 2009. 30

[94] P. A. Cirkel, J. P. M. van der Ploeg, and G. J. M. Koper, “Electrode effects in dielectric spectroscopy of colloidal suspensions,” *Physica A: Statistical and Theoretical Physics*, vol. 235, no. 1-2, pp. 269 – 278, 1997, proceedings of the Workshop on Colloid Physics. 30, 34

[95] K. Asami, “Design of a measurement cell for low-frequency dielectric spectroscopy of biological cell suspensions,” *Measurement Science & Technology*, vol. 22, no. 8, p. 085801, Aug. 2011. 30, 46

[96] H. Maruska and J. Stevens, “Technique for extracting dielectric permittivity from data obscured by electrode polarization,” *Electrical Insulation, IEEE Transactions on*, vol. 23, no. 2, pp. 197 –200, apr 1988. 30

[97] S. L. Srivastava and R. Dhar, “Characteristic time of ionic conductance and electrode polarization capacitance in some organic liquids by low-frequency dielectric-spectroscopy,” *Indian Journal Of Pure & Applied Physics*, vol. 29, no. 11, pp. 745–751, Nov. 1991. 30

[98] B. A. Mazzeo, S. Chandra, B. L. Mellor, and J. Arellano, “Temperature-stable parallel-plate dielectric cell for broadband liquid impedance measurements,” *Review of Scientific Instruments*, vol. 81, p. 125103, 2010. 30, 37, 53, 133

[99] Z. Kerner and T. Pajkossy, “On the origin of capacitance dispersion of rough electrodes,” *Electrochimica Acta*, vol. 46, no. 2-3, pp. 207–211, 2000. 35

[100] *Galinstan Material Safety Datasheet*, Geratherm Medical AG, Geratherm Medical AG Fahrenheitstraff 1 D-98716 Geschwenda, September 2006. 35

[101] M. D. Dickey, R. C. Chiechi, R. J. Larsen, E. A. Weiss, D. A. Weitz, and G. M. Whitesides, “Eutectic gallium-indium (EGAIN): A liquid metal alloy for the formation of stable structures in microchannels at room temperature,” *Advanced Functional Materials*, vol. 18, no. 7, pp. 1097–1104, 2008. 35, 40

[102] R. C. Chiechi, E. A. Weiss, M. D. Dickey, and G. M. Whitesides, "Eutectic gallium-indium (egain): A moldable liquid metal for electrical characterization of self-assembled monolayers," *Angewandte Chemie-international Edition*, vol. 47, no. 1, pp. 142–144, 2008. 35

[103] J. H. So, J. Thelen, A. Qusba, G. J. Hayes, G. Lazzi, and M. D. Dickey, "Reversibly deformable and mechanically tunable fluidic antennas," *Advanced Functional Materials*, vol. 19, no. 22, pp. 3632–3637, 2009. 35, 40

[104] E. F. Borra, G. Tremblay, Y. Huot, and J. Gauvin, "Gallium liquid mirrors: Basic technology, optical-shop tests and observations," *Publications of the Astronomical Society of the Pacific*, vol. 109, no. 733, pp. 319–325, 1997. 35, 40

[105] P. Surmann and H. Zeyat, "Voltammetric analysis using a self-renewable non-mercury electrode," *Analytical and Bioanalytical Chemistry*, vol. 383, no. 6, pp. 1009–1013, Nov. 2005. 35

[106] R. T. Kelly and A. T. Woolley, "Thermal bonding of polymeric capillary electrophoresis microdevices in water," *Analytical Chemistry*, vol. 75, no. 8, pp. 1941–1945, 2003. 37

[107] X. Y. Cui, J. F. Hetke, J. A. Wiler, D. J. Anderson, and D. C. Martin, "Electrochemical deposition and characterization of conducting polymer polypyrrole/PSS on multichannel neural probes," *Sensors And Actuators A-Physical*, vol. 93, no. 1, pp. 8–18, Aug. 2001. 44

[108] P. M. George, A. W. Lyckman, D. A. LaVan, A. Hegde, Y. Leung, R. Avasare, C. Testa, P. M. Alexander, R. Langer, and M. Sur, "Fabrication and biocompatibility of polypyrrole implants suitable for neural prosthetics," *Biomaterials*, vol. 26, no. 17, pp. 3511–3519, June 2005. 44

[109] G. Nimrod, A. Szilagyi, C. Leslie, and N. Ben-Tal, "Identification of DNA-binding proteins using structural, electrostatic and evolutionary features," *Journal of Molecular Biology*, vol. 387, no. 4, pp. 1040–1053, 2009. 51

[110] Z. X. Liang, I. V. Kurnikov, J. M. Nocek, A. G. Mauk, D. N. Beratan, and B. M. Hoffman, "Dynamic docking and electron-transfer between cytochrome B(5) and a suite of myoglobin surface-charge mutants. Introduction of a functional-docking algorithm for protein-protein complexes," *Journal of the American Chemical Society*, vol. 126, pp. 2785–2798, 2004. 51

[111] G. G. Tartaglia, A. Cavalli, R. Pellarin, and A. Caflisch, "The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates," *Protein Science*, vol. 13, no. 7, pp. 1939–1941, 2004. 51, 140

[112] S. E. Keefe and E. H. Grant, "Dipole-moment and relaxation-time of ribonuclease," *Physics in Medicine and Biology*, vol. 19, no. 5, pp. 701–707, 1974. 51

[113] B. L. Mellor, E. Cruz Cortes, D. D. Busath, and B. A. Mazzeo, "Method for estimating the internal dielectric constant of proteins using dielectric spectroscopy," *Journal Of Physical Chemistry B*, vol. 115, no. 10, pp. 2205–2213, March 2011. 51, 87, 124, 133, 136

[114] L. Keszthelyi, "Electrooptical and biophysical measurements on membrane fragments," *Colloids And Surfaces A-Physicochemical And Engineering Aspects*, vol. 209, no. 2-3, pp. 173–183, Sept. 2002. 51

[115] J. L. Oncley, "The investigation of proteins by dielectric measurements." *Chem. Rev.*, vol. 30, no. 3, pp. 433–450, June 1942. 52

[116] P. Drude, *Zeitschrift fr physikalische Chemie*, vol. 23, p. 267, 1897. 52

[117] J. B. Hasted, "Aqueous dielectrics," *Journal of Molecular Structure*, vol. 28, p. 227, 1973. 53, 100

[118] M. Nigen, T. Croguennec, D. Renard, and S. Bouhallab, "Temperature affects the supramolecular structures resulting from alpha-lactalbumin-lysozyme interaction," *Biochemistry*, vol. 46, no. 5, pp. 1248–1255, Feb. 2007. 53, 135, 146

[119] K. H. Valkonen, N. Marttinen, and T. Alatossava, "Electrophoretic methods for fractionation of native and heat-denatured bovine beta-lactoglobulin," *Bioseparation*, vol. 10, pp. 145–152, 2002. 54

[120] S. S. Rogers, P. Venema, J. P. M. Van der Ploeg, E. Van der Linden, L. M. C. Sagis, and A. M. Donald, "Investigating the permanent electric dipole moment of beta-lactoglobulin fibrils, using transient electric birefringence," *Biopolymers*, vol. 82, no. 3, pp. 241–252, June 2006. 54

[121] L. Korson, W. Drost-Hansen, and F. J. Millero, "Viscosity of water at various temperatures," *The Journal of Physical Chemistry*, vol. 73, no. 1, pp. 34–39, 1969. 57

[122] A. Bonincontro, A. De Francesco, and G. Onori, "Influence of pH on lysozyme conformation revealed by dielectric spectroscopy," *Colloids and Surfaces B-Biointerfaces*, vol. 12, no. 1, pp. 1–5, Oct. 1998. 59

[123] A. Warshel and J. Aqvist, "Electrostatic energy and macromolecular function," *Annual Review Of Biophysics And Biophysical Chemistry*, vol. 20, pp. 267–298, 1991. 63

[124] K. A. Sharp and B. Honig, "Electrostatic interactions in macromolecules - theory and applications," *Annual Review Of Biophysics And Biophysical Chemistry*, vol. 19, pp. 301–332, 1990. 63, 75

[125] A. Mozo-Villarias, J. Cedano, and E. Querol, "A simple electrostatic criterion for predicting the thermal stability of proteins," *Protein Engineering*, vol. 16, no. 4, pp. 279–286, Apr. 2003. 63

[126] S. C. Harvey and J. Garcia de la Torre, "Coordinate systems for modeling the hydrodynamic resistance and diffusion-coefficients of irregularly shaped rigid macromolecules," *Macromolecules*, vol. 13, no. 4, pp. 960–964, 1980. 65

[127] J. Antosiewicz, "Computation of the dipole-moments of proteins," *Biophysical Journal*, vol. 69, no. 4, pp. 1344–1354, Oct. 1995. 65

[128] S. Takashima, "Use of protein database for the computation of the dipole-moments of normal and abnormal-hemoglobins," *Biophysical Journal*, vol. 64, no. 5, pp. 1550–1558, May 1993. 65

[129] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, Jan. 2000. 66

[130] C. P. Smyth, *Dielectric behavior and structure: Dielectric constant and loss, dipole moment and molecular structure.* McGraw-Hill, 1955. 67

[131] K. A. Hasselbalch, "The calculation of the hydrogen number of the blood from the free and bound carbon dioxide of the same and the binding of oxygen by the blood as a function of the hydrogen number," *Biochemistry*, vol. 78, pp. 112–144, 1916. 68

[132] D. Bashford and M. Karplus, "pKas of ionizable groups in proteins - Atomic detail from a continuum electrostatic model," *Biochemistry*, vol. 29, no. 44, pp. 10 219–10 225, Nov. 1990. 69, 70, 74, 79, 80

[133] D. Bashford and K. Gerwert, "Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin," *Journal Of Molecular Biology*, vol. 224, no. 2, pp. 473–486, Mar. 1992. 69

[134] E. Demchuk and R. C. Wade, "Improving the continuum dielectric approach to calculating pK(a)s of ionizable groups in proteins," *Journal Of Physical Chemistry*, vol. 100, no. 43, pp. 17 373–17 387, Oct. 1996. 69, 75, 79

[135] B. Honig and A. Nicholls, "Classical electrostatics in biology and chemistry," *Science*, vol. 268, no. 5214, pp. 1144–1149, May 1995. 69

[136] H. Nakamura, T. Sakamoto, and A. Wada, "A theoretical-study of the dielectric-constant of protein," *Protein Engineering*, vol. 2, no. 3, pp. 177–183, Sept. 1988. 70, 75

[137] D. Bashford, "An object-oriented programming suite for electrostatic effects in biological molecules an experience report on the MEAD project," *Scientific Computing in Object-Oriented Parallel Environments*, vol. 1343, pp. 233–240, 1997. 70

[138] G. P. South and E. H. Grant, "Dielectric dispersion and dipole moment of myoglobin in water," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 328, no. 1574, pp. 371–387, 1972. 71, 90

[139] S. A. Hassan and E. L. Mehler, “A critical analysis of continuum electrostatics: The screened coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins,” *Proteins-Structure Function And Genetics*, vol. 47, no. 1, pp. 45–61, 2002. 74

[140] X. F. Li, S. A. Hassan, and E. L. Mehler, “Long dynamics simulations of proteins using atomistic force fields and a continuum representation of solvent effects: Calculation of structural and dynamic properties,” *Proteins-Structure Function And Bioinformatics*, vol. 60, no. 3, pp. 464–484, Aug. 2005. 74

[141] J. G. Mandell, V. A. Roberts, M. E. Pique, V. Kotlovyi, J. C. Mitchell, E. Nelson, I. Tsigelny, and L. F. Ten Eyck, “Protein docking using continuum electrostatics and geometric fit,” *Protein Engineering*, vol. 14, no. 2, pp. 105–113, 2001. 74

[142] S. Hofinger and T. Simonson, “Dielectric relaxation in proteins: A continuum electrostatics model incorporating dielectric heterogeneity of the protein and time-dependent charges,” *Journal Of Computational Chemistry*, vol. 22, no. 3, pp. 290–305, Feb. 2001. 74

[143] M. R. Gunner, J. J. Mao, Y. F. Song, and J. Kim, “Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations?” *Biochimica Et Biophysica Acta-Bioenergetics*, vol. 1757, no. 8, pp. 942–968, Aug. 2006. 74

[144] T. Simonson, “Dielectric relaxation in proteins: The computational perspective,” *Photosynthesis Research*, vol. 97, no. 1, pp. 21–32, July 2008. 74

[145] Y. Hao, M. R. Pear, and D. D. Busath, “Molecular dynamics study of free energy profiles for organic cations in gramicidin A channels,” *Biophysical Journal*, vol. 73, no. 4, pp. 1699–1716, 1997. 74

[146] A. Parsegian, “Energy of an ion crossing a low dielectric membrane: Solutions to four relevant electrostatic problems.” *Nature*, vol. 221, pp. 844–846, 1969. 74

[147] D. Boda, D. D. Busath, D. Henderson, and S. Sokolowski, “Monte Carlo simulations of the mechanism for channel selectivity: The competition between volume exclusion and charge neutrality,” *Journal Of Physical Chemistry B*, vol. 104, no. 37, pp. 8903–8910, 2000. 74

[148] T. Vora, B. Corry, and S. H. Chung, “Brownian dynamics study of flux ratios in sodium channels,” *European Biophysics Journal With Biophysics Letters*, vol. 38, no. 1, pp. 45–52, 2008. 74

[149] C. Song and B. Corry, “Ion conduction in ligand-gated ion channels: Brownian dynamics studies of four recent crystal structures,” *Biophysical Journal*, vol. 98, no. 3, pp. 404–411, 2010. 74

[150] R. E. Georgescu, E. G. Alexov, and M. R. Gunner, “Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins,” *Biophysical Journal*, vol. 83, no. 4, pp. 1731–1748, 2002. 74, 80

[151] K. L. Baran, M. S. Chimenti, J. L. Schlessman, C. A. Fitch, K. J. Herbst, and B. E. Garcia-Moreno, “Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease,” *Journal Of Molecular Biology*, vol. 379, no. 5, pp. 1045–1062, 2008. 74

[152] C. A. Castaneda, C. A. Fitch, A. Majumdar, V. Khangulov, J. L. Schlessman, and B. E. Garcia-Moreno, “Molecular determinants of the pK(a) values of Asp and Glu residues in staphylococcal nuclease,” *Proteins-Structure Function And Bioinformatics*, vol. 77, no. 3, pp. 570–588, 2009. 74

[153] J. B. Matthew, F. R. N. Gurd, E. B. Garciamoreno, M. A. Flanagan, K. L. March, and S. J. Shire, “pH-dependent processes in proteins,” *Crc Critical Reviews In Biochemistry*, vol. 18, no. 2, pp. 91–197, 1985. 75

[154] R. H. Tredgold and P. N. Hole, “Dielectric behavior of dry synthetic polypeptides,” *Biochimica Et Biophysica Acta*, vol. 443, no. 1, pp. 137–142, 1976. 75

[155] M. K. Gilson and B. H. Honig, “The dielectric-constant of a folded protein,” *Biopolymers*, vol. 25, no. 11, pp. 2097–2119, Nov. 1986. 75

[156] A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. F. Hendrickson, and W. C. Still, “Electrostatic contributions to solvation energies - Comparison of free-energy perturbation and continuum calculations,” *Journal Of The American Chemical Society*, vol. 113, no. 4, pp. 1454–1455, Feb. 1991. 75, 91

[157] J. A. Rupley and G. Careri, “Protein hydration and function,” *Advances In Protein Chemistry*, vol. 41, pp. 37–172, 1991. 75

[158] T. Simonson, D. Perahia, and A. T. Brunger, “Microscopic theory of the dielectric-properties of proteins,” *Biophysical Journal*, vol. 59, no. 3, pp. 670–690, Mar. 1991. 75

[159] T. Simonson, D. Perahia, and G. Bricogne, “Intramolecular dielectric screening in proteins,” *Journal Of Molecular Biology*, vol. 218, no. 4, pp. 859–886, Apr. 1991. 75

[160] J. Warwicker and H. C. Watson, “Calculation of the electric-potential in the active-site cleft due to alpha-helix dipoles,” *Journal Of Molecular Biology*, vol. 157, no. 4, pp. 671–679, 1982. 75

[161] A. S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig, “On the calculation of pK(a)s in proteins,” *Proteins-Structure Function And Genetics*, vol. 15, no. 3, pp. 252–265, Mar. 1993. 75

[162] H. Oberoi and N. M. Allewell, “Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves,” *Biophysical Journal*, vol. 65, no. 1, pp. 48–55, July 1993. 75

[163] S. Kundu and P. Gupta-Bhaya, “Electrostatic energy in fields provides a route to dielectric permittivity of proteins,” *Journal Of Molecular Structure-Theochem*, vol. 639, pp. 21–26, Nov. 2003. 75

[164] B. Pennock and H. Schwan, “Further oberservations on the electrical properties of hemoglobin-bound water,” *Journal of Physical Chemistry*, vol. 73, pp. 2600–2610, 1969. 75

[165] S. Takashima and H. P. Schwan, “Dielectric dispersion of crystalline powders of amino acids, peptides, and proteins1,” *The Journal of Physical Chemistry*, vol. 69, no. 12, pp. 4176–4182, 1965. 75

[166] J. Antosiewicz, J. A. Mccammon, and M. K. Gilson, “Prediction of pH-dependent properties of proteins,” *Journal Of Molecular Biology*, vol. 238, no. 3, pp. 415–436, May 1994. 75

[167] B. GarciaMoreno, J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, and W. E. Stites, “Experimental measurement of the effective dielectric in the hydrophobic core of a protein,” *Biophysical Chemistry*, vol. 64, no. 1-3, pp. 211–224, Feb. 1997. 75

[168] J. J. Dwyer, A. G. Gittis, D. A. Karp, E. E. Lattman, D. S. Spencer, W. E. Stites, and B. Garcia-Moreno, “High apparent dielectric constants in the interior of a protein reflect water penetration,” *Biophysical Journal*, vol. 79, no. 3, pp. 1610–1620, Sept. 2000. 75, 77, 88

[169] D. A. Karp, A. G. Gittis, M. R. Stahley, C. A. Fitch, W. E. Stites, and B. Garcia-Moreno, “High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp,” *Biophysical Journal*, vol. 92, no. 6, pp. 2041–2053, Mar. 2007. 75

[170] T. Simonson and C. L. Brooks, “Charge screening and the dielectric constant of proteins: Insights from molecular dynamics,” *Journal Of The American Chemical Society*, vol. 118, no. 35, pp. 8452–8458, Sept. 1996. 75

[171] G. Loffler, H. Schreiber, and O. Steinhauser, “Calculation of the dielectric properties of a protein and its solvent: Theory and a case study,” *Journal Of Molecular Biology*, vol. 270, no. 3, pp. 520–534, July 1997. 75, 94

[172] H. MacKenzie, *Milk proteins: Chemistry and molecular biology.* Academic Press, New York, 1971, vol. 2. 77, 82, 133

[173] J. J. Birktoft and D. M. Blow, “Structure of crystalline alpha-chymotrypsin : V. The atomic structure of tosyle-alpha-chymotrypsin at 2 Å,” *Journal of Molecular Biology*, vol. 68, pp. 187–240, 1972. 77

[174] J. G. Kirkwood and J. B. Shumaker, “The influence of dipole moment fluctuations on the dielectric increment of proteins in solution,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38, no. 10, pp. 855–862, 1952. 78

[175] S. Takashima, “A study of proton fluctuation in protein. Experimental study of the Kirkwood-Shumaker theory,” *The Journal Of Physical Chemistry*, vol. 69, no. 7, pp. 2281–2286, 1965. 78

[176] K. W. Wagner, “Erklarung der dielektrischen nachwirkungs - vorgange auf. grund maxwellscher vorstellungen,” *Arch. Elektrotech.*, vol. 2, pp. 371–387, 1914. 78

[177] H. Fricke, “The Maxwell-Wagner dispersion in a suspension of ellipsoids,” *Journal of Physical Chemistry*, vol. 57, no. 9, pp. 934–937, 1953. 78

[178] G. Schwarz, “A theory of the low-frequency dielectric dispersion of colloid particles in electrolyte solution,” *Journal of Physical Chemistry*, vol. 66, no. 12, pp. 2636–2642, 1962. 78

[179] C. L. Stanton and K. N. Houk, “Benchmarking pK(a) prediction methods for residues in proteins,” *Journal Of Chemical Theory And Computation*, vol. 4, no. 6, pp. 951–966, June 2008. 79

[180] J. Antosiewicz and D. Porschke, “The nature of protein dipole-moments - experimental and calculated permanent dipole of alpha-chymotrypsin,” *Biochemistry*, vol. 28, no. 26, pp. 10 072–10 078, Dec. 1989. 79

[181] C. E. Felder, J. Prilusky, I. Silman, and J. L. Sussman, “A server and database for dipole moments of proteins,” *Nucleic Acids Research*, vol. 35, pp. W512–W521, July 2007. 79, 80, 88

[182] J. E. Nielsen, “Analysing the pH-dependent properties of proteins using pKa calculations,” *Journal of Molecular Graphics and Modelling*, vol. 25, no. 5, pp. 691 – 699, 2007. 79

[183] A. S. Yang and B. Honig, “On the pH-dependence of protein stability,” *Journal Of Molecular Biology*, vol. 231, no. 2, pp. 459–474, May 1993. 79

[184] P. Beroza and D. A. Case, “Including side chain flexibility in continuum electrostatic calculations of protein titration,” *Journal Of Physical Chemistry*, vol. 100, no. 51, pp. 20 156–20 163, Dec. 1996. 79

[185] J. E. Nielsen and G. Vriend, “Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations,” *Proteins-Structure Function And Genetics*, vol. 43, no. 4, pp. 403–412, June 2001. 79

[186] J. A. Grant, B. T. Pickup, and A. Nicholls, “A smooth permittivity function for Poisson Boltzmann solvation methods,” *Journal of Computational Chemistry*, vol. 22, no. 6, pp. 608–640, 2001. 79

[187] H. Li, A. D. Robertson, and J. H. Jensen, “Very fast empirical prediction and rationalization of protein pK(a) values,” *Proteins-Structure Function And Bioinformatics*, vol. 61, no. 4, pp. 704–721, Dec. 2005. 79, 81

[188] Y. He, J. Xu, and X. M. Pan, “A statistical approach to the prediction of pK(a) values in proteins,” *Proteins-Structure Function And Bioinformatics*, vol. 69, no. 1, pp. 75–82, 2007. 79

[189] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. Mccammon, “Electrostatics and diffusion of molecules in solution - Simulations with the University-of-Houston brownian dynamics program,” *Computer Physics Communications*, vol. 91, no. 1-3, pp. 57–95, Sept. 1995. 80

[190] M. N. Davies, C. P. Toseland, D. S. Moss, and D. R. Flower, “Benchmarking pK(a) prediction,” *BMC Biochemistry*, vol. 7, p. 18, June 2006. 80, 81

[191] G. Nimrod, M. Schushan, A. Szilgyi, C. Leslie, and N. Ben-Tal, “IDBPS: A web server for the identification of dna binding proteins.” *Bioinformatics*, vol. 26, no. 5, pp. 692–693, 2010. 80

[192] A. A. Kantardjiev and B. P. Atanasov, “PHEMTO: Protein pH-dependent electric moment tools,” *Nucleic Acids Research*, vol. 37, pp. W422–W427, July 2009. 80

[193] S. Ahmad and A. Sarai, “Moment-based prediction of DNA-binding proteins,” *Journal Of Molecular Biology*, vol. 341, no. 1, pp. 65–71, July 2004. 80

[194] Y. Nozaki and C. Tanford, “Examination of titration behavior,” *Methods Enzymol*, vol. 11, pp. 715–734, 1967. 80, 81

[195] E. J. Cohn and J. T. Edsall, *Proteins, Amino Acids and Peptides.* Reinhold Publishing Corporation, 1943. 80, 81

[196] T. E. Creighton, *Proteins: Structures and molecular properties*, 2, Ed. W. H. Freeman, New York, 1993. 80, 81

[197] A. Warshel and S. T. Russell, “Calculations of electrostatic interactions in biological-systems and in solutions,” *Quarterly Reviews Of Biophysics*, vol. 17, no. 3, pp. 283–422, 1984. 88

[198] A. Wlodawer, N. Borkakoti, D. S. Moss, and B. Howlin, “Comparison of two independently refined models of ribonuclease-A,” *Acta Crystallographica Section B*, vol. 42, no. 4, pp. 379–387, Aug 1986. 90, 91

[199] K. Osapay, Y. Theriault, P. E. Wright, and D. A. Case, “Solution structure of carbonmonoxy myoglobin determined from nuclear-magnetic-resonance distance and chemical-shift constraints,” *Journal Of Molecular Biology*, vol. 244, no. 2, pp. 183–197, Nov. 1994. 90

[200] R. Diamond, “Real-space refinement of structure of hen egg-white lysozyme,” *Journal Of Molecular Biology*, vol. 82, no. 3, pp. 371–&, 1974. 90

[201] J. McCammon and S. Harvey, *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, 1988. 93

[202] C. Brooks, M. Karplus, and B. Pettitt, *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*, ser. Advances in chemical physics. Wiley, 1988. 93

[203] R. Hockney and J. Eastwood, *Computer simulation using particles.* Institute of Physics, 1992. 93

[204] M. Allen and D. Tildesley, *Computer simulation of liquids.* Clarendon Press, 1999. 93

[205] W. Gunsteren, P. Weiner, and A. Wilkinson, *Computer simulation of biomolecular systems: Theoretical and experimental applications*, ser. Computer Simulation of Biomolecular Systems. Published under the KLUWER/ESCOM imprint by Kluwer Academic Publishers, 1997. 93

[206] O. Becker, A. MacKerell, B. Roux, and M. Watanabe, *Computational Biochemistry and Biophysics.* Marcel Dekker Incorporated, 2001. 93

[207] D. Rapaport, *The art of molecular dynamics simulation.* Cambridge University Press, 2004. 93

[208] C. Sagui and T. A. Darden, “Molecular dynamics simulations of biomolecules: Long-range electrostatic effects,” *Annual Review Of Biophysics And Biomolecular Structure*, vol. 28, pp. 155–179, 1999. 93

[209] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, Sept. 2002. 93

[210] M. Karplus and J. Kuriyan, “Molecular dynamics and protein function,” *Proceedings Of The National Academy Of Sciences Of The United States Of America*, vol. 102, no. 19, pp. 6679–6685, May 2005. 93

[211] H. Z. Mao and J. Weber, “Identification of the beta(tp) site in the X-ray structure of f-1-atpase as the high-affinity catalytic site,” *Proceedings Of The National Academy Of Sciences Of The United States Of America*, vol. 104, no. 47, pp. 18 478–18 483, Nov. 2007. 93

[212] S. Boresch, P. Hochtl, and O. Steinhauser, “Studying the dielectric properties of a protein solution by computer simulation,” *Journal Of Physical Chemistry B*, vol. 104, no. 36, pp. 8743–8752, Sept. 2000. 93, 99, 102

[213] T. N. Heinz, W. F. van Gunsteren, and P. H. Hunenberger, “Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations,” *Journal Of Chemical Physics*, vol. 115, no. 3, pp. 1125–1136, July 2001. 93

[214] S. Boresch, S. Ringhofer, P. Hochtl, and O. Steinhauser, "Towards a better description and understanding of biomolecular solvation," *Biophysical Chemistry*, vol. 78, no. 1-2, pp. 43–68, Apr. 1999. 93

[215] H. Weingartner, A. Knocks, S. Boresch, P. Hochtl, and O. Steinhauser, "Dielectric spectroscopy in aqueous solutions of oligosaccharides: Experiment meets simulation," *Journal Of Chemical Physics*, vol. 115, no. 3, pp. 1463–1472, July 2001. 94

[216] I. M. Svishchev and P. G. Kusalik, "Dynamics in liquid water, water-D2, and water-T2: A comparative simulation study," *The Journal of Physical Chemistry*, vol. 98, no. 3, pp. 728–733, 1994. 94

[217] L. Q. Yang, S. Weerasinghe, P. E. Smith, and B. M. Pettitt, "Dielectric response of triplex DNA in ionic solution from simulations," *Biophysical Journal*, vol. 69, no. 4, pp. 1519–1527, Oct. 1995. 94

[218] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "Charmm: The biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009. 94, 95

[219] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187–217, 1983. 94

[220] M. Neumann and O. Steinhauser, "On the calculation of the frequency-dependent dielectric-constant in computer-simulations," *Chemical Physics Letters*, vol. 102, no. 6, pp. 508–513, 1983. 97

[221] ——, "On the calculation of the dielectric-constant using the Ewald-Kornfeld tensor," *Chemical Physics Letters*, vol. 95, no. 4-5, pp. 417–422, 1983. 97

[222] A. Knocks and H. Weingartner, "The dielectric spectrum of ubiquitin in aqueous solution," *The Journal of Physical Chemistry B*, vol. 105, no. 17, pp. 3635–3638, 2001. 102

[223] D. V. Matyushov, "Dipolar response of hydrated proteins," *Journal Of Chemical Physics*, vol. 136, no. 8, p. 085102, Feb. 2012. 102

[224] C. Cametti, S. Marchetti, C. Gambi, and G. Onori, "Dielectric relaxation spectroscopy of lysozyme aqueous solutions: Analysis of the δ-dispersion and the contribution of the hydration water," *J. Phys. Chem. B*, vol. 115, no. 21, pp. 7144–7153, May 2011. 102, 133, 134, 138

[225] S. Takashima, "The structure and dipole moment of globular proteins in solution and crystalline states: Use of NMR and X-ray databases for the numerical calculation of dipole moment," *Biopolymers*, vol. 58, no. 4, pp. 398–409, Apr. 2001. 116

[226] J. Lakowicz, *Principles of Fluorescence Spectroscopy*, ser. Principles of Fluorescence Spectroscopy. Springer, 2006, no. v. 1. 119

[227] S. W. Provencher and J. Glockner, "Estimation of globular protein secondary structure from circular-dichroism," *Biochemistry*, vol. 20, no. 1, pp. 33–37, 1981. 119

[228] G. Scapin, "Structural biology and drug discovery," *Current Pharmaceutical Design*, vol. 12, no. 17, pp. 2087–2097, 2006. 119

[229] S. P. Williams, L. F. Kuyper, and K. H. Pearce, "Recent applications of protein crystallography and structure-guided drug design," *Current Opinion In Chemical Biology*, vol. 9, no. 4, pp. 371–380, Aug. 2005. 119

[230] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch, and G. Chang, "Structure of p-glycoprotein reveals a molecular basis for poly-specific drug binding," *Science*, vol. 323, no. 5922, pp. 1718–1722, 2009. 119

[231] N. M. Green, "Avidin. 1. The use of biotin for kinetic studies and for assay," *Biochemical Journal*, vol. 89, p. 585, 1963. 119

[232] ——, "Avidin," ser. Advances in Protein Chemistry, J. T. E. C.B. Anfinsen, Jr. and F. M. Richards, Eds. Academic Press, 1975, vol. 29, pp. 85 – 133. 120

[233] Y. Pazy, Y. Eisenberg-Donlovich, O. H. Laitinen, M. S. Kulomaa, E. A. Bayer, M. Wilchek, and O. Livnah, "Dimer-tetramer transition between solution and crystalline states of streptavidin and avidin mutants," *Journal of Bacteriology*, vol. 185, no. 14, pp. 4050–4056, July 2003. 124

[234] T. Sano and C. R. Cantor, "Intersubunit contacts made by tryptophan-120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin," *Proceedings Of The National Academy Of Sciences Of The United States Of America*, vol. 92, no. 8, pp. 3180–3184, Apr. 1995. 124

[235] E. Van Der Linden and P. Venema, "Self-assembly and aggregation of proteins," *Current Opinion In Colloid & Interface Science*, vol. 12, no. 4-5, pp. 158–165, Oct. 2007. 130

[236] S. G. Zhang, "Fabrication of novel biomaterials through molecular self-assembly," *Nature Biotechnology*, vol. 21, no. 10, pp. 1171–1178, Oct. 2003. 132

[237] Y. H. Yong and E. A. Foegeding, "Caseins: Utilizing molecular chaperone properties to control protein aggregation in foods," *J. Agric. Food Chem.*, vol. 58, no. 2, pp. 685–693, Dec. 2009. 132

[238] P. Ringler and G. E. Schulz, "Self-assembly of proteins into designed networks," *Science*, vol. 302, no. 5642, pp. 106–109, Oct. 2003. 132

[239] Y. Desfougeres, T. Croguennec, V. Lechevalier, S. Bouhallab, and F. Nau, "Charge and size drive spontaneous self-assembly of oppositely charged globular proteins into microspheres," *J. Phys. Chem. B*, vol. 114, no. 12, pp. 4138–4144, Mar. 2010. 132, 133, 139, 140

[240] A. Parbhu, H. Lin, J. Thimm, and R. Lal, "Imaging real-time aggregation of amyloid beta protein (1-42) by atomic force microscopy," *Peptides*, vol. 23, no. 7, pp. PII S0196–9781(02)00 061–X, July 2002. 132, 143

[241] L. R. Helms and R. Wetzel, "Specificity of abnormal assembly in immunoglobulin light chain deposition disease and amyloidosis," *Journal Of Molecular Biology*, vol. 257, no. 1, pp. 77–86, Mar. 1996. 132, 143

[242] Y. Georgalis, E. B. Starikov, B. Hollenbach, R. Lurz, E. Scherzinger, W. Saenger, H. Lehrach, and E. E. Wanker, "Huntingtin aggregation monitored by dynamic light scattering," *Proceedings Of The National Academy Of Sciences Of The United States Of America*, vol. 95, no. 11, pp. 6118–6121, May 1998. 132

[243] B. S. Kendrick, J. L. Cleland, X. Lam, T. Nguyen, T. W. Randolph, M. C. Manning, and J. F. Carpenter, "Aggregation of recombinant human interferon gamma: Kinetics and structural transitions," *Journal Of Pharmaceutical Sciences*, vol. 87, no. 9, pp. 1069–1076, Sept. 1998. 132, 143

[244] A. C. Dong, S. J. Prestrelski, S. D. Allison, and J. F. Carpenter, "Infrared spectroscopic studies of lyophilization-induced and temperature-induced protein aggregation," *Journal Of Pharmaceutical Sciences*, vol. 84, no. 4, pp. 415–424, Apr. 1995. 132, 143

[245] B. A. Kerwin, M. C. Heller, S. H. Levin, and T. W. Randolph, "Effects of Tween 80 and sucrose on acute short-term stability and long-term storage at -20 degrees C of a recombinant hemoglobin," *Journal Of Pharmaceutical Sciences*, vol. 87, no. 9, pp. 1062–1068, Sept. 1998. 132, 143

[246] S. J. Tomski and R. M. Murphy, "Kinetics of aggregation of synthetic beta-amyloid peptide," *Archives Of Biochemistry And Biophysics*, vol. 294, no. 2, pp. 630–638, May 1992. 132, 143

[247] H. C. Mahler, W. Friess, U. Grauschopf, and S. Kiese, "Protein aggregation: Pathways, induction factors and analysis," *Journal Of Pharmaceutical Sciences*, vol. 98, no. 9, pp. 2909–2934, Sept. 2009. 132

[248] I. Choi, Y. S. Huh, and D. Erickson, "Ultra-sensitive, label-free probing of the conformational characteristics of amyloid beta aggregates with a SERS active nanofluidic device," *Microfluidics And Nanofluidics*, vol. 12, no. 1-4, pp. 663–669, Jan. 2012. 132

[249] B. Demeule, C. Palais, G. Machaidze, R. Gurny, and T. Arvinte, "New methods allowing the detection of protein aggregates a case study on trastuzumab," *Mabs*, vol. 1, no. 2, pp. 142–150, Mar. 2009. 132

[250] T. P. J. Knowles, W. Shu, F. Huber, H. P. Lang, C. Gerber, C. M. Dobson, and M. E. Welland, "Label-free detection of amyloid growth with microcantilever sensors," *Nanotechnology*, vol. 19, no. 38, p. 384007, Sept. 2008. 132

[251] D. E. Kuehner, J. Engmann, F. Fergg, M. Wernick, H. W. Blanch, and J. M. Prausnitz, "Lysozyme net charge and ion binding in concentrated aqueous electrolyte solutions," *J. Phys. Chem. B*, vol. 103, no. 8, pp. 1368–1374, Jan. 1999. 133

[252] T. Peters, "Serum-albumin," *Advances In Protein Chemistry*, vol. 37, pp. 161–245, 1985. 133

[253] S. Takashima, "Dielectric dispersion of albumins. Studies of denaturation by dielectric measurement," *Biochimica et Biophysica Acta (BBA) - Specialized Section on Biophysical Subjects*, vol. 79, no. 3, pp. 531 – 538, 1964. 133

[254] B. L. Mellor, E. Cruz Cortes, S. Khadka, and B. A. Mazzeo, "Increased bandwidth for dielectric spectroscopy of proteins through electrode surface preparation," *Review Of Scientific Instruments*, vol. 83, p. 015110, 2012. 133

[255] M. Wolf, R. Gulich, P. Lunkenheimer, and A. Loidl, "Relaxation dynamics of a protein solution investigated by dielectric spectroscopy," *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, vol. 1824, no. 5, pp. 723 – 730, 2012. 134

[256] P. R. Majhi, R. R. Ganta, R. P. Vanam, E. Seyrek, K. Giger, and P. L. Dubin, "Electrostatically driven protein aggregation: Beta-lactoglobulin at low ionic strength," *Langmuir*, vol. 22, no. 22, pp. 9150–9159, 2006, pMID: 17042523. 140

[257] E. Y. Chi, S. Krishnan, T. W. Randolph, and J. F. Carpenter, "Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation," *Pharmaceutical Research*, vol. 20, pp. 1325–1336, 2003, 10.1023/A:1025771421906. 140

[258] I. Campbell, *Biophysical Techniques*. Oxford University Press, 2012. 143

[259] J. Arrondo and A. Alonso, *Advanced Techniques in Biophysics*, ser. Springer Series in Biophysics. Springer-Verlag, 2006. 143

[260] H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel, and M. Mons, "Strength of Hydrogen bonds in methionine residues revealed by gas-phase IR/UV spectroscopy," *Journal Of Physical Chemistry Letters*, vol. 3, no. 6, pp. 755–759, Mar. 2012. 144

[261] R. Chari, S. N. Singh, S. Yadav, D. N. Brems, and D. S. Kalonia, "Determination of the dipole moments of RNASE SA wild type and a basic mutant," *Proteins-Structure Function And Bioinformatics*, vol. 80, no. 4, pp. 1041–1052, Apr. 2012. 144

[262] K. K. Irikura, J. K. Merle, and Y. Simon-Manso, “Tryptic Y(++) fragment ion distributions are guided by coulombic repulsion,” *Journal Of The American Society For Mass Spectrometry*, vol. 23, no. 3, pp. 483–488, Mar. 2012. 144

[263] M. M. Reif, P. H. Hunenberger, and C. Oostenbrink, “New interaction parameters for charged amino acid side chains in the GROMOS force field,” *Journal of Chemical Theory and Computation*, vol. 0, no. 0, p. 0, 2012. 144

[264] S. Hofinger, E. Yamamoto, Y. Hirano, F. Zerbetto, T. Narumi, K. Yasuoka, and M. Yasui, “Structural features of aquaporin 4 supporting the formation of arrays and junctions in biomembranes,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 9, pp. 2234–2243, 2012. 144

[265] R. Chen, L. Li, and Z. P. Weng, “Zdock: An initial-stage protein-docking algorithm,” *Proteins-Structure Function And Genetics*, vol. 52, no. 1, pp. 80–87, July 2003. 146

[266] W. W. Thompson, D. K. Shay, E. Weintraub, L. Brammer, N. Cox, L. J. Anderson, and K. Fukuda, “Mortality associated with influenza and respiratory syncytial virus in the united states,” *The Journal of the American Medical Association*, vol. 289, no. 2, pp. 179–186, January 2003. 149

[267] CDC weekly influenza report - week 35, 2009. 149

[268] P. Kumar and E. A. Yildirim, “Minimum-volume enclosing ellipsoids and core sets,” *Journal Of Optimization Theory And Applications*, vol. 126, no. 1, pp. 1–21, July 2005. 175

[269] M. J. Todd and E. A. Yildirim, “On Khachiyan’s algorithm for the computation of minimum-volume enclosing ellipsoids,” *Discrete Applied Mathematics*, vol. 155, no. 13, pp. 1731–1744, Aug. 2007. 175

Appendix A

Refereed Publications and Presentations

The following papers were published and presentations given during the research performed for completion of this dissertation:

1. Brett L. Mellor, Efrén Cruz Cortés, Shiul Khadka, and Brian A. Mazzeo. Increased bandwidth for dielectric spectroscopy of proteins through electrode surface preparation. *Review of Scientific Instruments*, 83, 015110 (2012).
2. Brett L. Mellor, Shiul Khadka, David D. Busath, and Brian A. Mazzeo. Influence of pKa Shifts on the Calculated Dipole Moments of Proteins. *The Protein Journal*, 30, 490 (2011).
3. Brett L. Mellor, Nathan A. Kellis, Steven J. Brewer, David D. Busath, and Brian A. Mazzeo. Protein-ligand biosensing: dielectric spectroscopy and numerical simulation of molecular interactions. *Four Corners Section Meeting of the American Physical Society*, Tucson, AZ. October 21, 2011.
4. Brian A. Mazzeo, Efrén Cruz Cortés, and Brett L. Mellor. Electrode impedance due to polarization: the influence of surface roughness. *Four Corners Section Meeting of the American Physical Society*, Tucson, AZ. October 21, 2011.
5. Brett L. Mellor, Nathan A. Kellis, Steven J. Brewer, David D. Busath, and Brian A. Mazzeo. Detection of Molecular Interactions Using Impedance-Based Biosensors (Poster). *Utah Biomedical Engineering Conference*, Salt Lake City, UT. September 10, 2011.
6. Brett L. Mellor, Nathan A. Kellis, and Brian A. Mazzeo. Dielectric Spectroscopy of Molecular Interactions Based on the Avidin-Biotin Complex. *32nd Meeting of the IEEE Engineering in Medicine and Biology Society*, Boston, MA. August 30-September 3, 2011.
7. Brett L. Mellor, Nathan A. Kellis, and Brian A. Mazzeo. Note: Electrode polarization of Galinstan electrodes for liquid impedance spectroscopy. *Review of Scientific Instruments*, 82, 046110 (2011).
8. Brett L. Mellor, Efrén Cruz Cortés, David D. Busath, and Brian A. Mazzeo. Method for Estimating the Internal Permittivity of Proteins Using Dielectric Spectroscopy. *Journal of Physical Chemistry B*, 115, 2205 (2011).

9. Brett L. Mellor, Efrén Cruz Cortés, David D. Busath, and Brian A. Mazzeo. Estimation of the internal dielectric constant of proteins using measured and simulated charge moments. *Four Corners Section Meeting of the American Physical Society*, Ogden, UT. October 15, 2010.

10. Brian A. Mazzeo, Satyan Chandra, Brett L. Mellor, and Jesus Arellano. Temperature-stable parallel-plate dielectric cell for broadband liquid impedance measurements. *Review of Scientific Instruments*, 81, 125103 (2010).

11. Brett L. Mellor and Brian A. Mazzeo. A Method to Calculate Protein Dipole Moments. *Four Corners Section Meeting of the American Physical Society*, Golden, CO. October 24, 2009.

Appendix B

Minimum-Volume Enclosing Ellipsoid (MVEE) Fitting Algorithm

This section outlines an algorithm for calculating molecular radii from a PDB file. The algorithm is based on the Khachiyan algorithm, a method to solve for the minimum-volume enclosing ellipsoid (MVEE) around a set of points $S = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m\}$ where $\mathbf{p}_i \in \mathbb{R}^{d\times 1}$.

An ellipsoid $\mathcal{E}$ is described by the equation

$$\mathcal{E} = \left\{ \mathbf{x} \in \mathbb{R}^{d\times 1} \mid (\mathbf{x} - \mathbf{c})^T \, \mathbf{E} \, (\mathbf{x} - \mathbf{c}) \leq 1 \right\} \tag{B.1}$$

where $\mathbf{c} \in \mathbb{R}^{d\times 1}$ is the center of the ellipsoid and $\mathbf{E}$ is a $d \times d$ positive-definite matrix. The volume of the ellipsoid is given by

$$\mathrm{vol}\,(\mathcal{E}) = V_0 \det \left[\mathbf{E}^{-1}\right]^{\frac{1}{2}} \tag{B.2}$$

where V_0 is the volume of the unit sphere in dimension d. In order to find the MVEE containing the points of S, one must obtain $\mathbf{c}$ and $\mathbf{E}$ that will minimize $\det\left[\mathbf{E}^{-1}\right]^{\frac{1}{2}}$ subject to the constraint $(\mathbf{x} - \mathbf{c})^T \, \mathbf{E} \, (\mathbf{x} - \mathbf{c}) \leq 1$.

A detailed derivation of this problem is found elsewhere [268, 269]; only its solution will be given here. Letting $\mathbf{P} \in \mathbb{R}^{d\times m}$ be a matrix whose ith column is $\mathbf{p}_i$, $\mathbf{u} \in \mathbb{R}^{m\times 1}$ be the optimal solution of the Lagrangian dual of $\mathbf{P}$, and $\mathbf{U} \in \mathbb{R}^{m\times m}$ be the diagonal matrix whose diagonal entries are given by vector $\mathbf{u}$ yields

$$\mathrm{MVEE}(S) = \left\{ \mathbf{x} \in \mathbb{R}^{d\times 1} \mid (\mathbf{x} - \hat{\mathbf{c}})^T \, \hat{\mathbf{E}} \, (\mathbf{x} - \hat{\mathbf{c}}) \leq 1 \right\} \tag{B.3}$$

where

$$\hat{\mathbf{E}} = \frac{1}{d} \left[\mathbf{P}\mathbf{U}\mathbf{P}^T - \mathbf{P}\mathbf{u} \, (\mathbf{P}\mathbf{u})^T \right]^{-1} \tag{B.4}$$

and

$$\hat{\mathbf{c}} = \mathbf{P}\mathbf{u}. \tag{B.5}$$

For three-dimensional coordinates, the three calculated radii r_1, r_2, and r_3 are obtained through the singular value decomposition of matrix $\hat{\mathbf{E}}$ of the form

$$\hat{\mathbf{E}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^* \tag{B.6}$$

where

$$
\begin{aligned}
r_1 &= \frac{1}{\sqrt{\boldsymbol{\Sigma}\,(1,1)}}, &\text{(B.7)}\\
r_2 &= \frac{1}{\sqrt{\boldsymbol{\Sigma}\,(2,2)}}, &\text{(B.8)}\\
r_3 &= \frac{1}{\sqrt{\boldsymbol{\Sigma}\,(3,3)}}. &\text{(B.9)}
\end{aligned}
$$

a_{calc} is then the radius of the equivalent volume sphere,

$$
a_{\text{calc}} = (r_1 r_2 r_3)^{\frac{1}{3}}\,. \qquad \text{(B.10)}
$$

Appendix C

List of PDB Codes for pK_a Shift Analysis of Section 5.5.4

1AA0
1AVE
1BBL
1BEB
1BEO
1CDC
1CMF
1DE3
1DG9
1DIV
1DK3
1DSB
1EGO
1ENH
1ERC
1ERT
1ERU
1FK6
1FKS
1FW4
1GDC
1GS9
1GYM
1HDJ
1HHP
1HPX
1I1B
1IGD
1J8Q
1L63
1LYZ
1LZ1
1MEG
1P2P

1PGA
1PLB
1PNB
1PNT
1POC
1POH
1QBR
1QBU
1RNZ
1SBT
1SSO
1UBQ
1UTG
1XNB
1YMB
2AVI
2CDV
2FX2
2LZT
2RN2
2TRX
3OGB
3RN3
3RNT
3SSI
4ICB
4MBN
4PTI
5P21
6LYZ
7RSA

Appendix D

MATLAB Code for Dipole Moment Calculation

```
function [myPH,myDipoles] = PDBreader(fname,use_shifted_pKas,model_set)

%%%               %%%
%%%% PDBreader.m %%%%
%%%               %%%
%
%Created By Brett Mellor on 5/13/09
%
% 1. Loads PDB file into ’data’
% 2. Displays information about the Protein
% 3. Plots Protein onto 3D space
% 4. Computes Core Dipole Moment
% 5. Computes Isoelectric Point (for either shifted or non shifted pKa)
% 6. Computes Surface Charge Dipole Moment
% 7. Plots Backbone of Protein with point charges and dipole moment
%

%%%Control Box%%%
chains = ’ABCDEF’;
PH = [4:.2:10];  %’i’ for isoelectric point

%%%Visual Control Box%%%
plotcharge_VS_pH = 0;
plotdipoles_VS_pH = 0;
plotbb = 1;
use_center_charge = 0;
dispDipoleMoment = 1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%   Load PDB file into ’data’  %%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
bbmain = [];
bbdivider = [];
modeldivider = [];
foundfirst = 0;

fid=fopen(fname);
i = 0;
while 1
    i = i+1;
    tline = fgetl(fid);
    if ˜ischar(tline)
        break
    end
    if ˜strcmp(tline(1:6),’ANISOU’)
        if i == 1
            datamain = tline;
            Title = tline(11:80);
            if strcmp(tline(1:4),’ATOM’)
                aaSTART = str2double(datamain(i,24:26));
                foundfirst = 1;
                bbmain = [bbmain;tline];
                bbdivider = 1;
```

```
            end
        else
            datamain = [datamain;tline];

            if strcmp(tline(1:5),'TITLE')
                Title = [Title;tline(11:80)];
            end
            if strcmp(tline(1:4),'ATOM') && foundfirst==0
                aaSTART = str2double(datamain(i,24:26));
                foundfirst = 1;
            end
            if strcmp(tline(1:5),'MODEL')
                modeldivider = [modeldivider i];
                bbdivider = [bbdivider length(bbmain)+1];
            end
            if strcmp(tline(1:4),'ATOM') && ismember(tline(22),chains)...
                    && (strcmp(tline(13:16),' C  ') || strcmp(tline(13:16),' CA ')...
                    || strcmp(tline(13:16),' N  '))
                bbmain = [bbmain;tline];
            end
            if strcmp(tline(1:4),'ATOM') && strcmp(tline(13:16),' N  ')...
                    && ismember(tline(22),chains) && (str2double(tline(24:26))==aaSTART...
                    || ~strcmp(tline(22),bbmain(end-1,22)))
                bbdivider = [bbdivider length(bbmain(:,1))];
            end

            if strcmp(tline(1:4),'ATOM') && strcmp(tline(13:16),' N  ')...
                    && ismember(tline(22),chains) && ~strcmp(tline(22),bbmain(end-1,22))
                bbdivider = [bbdivider length(bbmain(:,1))];
            end
        end
    else
        i = i-1;
    end
end
fclose(fid);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%  Compute Core dipole moment of protein    %%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
myDipoles = [];
myPH = PH;
if strfind(atomnames,'H');
    containsH = 1;
    display('Single Model Contains Hydrogen Atoms')
else
    containsH = 0;
end

for PH = PH

    dind = 1;
    dipCO = 2.31;
    dipNH = 1.31;
    dipCN = .2;

    for i = 1:length(data)
        if containsH==0 %%NO HYDROGEN IN MODEL
            %%%Backbone C=0 bonds, N-H bonds, and  C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' O  ')...
                    && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = (dipCO+dipNH)*dipole(dind,1:3)/normfactor;
                dind = dind + 1;

```

```
            dipole(dind,1) = str2double(data(i-2,32:38))-str2double(data(i-3,32:38));
            dipole(dind,2) = str2double(data(i-2,40:46))-str2double(data(i-3,40:46));
            dipole(dind,3) = str2double(data(i-2,48:54))-str2double(data(i-3,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
            dipole(dind,1:3) = (dipCN)*dipole(dind,1:3)/normfactor;
            dind = dind + 1;
        end

        %%%Backbone C-N bond (2nd)
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' C  ')...
                && ismember(data(i,22),chains)
            for j = 2:length(data)-i
                if strcmp(data(i+j,13:16),' N  ')
                    dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+j,32:38));
                    dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+j,40:46));
                    dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+j,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = (dipCN)*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                    break
                end
            end
        end

        %%Sidechain Asn C=0 bonds
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OD1')...
                && strcmp(data(i,18:20),'ASN') && ismember(data(i,22),chains)
            dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
            dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
            dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
            dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
            dind = dind + 1;
        end
        %%%Sidechain Gln C=0 bonds
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OE1')...
                && strcmp(data(i,18:20),'GLN') && ismember(data(i,22),chains)
            dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
            dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
            dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
            dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
            dind = dind + 1;
        end
        %%%Sidechain Asp 2 partial C=0 bonds
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OD1')...
                && strcmp(data(i,18:20),'ASP') && ismember(data(i,22),chains)
            dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
            dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
            dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
            dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
            dind = dind + 1;

            dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i+1,32:38));
            dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i+1,40:46));
            dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i+1,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
            dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
            dind = dind + 1;
        end
        %%%Sidechain Glu 2 partial C=0 bonds
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OE1')...
                && strcmp(data(i,18:20),'GLU') && ismember(data(i,22),chains)
            dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
            dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
            dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
            normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
```

```
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;

                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Arg 1 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CD ')...
                    && strcmp(data(i,18:20),'ARG') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Asn 1 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CG ')...
                    && strcmp(data(i,18:20),'ASN') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Gln 1 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CD ')...
                    && strcmp(data(i,18:20),'GLN') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain His 3 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CD2')...
                    && strcmp(data(i,18:20),'HIS') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i-2,32:38))-str2double(data(i-1,32:38));
                dipole(dind,2) = str2double(data(i-2,40:46))-str2double(data(i-1,40:46));
                dipole(dind,3) = str2double(data(i-2,48:54))-str2double(data(i-1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+1,32:38))-str2double(data(i-1,32:38));
                dipole(dind,2) = str2double(data(i+1,40:46))-str2double(data(i-1,40:46));
                dipole(dind,3) = str2double(data(i+1,48:54))-str2double(data(i-1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Lys 1 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CE ')...
                    && strcmp(data(i,18:20),'LYS') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+1,48:54));
```

```
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Trp 2 C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' CD1')...
                    && strcmp(data(i,18:20),'TRP') && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+3,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i+3,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i+3,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        else
            %%%%%%FOR SINGLE MODELS CONTAINING HYDROGEN
            %%%Backbone C=0 bonds,and  C-N bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' O  ')...
                    && ismember(data(i,22),chains)
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = (dipCO)*dipole(dind,1:3)/normfactor;
                dind = dind + 1;

                dipole(dind,1) = str2double(data(i-2,32:38))-str2double(data(i-3,32:38));
                dipole(dind,2) = str2double(data(i-2,40:46))-str2double(data(i-3,40:46));
                dipole(dind,3) = str2double(data(i-2,48:54))-str2double(data(i-3,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = (dipCN)*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Backbone C-N bond (2nd)
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' C  ')...
                    && ismember(data(i,22),chains)
                for j = 2:length(data)-i
                    if strcmp(data(i+j,13:16),' N  ')
                        dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+j,32:38));
                        dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+j,40:46));
                        dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+j,48:54));
                        normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                        dipole(dind,1:3) = (dipCN)*dipole(dind,1:3)/normfactor;
                        dind = dind + 1;
                        break
                    end
                end
            end
            %%%Backbone N-H bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ')...
                    && ismember(data(i,22),chains) && ~strcmp(data(i,24:26),'  1')
                for j = 2:length(data)-i
                    if strcmp(data(i+j,13:16),' H  ')
                        dipole(dind,1) = str2double(data(i+j,32:38))-str2double(data(i,32:38));
                        dipole(dind,2) = str2double(data(i+j,40:46))-str2double(data(i,40:46));
                        dipole(dind,3) = str2double(data(i+j,48:54))-str2double(data(i,48:54));
                        normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                        dipole(dind,1:3) = (dipNH)*dipole(dind,1:3)/normfactor;
                        dind = dind + 1;
                        break
                    end
                end
```

```
            end
            %%%Starting 3 N-H bond
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ')...
                    && ismember(data(i,22),chains) && strcmp(data(i,24:26),'  1')
                for j = 2:50
                    if strcmp(data(i+j,13:16),' H1 ')
                        dipole(dind,1) = str2double(data(i+j,32:38))-str2double(data(i,32:38));
                        dipole(dind,2) = str2double(data(i+j,40:46))-str2double(data(i,40:46));
                        dipole(dind,3) = str2double(data(i+j,48:54))-str2double(data(i,48:54));
                        normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                        dipole(dind,1:3) = (dipNH)*dipole(dind,1:3)/normfactor;
                        dind = dind + 1;
                        dipole(dind,1) = str2double(data(i+j+1,32:38))-str2double(data(i,32:38));
                        dipole(dind,2) = str2double(data(i+j+1,40:46))-str2double(data(i,40:46));
                        dipole(dind,3) = str2double(data(i+j+1,48:54))-str2double(data(i,48:54));
                        normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                        dipole(dind,1:3) = (dipNH)*dipole(dind,1:3)/normfactor;
                        dind = dind + 1;
                        dipole(dind,1) = str2double(data(i+j+2,32:38))-str2double(data(i,32:38));
                        dipole(dind,2) = str2double(data(i+j+2,40:46))-str2double(data(i,40:46));
                        dipole(dind,3) = str2double(data(i+j+2,48:54))-str2double(data(i,48:54));
                        normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                        dipole(dind,1:3) = (dipNH)*dipole(dind,1:3)/normfactor;
                        dind = dind + 1;
                        break
                    end
                end
            end

            %%%%%%%%                      %%%%%%%%
            %%%%%%%%    SIDECHAIN BONDS   %%%%%%%%
            %%%%%%%%                      %%%%%%%%

            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'ARG')...
                    && ismember(data(i,22),chains)
                %%%Sidechain Arg 5 N-H Bonds
                if strcmp(data(i,13:16),' NE ')
                    dipole(dind,1) = str2double(data(i+12,32:38))-str2double(data(i,32:38));
                    dipole(dind,2) = str2double(data(i+12,40:46))-str2double(data(i,40:46));
                    dipole(dind,3) = str2double(data(i+12,48:54))-str2double(data(i,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                    dipole(dind,1) = str2double(data(i+15,32:38))-str2double(data(i+3,32:38));
                    dipole(dind,2) = str2double(data(i+15,40:46))-str2double(data(i+3,40:46));
                    dipole(dind,3) = str2double(data(i+15,48:54))-str2double(data(i+3,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                    dipole(dind,1) = str2double(data(i+16,32:38))-str2double(data(i+3,32:38));
                    dipole(dind,2) = str2double(data(i+16,40:46))-str2double(data(i+3,40:46));
                    dipole(dind,3) = str2double(data(i+16,48:54))-str2double(data(i+3,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                    dipole(dind,1) = str2double(data(i+13,32:38))-str2double(data(i+2,32:38));
                    dipole(dind,2) = str2double(data(i+13,40:46))-str2double(data(i+2,40:46));
                    dipole(dind,3) = str2double(data(i+13,48:54))-str2double(data(i+2,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                    dipole(dind,1) = str2double(data(i+14,32:38))-str2double(data(i+2,32:38));
                    dipole(dind,2) = str2double(data(i+14,40:46))-str2double(data(i+2,40:46));
                    dipole(dind,3) = str2double(data(i+14,48:54))-str2double(data(i+2,48:54));
                    normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                    dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                    dind = dind + 1;
                end
```

```
            %%%Sidechain Arg 1 C-N Bond
            if strcmp(data(i,13:16),' CD ')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'ASP')...
                && ismember(data(i,22),chains)
            %%%Sidechain Asp 2 partial C=0 bonds
            if strcmp(data(i,13:16),' OD1')
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'GLU')...
                && ismember(data(i,22),chains)
            %%%Sidechain Glu 2 partial C=O bonds
            if strcmp(data(i,13:16),' OE1')
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'ASN')...
                && ismember(data(i,22),chains)
            %%%Sidechain Asn1 C=O Bond
            if  strcmp(data(i,13:16),' OD1')
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Asn 2 N-H Bonds
            if  strcmp(data(i,13:16),' ND2')
                dipole(dind,1) = str2double(data(i+5,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+5,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+5,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
```

```
                dipole(dind,1) = str2double(data(i+6,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+6,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+6,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Asn 1 C-N Bond
            if strcmp(data(i,13:16),' CG ')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'GLN')...
                && ismember(data(i,22),chains)
            %%%Sidechain Gln 1 C=O Bond
            if strcmp(data(i,13:16),' OE1')
                dipole(dind,1) = str2double(data(i-1,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i-1,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i-1,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCO*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Gln 2 N-H Bonds
            if  strcmp(data(i,13:16),' NE2')
                dipole(dind,1) = str2double(data(i+7,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+7,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+7,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+8,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+8,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+8,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            if strcmp(data(i,13:16),' CD ')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'HIS')...
                && ismember(data(i,22),chains)
            %%%Sidechain His 2 N-H Bonds
            if  strcmp(data(i,13:16),' ND1')
                dipole(dind,1) = str2double(data(i+8,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+8,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+8,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+11,32:38))-str2double(data(i+3,32:38));
                dipole(dind,2) = str2double(data(i+11,40:46))-str2double(data(i+3,40:46));
                dipole(dind,3) = str2double(data(i+11,48:54))-str2double(data(i+3,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
```

```
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain His 3 C-N Bonds
            if strcmp(data(i,13:16),' CD2')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i-2,32:38))-str2double(data(i-1,32:38));
                dipole(dind,2) = str2double(data(i-2,40:46))-str2double(data(i-1,40:46));
                dipole(dind,3) = str2double(data(i-2,48:54))-str2double(data(i-1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+1,32:38))-str2double(data(i-1,32:38));
                dipole(dind,2) = str2double(data(i+1,40:46))-str2double(data(i-1,40:46));
                dipole(dind,3) = str2double(data(i+1,48:54))-str2double(data(i-1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'LYS')...
                && ismember(data(i,22),chains)
            %%%Sidechain Lys 3 N-H Bonds
            if  strcmp(data(i,13:16),' NZ ')
                dipole(dind,1) = str2double(data(i+2,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+2,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+2,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+3,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+3,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+3,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+4,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+4,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+4,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Lys 1 C-N Bond
            if strcmp(data(i,13:16),' CE ')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+1,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+1,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+1,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end

        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,18:20),'TRP')...
                && ismember(data(i,22),chains)
            %%%Sidechain Trp 1 N-H Bond
            if  strcmp(data(i,13:16),' NE1')
                dipole(dind,1) = str2double(data(i+11,32:38))-str2double(data(i,32:38));
                dipole(dind,2) = str2double(data(i+11,40:46))-str2double(data(i,40:46));
                dipole(dind,3) = str2double(data(i+11,48:54))-str2double(data(i,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
```

```
                dipole(dind,1:3) = dipNH*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
            %%%Sidechain Trp 2 C-N Bonds
            if strcmp(data(i,13:16),' CD1')
                dipole(dind,1) = str2double(data(i,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
                dipole(dind,1) = str2double(data(i+3,32:38))-str2double(data(i+2,32:38));
                dipole(dind,2) = str2double(data(i+3,40:46))-str2double(data(i+2,40:46));
                dipole(dind,3) = str2double(data(i+3,48:54))-str2double(data(i+2,48:54));
                normfactor = sqrt(dipole(dind,1)^2 + dipole(dind,2)^2 + dipole(dind,3)^2);
                dipole(dind,1:3) = dipCN*dipole(dind,1:3)/normfactor;
                dind = dind + 1;
            end
        end
    end
end

core_moment = [sum(dipole(:,1)); sum(dipole(:,2)); sum(dipole(:,3))];
core_moment_mag = sqrt(dot(core_moment,core_moment));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%  Compute surface charge dipole moment of protein %%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
charge = [];
mass = [];

if use_shifted_pKas   %%%%%%%%%%%%%%%%USING SHIFTED PKAS%%%%%%%%%%%%%%%%
    [pKa_TYR,pKa_CYS,pKa_LYS,pKa_ARG,...
     pKa_HIS,pKa_ASP,pKa_GLU,pKa_Nterm,pKa_Cterm] = get_pKaShifts();

    %Ionization fractions
    pH = (0:.1:18);
    a = ones(length(pKa_TYR),1)*pH-pKa_TYR'*ones(1,length(pH));
    ionf_TYR = (10.^(a))./(1+10.^(a));                                          %-1
    a = ones(length(pKa_CYS),1)*pH-pKa_CYS'*ones(1,length(pH));
    ionf_CYS = (10.^(a))./(1+10.^(a));                                          %-1
    a = ones(length(pKa_LYS),1)*pH-pKa_LYS'*ones(1,length(pH));
    ionf_LYS = 1./(1+10.^(a));                                                  % 1
    a = ones(length(pKa_ARG),1)*pH-pKa_ARG'*ones(1,length(pH));
    ionf_ARG = 1./(1+10.^(a));                                                  % 1
    a = ones(length(pKa_HIS),1)*pH-pKa_HIS'*ones(1,length(pH));
    ionf_HIS = 1./(1+10.^(a));                                                  % 1
    a = ones(length(pKa_ASP),1)*pH-pKa_ASP'*ones(1,length(pH));
    ionf_ASP = (10.^(a))./(1+10.^(a));                                          %-1
    a = ones(length(pKa_GLU),1)*pH-pKa_GLU'*ones(1,length(pH));
    ionf_GLU = (10.^(a))./(1+10.^(a));                                          %-1
    a = ones(length(pKa_Nterm),1)*pH-pKa_Nterm'*ones(1,length(pH));
    ionf_Nterm = 1./(1+10.^(a));                                                % 1
    a = ones(length(pKa_Cterm),1)*pH-pKa_Cterm'*ones(1,length(pH));
    ionf_Cterm = (10.^(a))./(1+10.^(a));                                        %-1

    %Count the number of each amino acid
    index = zeros(1,9);
    for i = 1:length(data)
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                strcmp(data(i,18:20),'TYR') && ismember(data(i,22),chains)
            index(1) = index(1)+1;
        end
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                strcmp(data(i,18:20),'CYS') && ismember(data(i,22),chains)
            index(2) = index(2)+1;
        end
        if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
```

```
                    strcmp(data(i,18:20),'LYS') && ismember(data(i,22),chains)
                index(3) = index(3)+1;
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                    strcmp(data(i,18:20),'ARG') && ismember(data(i,22),chains)
                index(4) = index(4)+1;
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                    strcmp(data(i,18:20),'HIS') && ismember(data(i,22),chains)
                index(5) = index(5)+1;
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                    strcmp(data(i,18:20),'ASP') && ismember(data(i,22),chains)
                index(6) = index(6)+1;
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                    strcmp(data(i,18:20),'GLU') && ismember(data(i,22),chains)
                index(7) = index(7)+1;
            end
            %N terminal
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') && ...
                    str2double(data(i,24:26))==aaSTART && ismember(data(i,22),chains)
                index(8) = index(8)+1;
            end
            %C terminal
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OXT') && ...
                    ismember(data(i,22),chains)
                index(9) = index(9)+1;
            end
        end

        ionf = [-ionf_TYR(1:index(1),:); -ionf_CYS(1:index(2),:);    ionf_LYS(1:index(3),:);...
            ionf_ARG(1:index(4),:);  ionf_HIS(1:index(5),:);   -ionf_ASP(1:index(6),:);...
            -ionf_GLU(1:index(7),:);  ionf_Nterm(1:index(8),:); -ionf_Cterm(1:index(9),:)];

        chargepH = sum(ionf);
        [b,a] = min(abs(chargepH));
        iso = pH(a);

        if plotcharge_VS_pH && PH==myPH(1)
            figure(5);
            plot(pH,chargepH)
            title([fname '  iso=' num2str(iso)]);
            xlabel('pH');
            ylabel('Charge (e)');
            hold on;
            plot(0:18,zeros(1,19));
            scatter(iso,0,'o');
        end

        %%%ionization factors for given pH
        if strcmp(PH,'i')
            pH = iso;
        else
            pH = PH;
        end
ionf_TYR = 10.^(pH*ones(1,length(pKa_TYR))-pKa_TYR)./(1+10.^(pH*ones(1,length(pKa_TYR))-pKa_TYR));
ionf_CYS = 10.^(pH*ones(1,length(pKa_CYS))-pKa_CYS)./(1+10.^(pH*ones(1,length(pKa_CYS))-pKa_CYS));
ionf_LYS = 1./(1+10.^(pH*ones(1,length(pKa_LYS))-pKa_LYS));
ionf_ARG = 1./(1+10.^(pH*ones(1,length(pKa_ARG))-pKa_ARG));
ionf_HIS = 1./(1+10.^(pH*ones(1,length(pKa_HIS))-pKa_HIS));
ionf_ASP = 10.^(pH*ones(1,length(pKa_ASP))-pKa_ASP)./(1+10.^(pH*ones(1,length(pKa_ASP))-pKa_ASP));
ionf_GLU = 10.^(pH*ones(1,length(pKa_GLU))-pKa_GLU)./(1+10.^(pH*ones(1,length(pKa_GLU))-pKa_GLU));
ionf_Nterm = 1./(1+10.^(pH*ones(1,length(pKa_Nterm))-pKa_Nterm));
ionf_Cterm = 10.^(pH*ones(1,length(pKa_Cterm))-pKa_Cterm)./...
            (1+10.^(pH*ones(1,length(pKa_Cterm))-pKa_Cterm));

        index = ones(1,9);
```

```
        for i = 1:length(data)

            %%%Find polar side chains
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ OH ’)...
                    && strcmp(data(i,18:20),’TYR’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_TYR(index(1)) * -1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(1) = index(1)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ SG ’)...
                    && strcmp(data(i,18:20),’CYS’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_CYS(index(2)) * -1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(2) = index(2)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ NZ ’)...
                    && strcmp(data(i,18:20),’LYS’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_LYS(index(3)) * 1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(3) = index(3)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ NE ’)...
                    && strcmp(data(i,18:20),’ARG’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_ARG(index(4)) * 1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(4) = index(4)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ NE2’)...
                    && strcmp(data(i,18:20),’HIS’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_HIS(index(5)) * 1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(5) = index(5)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ OD2’)...
                    && strcmp(data(i,18:20),’ASP’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_ASP(index(6)) * -1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(6) = index(6)+1;
            end
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ OE2’)...
                    && strcmp(data(i,18:20),’GLU’) && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_GLU(index(7)) * -1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(7) = index(7)+1;
            end
            %N terminal
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ N  ’)...
                    && str2double(data(i,24:26))==aaSTART && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_Nterm(index(8)) * 1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(8) = index(8)+1;
            end
            %C terminal
            if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,13:16),’ OXT’)...
                    && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_Cterm(index(9)) *-1;   str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
                index(9) = index(9)+1;
            end
            %%%
            %%%Create an array of all atoms with atomic mass
            %%%
          if strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,14),’H’) && ismember(data(i,22),chains)
                mass(:,end+1) = [1.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),’ATOM’) && strcmp(data(i,14),’C’) && ismember(data(i,22),chains)
                mass(:,end+1) = [12.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
```

```
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'N') && ismember(data(i,22),chains)
                mass(:,end+1) = [14.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'O') && ismember(data(i,22),chains)
                mass(:,end+1) = [16.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'P') && ismember(data(i,22),chains)
                mass(:,end+1) = [31.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'S') && ismember(data(i,22),chains)
                mass(:,end+1) = [32.07;                               str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          end
        end
    else                    %%%%%%%%%%%%%%%%NOT USING SHIFTED PKAS%%%%%%%%%%%%%%%%
        %%%Find Isoelectric point
        pH = 0:.001:18;
        switch(model_set)
            case 'Nozaki'
                pKa_TYR = 9.6;
                pKa_CYS = 9.5;
                pKa_LYS = 10.4;
                pKa_ARG = 12.0;
                pKa_HIS = 6.3;
                pKa_ASP = 4.0;
                pKa_GLU = 4.4;
                pKa_Nterm = 7.5;
                pKa_Cterm = 3.8;

            case 'Thurkill'
                pKa_TYR = 9.84;
                pKa_CYS = 8.55;
                pKa_LYS = 10.4;
                pKa_ARG = 12.0;
                pKa_HIS = 6.54;
                pKa_ASP = 3.67;
                pKa_GLU = 4.25;
                pKa_Nterm = 8.0;
                pKa_Cterm = 3.67;
            case 'Cohn'
                pKa_TYR = 10.1;
                pKa_CYS = 9.95;
                pKa_LYS = 10.0;
                pKa_ARG = 12.1;
                pKa_HIS = 6.3;
                pKa_ASP = 3.85;
                pKa_GLU = 4.4;
                pKa_Nterm = 8.0;
                pKa_Cterm = 3.1;
            case 'Creighton'
                pKa_TYR = 10.15;
                pKa_CYS = 9.25;
                pKa_LYS = 10.75;
                pKa_ARG = 12.0;
                pKa_HIS = 6.5;
                pKa_ASP = 3.95;
                pKa_GLU = 4.4;
                pKa_Nterm = 7.4;
                pKa_Cterm = 3.9;
        end

        %Ionization fractions
        ionf_TYR = 10.^(pH-pKa_TYR)./(1+10.^(pH-pKa_TYR));       %-1
        ionf_CYS = 10.^(pH-pKa_CYS)./(1+10.^(pH-pKa_CYS));       %-1
        ionf_LYS = 1./(1+10.^(pH-pKa_LYS));                      % 1
        ionf_ARG = 1./(1+10.^(pH-pKa_ARG));                      % 1
        ionf_HIS = 1./(1+10.^(pH-pKa_HIS));                      % 1
        ionf_ASP = 10.^(pH-pKa_ASP)./(1+10.^(pH-pKa_ASP));       %-1
```

```
        ionf_GLU = 10.^(pH-pKa_GLU)./(1+10.^(pH-pKa_GLU));       %-1
        ionf_Nterm = 1./(1+10.^(pH-pKa_Nterm));                  % 1
        ionf_Cterm = 10.^(pH-pKa_Cterm)./(1+10.^(pH-pKa_Cterm)); %-1

        ionf = [ionf_TYR; ionf_CYS; ionf_LYS; ionf_ARG; ionf_HIS;...
                ionf_ASP; ionf_GLU; ionf_Nterm; ionf_Cterm];

        aanum = [zeros(1,7) 1 -1];
        for i = 1:length(data)
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'TYR') && ismember(data(i,22),chains)
                aanum(1,1) = aanum(1,1) - 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'CYS') && ismember(data(i,22),chains)
                aanum(1,2) = aanum(1,2) - 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'LYS') && ismember(data(i,22),chains)
                aanum(1,3) = aanum(1,3) + 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'ARG') && ismember(data(i,22),chains)
                aanum(1,4) = aanum(1,4) + 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'HIS') && ismember(data(i,22),chains)
                aanum(1,5) = aanum(1,5) + 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'ASP') && ismember(data(i,22),chains)
                aanum(1,6) = aanum(1,6) - 1;
            elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    strcmp(data(i,18:20),'GLU') && ismember(data(i,22),chains)
                aanum(1,7) = aanum(1,7) - 1;
            end
        end

        chargepH = aanum(1,:)*ionf;
        [b,a] = min(abs(chargepH));
        iso = pH(a);
        if plotcharge_VS_pH && PH==myPH(1)
            figure(5);
            plot(pH,chargepH)
            title([fname '  iso=' num2str(iso)]);
            xlabel('pH');
            ylabel('Charge (e)');
            hold on;
            plot(0:18,zeros(1,19));
            scatter(iso,0,'o');
        end
        %%%%%%%%%%%%%%
        %%%ionization factors for given pH
        if strcmp(PH,'i')
            pH = iso;
        else
            pH = PH;
        end
        ionf_TYR = 10^(pH-pKa_TYR)/(1+10^(pH-pKa_TYR));       %-1
        ionf_CYS = 10^(pH-pKa_CYS)/(1+10^(pH-pKa_CYS));       %-1
        ionf_LYS = 1/(1+10^(pH-pKa_LYS));                     % 1
        ionf_ARG = 1/(1+10^(pH-pKa_ARG));                     % 1
        ionf_HIS = 1/(1+10^(pH-pKa_HIS));                     % 1
        ionf_ASP = 10^(pH-pKa_ASP)/(1+10^(pH-pKa_ASP));       %-1
        ionf_GLU = 10^(pH-pKa_GLU)/(1+10^(pH-pKa_GLU));       %-1
        ionf_Nterm = 1/(1+10^(pH-pKa_Nterm));                 % 1
        ionf_Cterm = 10^(pH-pKa_Cterm)/(1+10^(pH-pKa_Cterm)); %-1

        for i = 1:length(data)
            %%%Find polar side chains
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OH ')...
                    && strcmp(data(i,18:20),'TYR') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_TYR * -1;                 str2double(data(i,32:38));
```

```
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' SG ')...
                    && strcmp(data(i,18:20),'CYS') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_CYS * -1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' NZ ')...
                    && strcmp(data(i,18:20),'LYS') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_LYS * 1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' NE ')...
                    && strcmp(data(i,18:20),'ARG') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_ARG * 1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' NE2')...
                    && strcmp(data(i,18:20),'HIS') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_HIS * 1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OD2')...
                    && strcmp(data(i,18:20),'ASP') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_ASP * -1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OE2')...
                    && strcmp(data(i,18:20),'GLU') && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_GLU * -1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            %N terminal
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' N  ') &&...
                    str2double(data(i,24:26))==aaSTART && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_Nterm * 1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            %C terminal
            if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,13:16),' OXT')...
                    && ismember(data(i,22),chains)
                charge(:,end+1) = [ ionf_Cterm *-1;                 str2double(data(i,32:38));
                    str2double(data(i,40:46));     str2double(data(i,48:54));  ];
            end
            %%%
            %%%Create an array of all atoms with atomic mass
            %%%
          if strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'H') && ismember(data(i,22),chains)
                mass(:,end+1) = [1.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'C') && ismember(data(i,22),chains)
                mass(:,end+1) = [12.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'N') && ismember(data(i,22),chains)
                mass(:,end+1) = [14.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'O') && ismember(data(i,22),chains)
                mass(:,end+1) = [16.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'P') && ismember(data(i,22),chains)
                mass(:,end+1) = [31.0;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          elseif strcmp(data(i,1:4),'ATOM') && strcmp(data(i,14),'S') && ismember(data(i,22),chains)
                mass(:,end+1) = [32.07;                                str2double(data(i,32:38));
                    str2double(data(i,40:46));        str2double(data(i,48:54));  ];
          end
        end
```

```
    end

    center_mass = sum(mass(2:4,:).*(ones(3,1)*mass(1,:)),2)/sum(mass(1,:));
    center_charge = sum(ones(3,1)*abs(charge(1,:)).*charge(2:4,:),2)/sum(abs(charge(1,:)));

    if use_center_charge
        surface_charge_moment = sum((charge(2:4,:)-center_charge*ones(1,length(charge)))...
            .*(ones(3,1)*charge(1,:)),2);
    else
        surface_charge_moment = sum((charge(2:4,:)-center_mass*ones(1,length(charge)))...
            .*(ones(3,1)*charge(1,:)),2);
    end

    surface_charge_moment = surface_charge_moment*1.6e-19*1e-10/(3.3356643e-30);
    surface_charge_moment_mag = sqrt(dot(surface_charge_moment,surface_charge_moment));

    dipole_moment = core_moment + surface_charge_moment;
    dipole_moment_mag = sqrt(dot(dipole_moment,dipole_moment));
    if strcmp(myPH,'i')
        disp(['dipole_moment_mag at pH of ' num2str(iso) ':    ' num2str(dipole_moment_mag)]);
    else
        disp(['dipole_moment_mag at pH of ' num2str(PH) ':    ' num2str(dipole_moment_mag)]);
    end


    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%  Plot Backbone of Protein with charge centers and dipole moment     %%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    if (plotbb)&& PH==myPH(1)
        xdata = [];
        ydata = [];
        zdata = [];

        for i = 1:size(bb,1)
            xdata(end+1) = str2double(bb(i,32:38));
            ydata(end+1) = str2double(bb(i,40:46));
            zdata(end+1) = str2double(bb(i,48:54));
        end

        figure(6);   %%%In 3d plane
        for i = 1:length(bbdivider)

            if i ~= length(bbdivider)
                plot3(xdata(bbdivider(i):bbdivider(i+1)-1),ydata(bbdivider(i):bbdivider(i+1)-1),...
                    zdata(bbdivider(i):bbdivider(i+1)-1),'-');
            else
                plot3(xdata(bbdivider(i):end),ydata(bbdivider(i):end),zdata(bbdivider(i):end),'-');
            end
            grid on;
            hold on;
        end

        charge1 = [];
        charge2 = [];
        for i = 1:length(charge)
            if charge(1,i)>0
                charge1(:,end+1) = charge(:,i);
            else
                charge2(:,end+1) = charge(:,i);
            end
        end

        xlabel('X-AXIS [A]');
        ylabel('Y-AXIS [A]');
        zlabel('Z-AXIS [A]');
        if (dispDipoleMoment)
            scatter3(charge1(2,:),charge1(3,:),charge1(4,:),50*charge1(1,:)+10,'+r');
            scatter3(charge2(2,:),charge2(3,:),charge2(4,:),50*abs(charge2(1,:))+10,'k');
```

```
            arrow(center_mass,center_mass+1/100*dipole_moment,'wid',2,'tip',30,'base',70,'len',10)
            text(center_mass(1),center_mass(2),center_mass(3),...
                [' ' num2str(round(dipole_moment_mag)) ' D'])
        end

        figure(2);
        %%%In xy plane
        for i = 1:length(bbdivider)
            subplot(311)
            title(fname);
            ylabel('XY-PLANE(A)');
            if i ~= length(bbdivider)
                plot(xdata(bbdivider(i):bbdivider(i+1)-1),ydata(bbdivider(i):bbdivider(i+1)-1));
            else
                plot(xdata(bbdivider(i):end),ydata(bbdivider(i):end));
            end
            grid on;
            hold on;
        end
        %%%In yz plane
        for i = 1:length(bbdivider)
            subplot(312)
            ylabel('YZ-PLANE(A)');
            if i ~= length(bbdivider)
                plot(ydata(bbdivider(i):bbdivider(i+1)-1),zdata(bbdivider(i):bbdivider(i+1)-1));
            else
                plot(ydata(bbdivider(i):end),zdata(bbdivider(i):end));
            end
            grid on;
            hold on;
        end
        %%%In xz plane
        for i = 1:length(bbdivider)
            subplot(313)
            ylabel('XZ-PLANE(A)');
            if i ~= length(bbdivider)
                plot(xdata(bbdivider(i):bbdivider(i+1)-1),zdata(bbdivider(i):bbdivider(i+1)-1));
            else
                plot(xdata(bbdivider(i):end),zdata(bbdivider(i):end));
            end
            grid on;
            hold on;
        end
    end
    myDipoles(end+1) = dipole_moment_mag;
end
if(plotdipoles_VS_pH)&& length(myPH) > 1
    figure(7)
    plot(myPH,myDipoles)
    title(fname);
    ylabel('Dipole Moment (Debye)');
    xlabel('pH');

    [maxDipole,maxpH_ind] = max(myDipoles);
    hold on;
    plot([myPH(maxpH_ind),myPH(maxpH_ind)],[0 , maxDipole],'r--');
    plot([min(myPH) myPH(maxpH_ind)],[maxDipole maxDipole],'r:');
    scatter(myPH(maxpH_ind),maxDipole,'r');

    z = abs(iso*ones(1,length(myPH)) - myPH);
    if min(z) <= .5
        ind = find(min(z)==z,1);
        plot([myPH(ind),myPH(ind)],[0 , myDipoles(ind)],'r--');
        plot([min(myPH) myPH(ind)],[myDipoles(ind),myDipoles(ind)],'r:');
        scatter(myPH(ind),myDipoles(ind),'r');
    end
end

```

```


function [pKa_TYR,pKa_CYS,pKa_LYS,...
          pKa_ARG,pKa_HIS,pKa_ASP,pKa_GLU,pKa_Nterm,pKa_Cterm]=get_pKaShifts()
    fid=fopen(use_shifted_pKas);
    iii = 0;
    while 1
        iii = iii+1;
        tline = fgetl(fid);
        if ~ischar(tline) || strcmp(tline(1:4),'Site')
            break
        end
        while length(tline) < 50
            tline = [tline ' '];
        end
        if iii == 1;
            data3 = tline;
        else
            data3 = [data3; tline];
        end
    end
    fclose(fid);
    pKa_TYR = [];
    pKa_CYS = [];
    pKa_LYS = [];
    pKa_ARG = [];
    pKa_HIS = [];
    pKa_ASP = [];
    pKa_GLU = [];
    pKa_Nterm = [];
    pKa_Cterm = [];

    if strcmp(use_shifted_pKas(end-9:end-4),'propKa')
        %%%%%%propKa pKas
        for jjj = 1:length(data3(:,1))
            type = data3(jjj,4:6);

            if strcmp(type,'TYR')
                pKa_TYR = [pKa_TYR str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'CYS')
                pKa_CYS = [pKa_CYS str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'LYS')||strcmp(type(1:3),'lys')
                pKa_LYS = [pKa_LYS str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'ARG')
                pKa_ARG = [pKa_ARG str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'HIS')
                pKa_HIS = [pKa_HIS str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'ASP')
                pKa_ASP = [pKa_ASP str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'GLU')
                pKa_GLU = [pKa_GLU str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'N+ ')
                pKa_Nterm = [pKa_Nterm str2num(data3(jjj,13:18))];
            end
            if strcmp(type,'C- ')
                pKa_Cterm = [pKa_Cterm str2num(data3(jjj,13:18))];
            end
        end

    else
        %%%%%H++ pKas
```

```
            for jjj = 1:length(data3(:,1))
                type = data3(jjj,1:5);
                if strcmp(type(1),' ')
                    type = type(2:5);
                    if strcmp(type(1),' ')
                        type = type(2:4);
                    end
                end
                if strcmp(type(1:3),'TYR')
                    pKa_TYR = [pKa_TYR str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'CYS')
                    pKa_CYS = [pKa_CYS str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'LYS')||strcmp(type(1:3),'lys')
                    pKa_LYS = [pKa_LYS str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'ARG')
                    pKa_ARG = [pKa_ARG str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'HIS')
                    pKa_HIS = [pKa_HIS str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'ASP')
                    pKa_ASP = [pKa_ASP str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:3),'GLU')
                    pKa_GLU = [pKa_GLU str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:2),'NT')
                    pKa_Nterm = [pKa_Nterm str2num(data3(jjj,24:end))];
                end
                if strcmp(type(1:2),'CT')
                    pKa_Cterm = [pKa_Cterm str2num(data3(jjj,24:end))];
                end
            end
        end

        if ~length(pKa_TYR)
            disp('No TYR information included');
        end
        if ~length(pKa_CYS)
            disp('No CYS information included');
        end
        if ~length(pKa_LYS)
            disp('No LYS information included');
        end
        if ~length(pKa_ARG)
            disp('No ARG information included');
        end
        if ~length(pKa_HIS)
            disp('No HIS information included');
        end
        if ~length(pKa_ASP)
            disp('No ASP information included');
        end
        if ~length(pKa_GLU)
            disp('No GLU information included');
        end
        if ~length(pKa_Nterm)
            disp('No N-terminus information included');
        end
        if ~length(pKa_Cterm)
            disp('No C-terminus information included');
        end
        pKa_TYR = [pKa_TYR 9.84*ones(1,10)];
        pKa_CYS = [pKa_CYS 8.55*ones(1,30)];
        pKa_LYS = [pKa_LYS 10.4*ones(1,10)];
```

```
        pKa_ARG = [pKa_ARG 7*   ones(1,10)];
        pKa_HIS = [pKa_HIS 6.54*ones(1,10)];
        pKa_ASP = [pKa_ASP 4.0 *ones(1,10)];
        pKa_GLU = [pKa_GLU 4.25*ones(1,10)];
        pKa_Nterm =        10.6*ones(1,10);
        pKa_Cterm =         4.8*ones(1,10);
    end
end
```

Appendix E

Detailed Instructions of a Simple Protein Titration Experiment

What follows is a step-by-step guide to perform a simple dielectric spectroscopy experiment with the proteins β-Lg and HENL. This guide is aimed at newcomers to dielectric spectroscopy who wish to familiarize themselves with some of the basic experimental techniques. By the end of this tutorial, one should be able to reproduce Figure E.1 using the temperature stable dielectric spectrometer described in Section 2.3.3.

E.1 Preparation

1. Prepare $\sim$ 3 ml of 0.1 mM HCl.
2. Obtain from Sigma-Aldrich β-Lg (L3908) and HENL (L6876) protein powders.
3. Using a high precision scale, such as the Cole-Parmer Symmetry Analytical Balance, weight out at least 4 mg of each powder into microcentrifuge tubes.
4. Pipette 0.1 mM HCl into both microcentrifuge tubes to make protein concentrations of 20 mg/ml. For example, if 4 mg of protein powder was weighed out, 200 μl of 0.1 mM HCl should be added.
5. Use a vortex mixer on a high rpm setting to mix the microcentrifuge tubes for at least 20 seconds or until the powders are dissolved.
6. Connect the dielectric cell to the impedance analyzer and connect the plastic tubes from the dielectric cell to the thermal bath. Turn the bath on and set the temperature to 25 °C. Wait for the bath to arrive at the temperature before starting the experiment.
7. Change the following settings on the Agilent 4294A Impedance Analyzer: measurement parameters → Cp-G, sweep type → log, oscillator strength → 500 mV, bandwidth → 5, start frequency → 40 Hz, stop frequency → 110 MHz, number of points → 401.

E.2 Experiment

1. With the cell empty, take at least 3 minutes of sweeps.
2. Pipette 800 μl of 0.1 mM into the cell. Take at least 10 minutes of sweeps.
3. Remove 60 μl of solution and replace with 60 μl of the β-Lg solution (results in 1.5 mg/ml β-Lg). Take at least 10 minutes of sweeps.

4. Remove 60 μl of solution and replace with 60 μl of the β-Lg solution (results in ~ 3 mg/ml β-Lg). Take at least 10 minutes of sweeps.

5. Remove 60 μl of solution and replace with 60 μl of the β-Lg solution (results in ~ 4.5 mg/ml β-Lg). Take at least 10 minutes of sweeps.

6. Remove 60 μl of solution and replace with 60 μl of the HENL solution (results in 1.5 mg/ml HENL, ~ 4.5 mg/ml β-Lg). Take at least 10 minutes of sweeps.

7. Remove 60 μl of solution and replace with 60 μl of the HENL solution (results in ~ 3 mg/ml HENL, ~ 4.5 mg/ml β-Lg). Take at least 10 minutes of sweeps.

8. Remove 60 μl of solution and replace with 60 μl of the HENL solution (results in ~ 4.5 mg/ml HENL, ~ 4.5 mg/ml β-Lg). Take at least 10 minutes of sweeps.

E.3 Data Processing

1. Import the capacitance data from the sweeps into MATLAB.

2. Create two new vectors, `cell_constant` and `parasitic_capacitance`, as follows:

```
E_water = 78.4;
E_air = 1;
cell_constant = (C_water - C_air) / (E_water - E_air);
parasitic_capacitance = C_water - (E_water * cell_constant);
```

3. Create a permittivity matrix, `E_data`, as follows:

```
E_data = (C_data - parasitic_capacitance) / cell_constant;
```

4. Modify `E_data` to create a normalized permittivity matrix, `E_data_normalized`:

```
for i = 1:num_sweeps
   E_data_normalized(:,i) = E_data(:,i) - E_data(end,i);
end
```

5. Plot `E_data_normalized` using mesh.m or other 3D plotting tool.

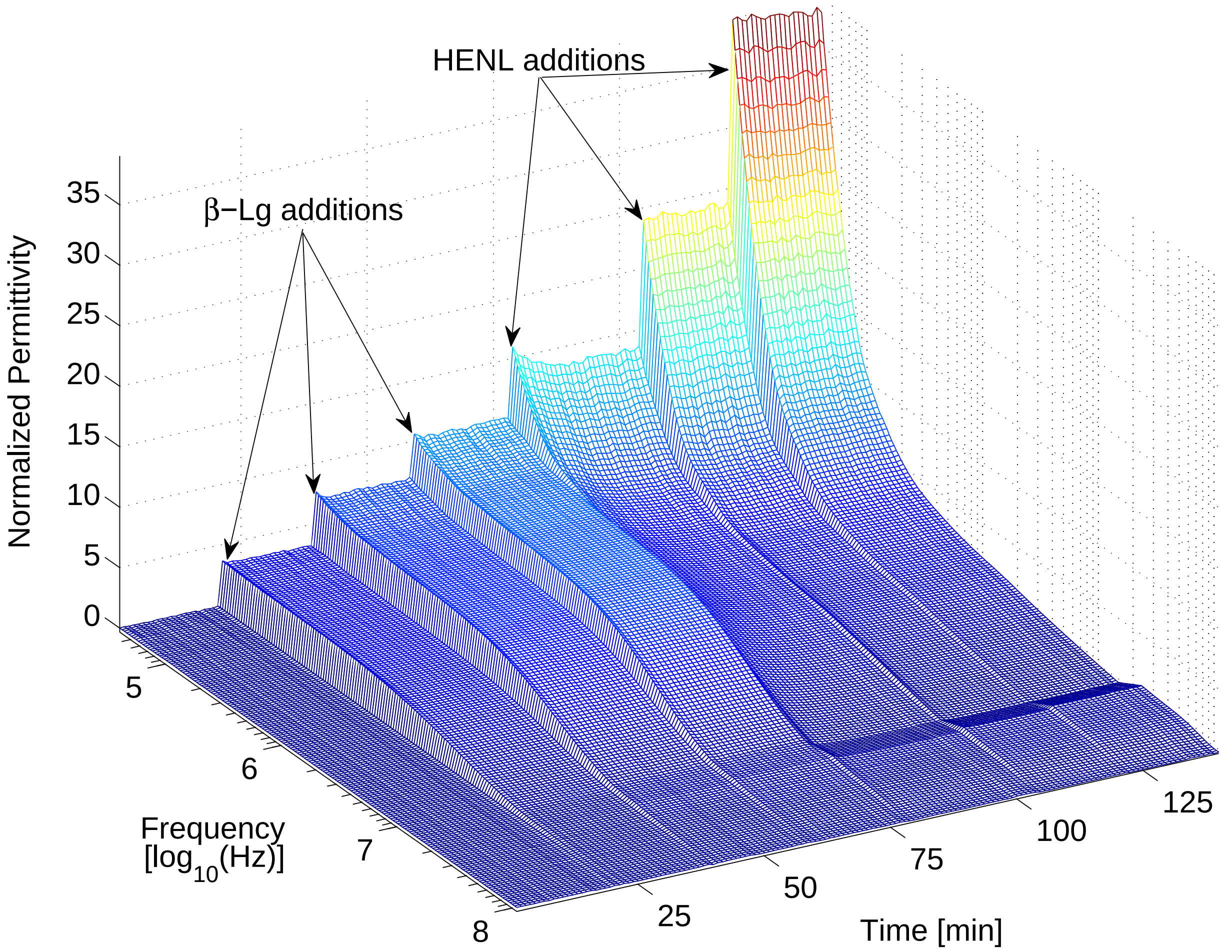


Figure E.1: Dielectric titration of β-Lg and HENL over time at 25 °C. The relaxations of the proteins are visible in this plot. It is noticeable that the time constant of the combined proteins shifts considerably on addition of HENL, corresponding to increased hydrodynamic volume of the complex.

Appendix F

PPy/PSS Preparation and Deposition

F.1 Preparation

1. Obtain Poly(4-styrenesulfonic acid) solution (561223) and Pyrrole reagent grade (131709) from Sigma-Aldrich.

2. Combine the following in a glass container of at least 15 ml:

    (a) 8.008 ml of DI water (8.008 mg)
    (b) 1.85 ml of Poly(4-styrenesulfonic acid) solution (2.053 mg)
    (c) 0.142 ml of Pyrrole

3. Mix the solution for 15 minutes using a magnetic stirrer.

4. Obtain a DC power supply, several alligator clips, and a multimeter.

F.2 Deposition

1. Pipette enough PPy/PSS solution to fill the measurement cell.

2. Attach the cell electrodes to the positive terminal of the power supply using alligator clips.

3. Connect the negative terminal of the power supply to a copper wire. Immerse the copper wire in the solution through the open top hole so that it is equally spaced between both electrode surfaces. The wire may also be placed above both electrodes if there is a chance it may come in contact with either of them.

4. Connect the multimeter in series with the circuit and set it to measure current.

5. Adjust the power supply voltage to allow a current density of 1.5 mA/cm^2. For 1/8" and 1/4" diameter electrodes, this is 0.24 mA and 0.95 mA, respectively. Turn off the power supply after 3 minutes.

6. Remove the remaining solution and rinse the cell with DI water.

Appendix G

U-cell Calibration and Simulation

A new dielectric cell was designed to lessen the electrode polarization effect by increasing the electrode spacing d by a factor of 4 from the design of Section 7.1.1. To facilitate the addition of liquids and limit the cell volume, the base acrylic block was cut in such a way that left a “U” shaped cavity between the two electrodes. This is depicted in Figure G.1. One drawback of the design is the increased amount of liquid needed before saturation due to fringing fields. Simulations with COMSOL Multiphysics were used to determine a volume threshold of $\sim 180\ \mu$l by computing capacitance at various solutions heights (data not shown). The electric field distribution inside the U-cell (Figure G.2) shows that fringing fields are significant outside the area immediately between the electrodes.

Another drawback of the U-cell design is that the non-uniformity of the electric field between the electrodes causes the capacitance to increase nonlinearly with solution permittivity, i.e., Equation 2.72 cannot be applied. Using COMSOL Multiphysics, the permittivity of the sample liquid was varied from 1 to 100 and the capacitance between left and right electrodes was simulated at 1 MHz. Results from this simulation in Figure G.3 show that the U-cell is sufficiently linear above $\varepsilon = 60$ to perform a temperature calibration with water. After cell construction, the cell was filled with DI water and lowered into a glass tempering beaker, which was connected to a thermal bath circulating water at one of three temperatures: 20, 25 or 30 °C. The temperature of the liquid was verified by a thermocouple. Capacitance measurements at these temperatures permit linear calibration as demonstrated in Figure G.3.

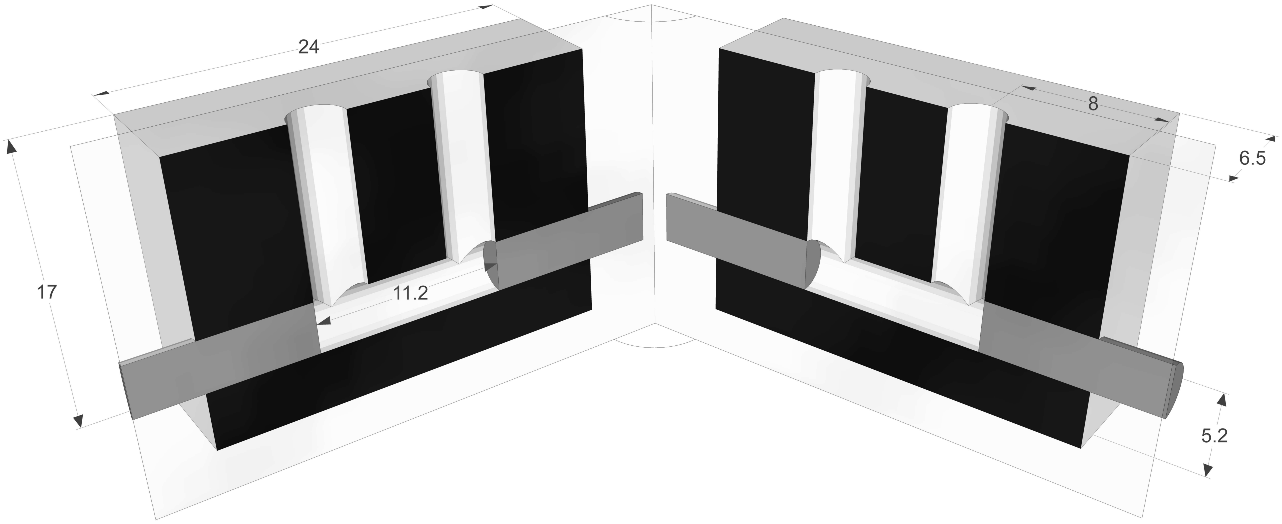


Figure G.1: Cross-sectional view of U-cell. Electrodes are cut from 1/8” stainless steel bar.

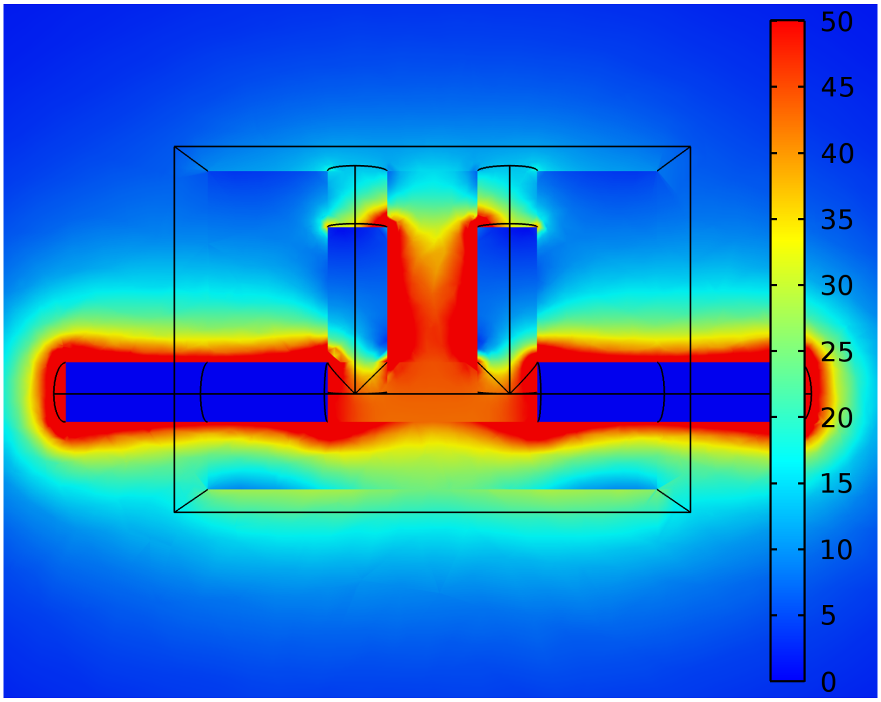


Figure G.2: Simulated electric field distribution inside water-filled U-cell at 1 MHz with a 500 mV voltage difference. Units are in $\mathrm{V \cdot m^{-1}}$.

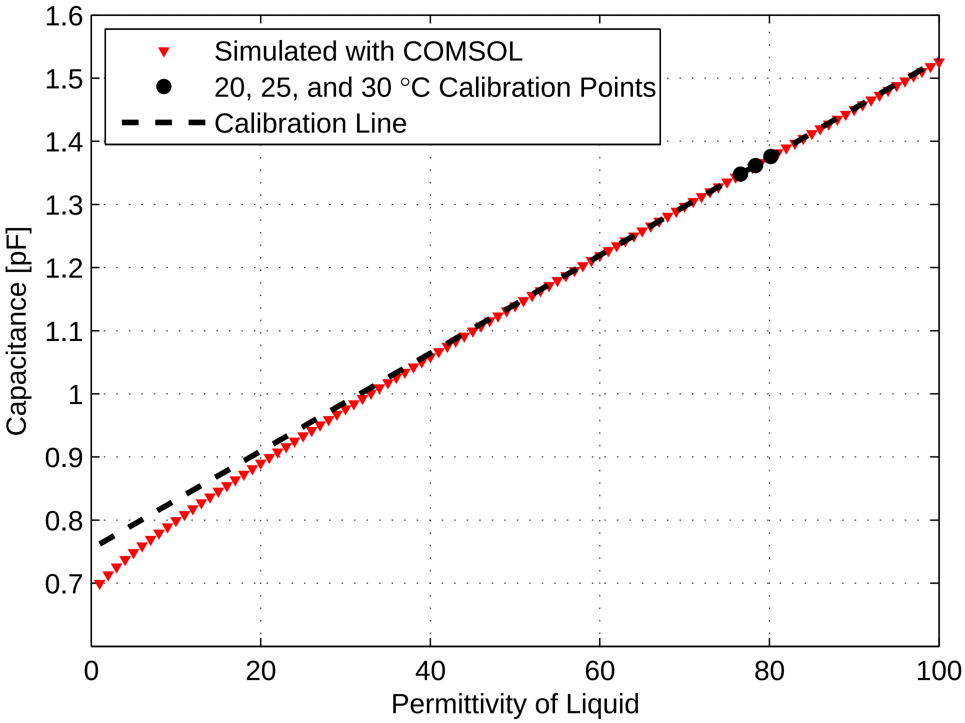


Figure G.3: Simulated capacitance versus solution permittivity of U-cell at 1 MHz alongside the calibration method. Dashed line is the three point temperature calibration line (based on simulated, not actual measured data).