January 2012

Theoretical Circular Dichroism Of Lysozyme, Myoglobin And Collagen And Experimental Circular Dichroism Of Myoglobin And Pea Lectin

Boris Anyangwa Sango

THEORETICAL CIRCULAR DICHROISM OF LYSOZYME, MYOGLOBIN AND COLLAGEN AND EXPERIMENTAL CIRCULAR DICHROISM OF MYOGLOBIN AND PEA LECTIN

By

Boris Anyangwa Sango,
Master of Science, University of Buea-Cameroon, 2007

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Sciences

Grand Forks, North Dakota
December
2012

This thesis, submitted by Boris Anyangwa Sango in partial fulfillment of the requirements for the degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____
Dr Kathryn Thomasson

_____
Dr Evguenii Kozliak

_____
Dr Jerome Delhommelle

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the Graduate School at the University of North Dakota, and is hereby approved.

_____
Dr Wayne Swisher
Dean of the Graduate School

_____
December 06, 2012

PERMISSION

Title Theoretical Circular Dichroism of Lysozyme, Myoglobin and Collagen and Experimental Circular Dichroism of Myoglobin and Pea Lectin

Department Chemistry

Degree Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in her absence by the Chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Boris Anyangwa Sango
October 19, 2012

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the members of my committee, who made improvements to this thesis and who shared valuable knowledge of chemistry with me. I will forever remain grateful to my advisor, Dr. Kathryn Thomasson for the selfless and continued input to my life as a student and a person. I also wish to thank Drs. Evguenii Kozliak and Jerome Delhommelle, who not only gave valuable suggestions during my oral progress reports but also guided my during course work in answering questions challenging questions that my had. I would also like to thank previous and current members of the Thomasson research group – Dr. Neville Forlemu, Dr. Eric Njabon, Dr. Victor Waingeh, Kristine Carlson, Sandeep Pothuganti, and Rahul Nori who in one way or another helped a great deal in my success as a graduate student. I would also like to thank friends the likes of but not limited Kenneth Ndumbe Ngale, Gregory Fondong, Patrick Tamukong, Dr. Chrsitian Jungong, Serges Tatsinkou Nguelo, Dr. Edmund Njua, Dr. Tsvetan Aleksandrov, Jason and Erica Hicks, among others with whom I attended seminar practice sessions and some of whom helped a lot in formatting this thesis. I cannot forget to thank my lovely wife Zinata Dienge Sango who spent sleepless nights sometimes just so that I could do academic work and live a better life. Many thanks go to my family members especially my dad, Mathew Sango and my uncle Harrison Abinteh for providing me with the moral and financial support I needed all these years I have lived. Finally all thanks to God Almighty for life, love, joy and salvation.

ABSTRACT

The need to understand protein structure and interaction is ever-growing and this has led to scientist and investigators utilizing numerous different techniques in order to obtain substantial insights and explanation to these structures and their interactions. Circular dichroism (CD), which is one of these techniques, is a powerful structural biology technique used to study protein and nucleic acid structures and their dynamics. This technique is important because it identifies the secondary, tertiary and even quaternary structures in proteins and can be used to study folding patterns in proteins. Theoretical and experimental methods are used to better understand and teach the phenomenon of circular dichroism. First, molecular mechanics allows for the energy calculations of different conformations in large molecules like peptides and proteins. Theoretical CD via the dipole interaction model (DInaMo) is used to relate the structural nature of peptides and proteins to the experimental CD observed. Minimization was done on lysozyme, myoglobin and collagen, using the molecular modeling software package, Insight®II to obtain minimum energy structures suitable for CD calculations. Molecular dynamics simulations were performed in water at 300K to create an ensemble of conformations. The program CDCALC was used to predict the CD spectrum of the proteins for comparison with experiment. The output from CDCALC was analyzed using OriginPro Version 7.5 and the analyzed data reported as plots with data from synchrotron radiation circular dichroism (SRCD). Theoretical CD plots showed agreement with SRCD for location, sign, and bandwidths of the peaks. Experimental CD spectra of horse heart myoglobin and pea lectin were measured on a JASCO spectropolarimeter and compared to that obtained from the Protein Circular Dichroism Data Bank (PCDDB) in the development of a physical chemistry laboratory exercise to teach secondary structure analysis.

CHAPTER 1

INTRODUCTION

1.1. Background

There are several methods that have been used to determine protein structure. Some of these include X–ray crystallography, nuclear magnetic resonance, electron microscopy and circular dichroism (CD). Of these methods, circular dichroism has advantages over the other methods in that it can be done both in solution and on all kinds of proteins.[1] Secondary structure determination is key to understanding the folding and dynamics of proteins. Circular dichroism is a very useful tool in such determination.[2] Protein secondary structure analysis involves analyzing proteins for the presence of different types of protein motifs. This helps in the refinement of protein structural models as well as rapid conformational changes in proteins that result from either ligand binding or macromolecular interaction.[2] This analysis can be done using a variety of different software packages that run on different servers around the world. One of these servers is DichroWeb,[3] which has been and is still used extensively to analyze protein circular dichroism data.

1.2. Circular Dichroism

Circular dichroism (CD) is one of many types of spectroscopies used for studying molecules. It is the difference in absorption between left and right circularly polarized

light when passed through a chiral molecule.[2] This can be expressed mathematically in Equation [1]

$$\Delta\varepsilon\ (\lambda) = \varepsilon_L\ (\lambda) - \varepsilon_R\ (\lambda) \qquad [1]$$

where ε_L and ε_R are the extinction coefficients for the left and right circularly polarized lights at wavelengths (λ) respectively (Figure 1).[2]

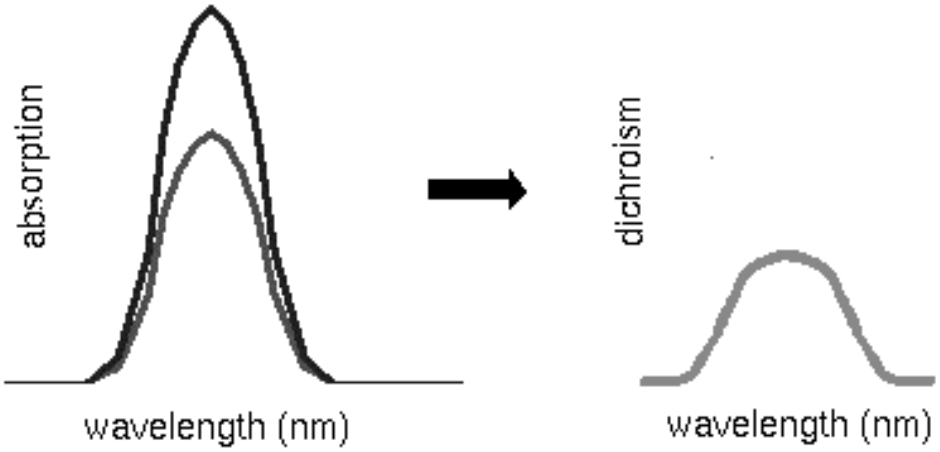


Figure 1. Circular Dichroism is a Difference in Absorption of Circularly Polarized Light. The band on the right shows CD, while the bands on the left represent left–handed (upper) and right–handed (lower) absorption for circularly polarized light through a molecular system.[2]

1.3. Working Principle of Circular Dichroism

Biological systems are often asymmetrical because of the differences in the chemical groups that attach to the central carbon. This is common in amino acids, which form proteins with four non–identical groups attached to the central carbon. This makes biological systems chiral such that when left and right circularly polarized light is passed through them, they interact with each component of the light to different extents. Their

ability to absorb light makes them to be referred to as chromophores. The difference that exists between these interactions is the basis for circular dichroism. This makes CD a powerful tool for examining protein conformation and dynamics, as well as for determining protein secondary structure.[4]

1.4. Classification of Protein Secondary Structure

The classification of protein structure makes use of four categories namely primary, secondary, tertiary, quaternary and quinary. The primary structure defines the amino acid sequence of the protein while the secondary structure defines specific kinds of folds that the primary sequence adopts using mostly hydrogen bonds. Tertiary structure defines a complete three–dimensional representation of the protein structure. Quaternary structure is used with some proteins that have more than one amino acid chain, while quinary structure is used to define the structure of an assembly of two different proteins interacting. CD is mostly used for examining secondary structure, as well as understanding the folding and unfolding (tertiary structure) of proteins. Spectra obtained in the far ultraviolet wavelength range (around 190 nm – 250 nm) give information on the polypeptide backbone conformation of proteins. However, when the vacuum ultraviolet (VUV) region (below 190 nm) is considered, further information can be obtained that provide more detail of protein secondary structure.[5]

1.5. Advantages of Circular Dichroism

Even though other techniques such as X–ray crystallography and nuclear magnetic resonance (NMR) exist for examining protein secondary structural elements, circular dichroism provides salient advantages that make it very useful as a method of choice. These advantages include that fact that it is non–destructive, simple and requires minute quantities of the sample.[1] Unlike X–ray crystallography, which is valuable for providing the detailed three–dimensional protein structure found in the protein databank, experimental CD does not require high–quality crystal structures; it requires a much smaller sample size and can be done in solution. Also, X–ray crystallography only provides a static structure while CD can provide the dynamics and structural fluctuations that explain the physiological functions of proteins. Both solid state and solution NMR have also been used to successfully determine protein structural details.[6] However, NMR requires high protein concentrations and is limited to small sized proteins. This is because with large proteins, there is need for multi–isotope labeling to assign huge resonance numbers, which makes the accuracy and application of NMR limited.[1]

1.6. Types of Circular Dichroism

CD can be classified as either theoretical CD or experimental CD.[5] Theoretical CD involves the prediction of CD by use of computer codes. These predictions are always done and compared to experimental CD data, which are obtained from either a conventional CD instrument that uses circularly polarized light or using a CD instrument that uses synchrotron radiation.[5]

1.6.1. Theoretical Circular Dichroism

Many methods, both theoretical and experimental have been used to study of proteins stability, dynamics and folding/unfolding. Some of these are discussed in this chapter.

Feng *et al.,* used circular dichroism to study the stability of the mini–protein FSD–1. In their study, measured the thermal unfolding of this protein using CD by following changes in CD. Even though their results did not show a fully folded baseline, they observed a broad transition with baselines very less clearly defined.[7] However, Honda *et al.,* have successfullyfollowed the thermal unfolding of a 10 residue mini–protein (CLN025), reporting a well–defined, folded–state baseline.[8] Feng *et al.,* also simulated the unfolding studies using replica exchange molecular dynamics (REMD). This would provide atomic details on the unfolding process with respect to temperature. [7]

Conformational studies to show aggregation in proteins have also been done in the past. Nath *et al.,* studied the aggregation of Synuclein using CD and recorded CD spectra that showed the formation of β–sheets after 6 hours of incubation at a concentration of 100 μM at 45 ^{o}C.[9] Greenfield has also used circular dichroism as a function of temperature to study the thermodynamics of protein unfolding and binding interactions. The study revealed the effects of mutations and ligands on protein and polypeptide stability, determine the intermediates involved in the unfolding process as well as the van't Hoff enthalpy and entropy of unfolding and the free energy of unfolding.[10] Greenfield also studied the folding of proteins as a function of denaturants,

osmolytes and ligands using circular dichroism. The study showed CD bands that were induced as result of folate binding to dihydrofolate reductase.[11]

Stjernschantz and Oostenbrink have used multiple binding modes to predict improved ligand–protein binding affinity using thiourea–containing compounds.[12] They used GROMACS[13] for the simulation with the force field GROMOS chosen. They successfully proposed to calculate weighted ensemble averages from multiple MD simulations and they reported obtaining better models in the binding studies when they used starting conformations from docking experiments than when they constructed the models manually.[12] Prediction of the stability of collagen as a result of contributions from salt bridges has been reported by Gurry *et al.* They did molecular dynamic simulations on collagen peptides using CHARMM[14] and reported triple–helical stability in the form of asymmetric interactions between chains of collagen. Kresten *et al.,* have also studied how fast–folding proteins fold using molecular dynamics.[15] In their study they did atomic–level molecular dynamic simulations over periods of ranging between 100 μs and 1 ms and reported spontaneous and repeated folding on twelve proteins representing the three major structural classes.[15]

Both theoretical and experimental circular dichroism have been used alongside Fourier transform infrared, and magnetic resonance spectroscopies to predict the folding and unfolding of hen egg–white lysozyme from molecular dynamics.[16] In this study, Meersman *et al.,* found that both theoretical and experimental CD successfully indicated reversible folding and unfolding of of hen egg–white lysozyme through a loss of β–structures followed by a reversible loss of the protein helical content from approximately 40% to 27% at 70^{o}C and approximately 20% at 77^{o}C.[16] Finke *et al.,* have also used CD

as well as steady–state and time–resolved Fluorescence resonance energy transfer (FRET) to study the equilibrium unfolding of the poly(glutamic acid)$_{20}$ helix.[17]

1.6.2. Experimental Circular Dichroism

Experimental CD is classified into two types namely conventional circular dichroism (cCD) and synchrotron radiation circular dichroism (SRCD).[18] Conventional CD makes use of circularly polarized light and measures and provides CD in the far UV region of the electromagnetic spectrum (typically from ~190 nm to 250 nm). It involves the use of a lab–based CD instrument, which limits the quality of information in the spectra by that fact that there is high absorption of the sample, buffer and solvents, and the intensity of the light source is low.[18] SRCD, however, makes use of a synchrotron radiation source that provides inherently linearly polarized light, at higher photon fluxes, particularly at shorter wavelengths than does conventional CD. This means that SRCD provides for collection of spectra data at lower wavelengths (150 – 190 nm) with much higher signal–to–noise ratio.[18] Not only does this provide extensive information on the secondary structure of the protein (information regarding turn structures and unordered structures in the protein), SRCD can also be used for examining protein samples in the presence of high buffer concentrations as well as other absorbing species.[18]

1.7. Objectives

Having discussed previous applications of circular dichroism in studying protein stability, protein folding and unfolding and aggregation, the rest of this thesis will focus

on the application of circular dichroism to study protein structure. This will involve using theory and experiment to study the structural elements found in some proteins and what conclusions can be drawn from the studies to better understand the proteins.

1.8. Analysis of Protein Secondary Structure

The analysis of protein structure using data obtained from circular dichroism is based on the fact that peptides and proteins that have different conformations do produce different spectra with different characteristics.[19] These spectra are representative of different types of secondary structure. These include α–helices, β–sheets, and random coils. The spectra come from different $\phi$and φ angles, which result in two major types of transitions occurring in the proteins, which are namely $n - \pi^*$ and $\pi - \pi^*$ transitions.[19] These transitions occur in the far ultraviolet region of the electromagnetic spectrum producing spectral peaks centered at different wavelengths and of different magnitudes, which are dependent on the backbone angles.

Different methods are useful for analyzing secondary structures of proteins based on circular dichroism data. These methods can be classified into three major categories namely visual inspection, empirical analysis, and *ab initio* analysis.[19]

Visual inspection is the most basic of all three methods and involves visual observation of the peaks in the CD spectra. Even though it is the most basic, it still can adequately quantitatively describe or identify proteins secondary structure mostly especially those that only have one kind of secondary structure.[19] This involves identifying the type of peaks if they are positive or negative and locating the wavelength

of the peaks. Generally, α–helices produce spectra with a large positive peak around 190 nm and a small negative peak around 208 nm both of which are a result of $\pi - \pi^*$ transitions. There is also a small negative peak around 222 nm that is a result of $n - \pi^*$ transitions.[19-20] β–sheets produce a small positive peak (20% the magnitude of the helical structure and red–shifted) and also a single negative peak at higher wavelengths.[20b] Spectra produced by poly–Proline II structures such as those in collagen have a large negative peak around 200 nm and a small positive one around 210 nm.[19, 20b]

Empirical analysis involves the use of linear and non–linear algorithms developed using reference databases derived from proteins of known structures. With this type of analysis, a large number of algorithms and reference databases are used to empirically analyze the percentage composition of α–helical, β–sheets, as well as other type of secondary structures obtained from circular dichroism data.[21] Empirical analysis can be done using least squares, neural network, principal component analysis, rigid regression, variable selection, and singular value deconvulation.[19] These are methods are employed in different algorithms that have been developed.

1.9. DichroWeb

DichroWeb[3, 22] is an interactive website established for the analysis of protein secondary structure from CD spectra. In order to use it, the individual has to register for an account on the server which can subsequently accessed via a login ID and password. This method of analysis has the advantage that in eliminates the need for software installations and even hardware compatibility issues that may arise with installation. Data

obtained from CD instruments such as the JASCO and Aviv are usually stored in ASCII text format. Any other ASCII formatted file can be used in DichroWeb and these could include output files from the Daresbury Synchrotron Radiation Circular Dichroism instrument[23], and both YY and BP output formats from the data processing program SUPER3.[24] Other file formats such as binary files cannot be used and if needed for analysis must be converted to text format using the manufacturer's software. Nonetheless, it accepts input in the following units; mean residue ellipticity, molar ellipticity, delta epsilon, and theta (machine units).[25] Analysis is done using the aforementioned algorithms and it also uses a wide range of reference databases.[26]

The reference databases are a set of data obtained from proteins of known secondary structure to which the secondary structure of other proteins can be determined through analysis. They are very vital in determining the quality of any empirical analysis. The major practical basis for choosing a reference database in the analysis of the secondary structure of a protein is largely dependent on the similarities that exist between the proteins within the database and that being analyzed. Usually, if there are structural differences between the protein being analyzed and those of the reference database being used, the analysis is less likely to be correct but if they are similar the analysis is better chanced at being successful.[19] With most analyses, the secondary structure of the protein is usually unknown and so there is absolute need to test different reference databases and to select the one that gives the best fit and a sum 1.0 for secondary structural elements using regression analysis. The good–of–fit parameter is used to determine this fit and it assesses the quality of the analysis by determining the correspondence between the experimental data and the back–calculated spectrum arising from the "best" calculated

secondary structure for the analysis.[19] This parameter is the normalized root mean square deviation (NRMSD)[27] and is defined mathematically as

$$NRMSD = \sum \left[\left(\theta_{exp} - \theta_{cal} \right)^2 / \left(\theta_{exp} \right)^2 \right]^{1/2} \quad [2]$$

where θ_{exp} and θ_{cal} are respectively the experimental ellipticities and the ellipticities of the back–calculated spectrum for the derived structure summed over all wavelengths. Generally, when the NRMSD is less than 0.10, there is good fit and hence successful analysis, when it is greater than 0.10 or less than 0.20, the calculated secondary structure is expected to be similar to that found in the protein being analyzed, and finally, when it is greater than 0.20, the analysis is considered to have failed and the secondary structure is likely incorrect.[19] With DichroWeb, some of these algorithms include SELCON 3,[28] VARSLC,[29] CDSSTR,[26, 29b] and K2D.[30]

CHAPTER 2

METHODS OF PREDICTING CIRCULAR DICHROISM

2.1. Theoretical Circular Dichroism

Structures used for the theoretical calculations were obtained from the RCSB Protein Data Bank (PDB).[31] The PDB is a database containing protein structures deposited from electron microscopy, X–ray crystallography, as well as Nuclear Magnetic Resonance (NMR). The files uploaded contain valuable information about the protein being deposited and could be anything from the name of the protein, number of chains and/or subunits it contains, resolution of the structure, organism from which it was obtained, authors who did the work, to about all the technical remarks about the protein structure that anyone intending to use the structure may need to know.

2.2. Dipole Interaction Model

Theoretical CD employs the Dipole Interaction Model (DInaMo). Theoretical CD via DInaMo was pioneered by Jon Applequist.[32] This model consists of N units that interact with each other by way of the fields of their induced electric dipole moments in the presence of a light wave.[32-33] A unit may be an atom, a group of atoms, or a whole molecule. For peptides and proteins, it is the amide group NC'O. Unit $\boldsymbol{i}$, located at position $\boldsymbol{r_i}$ has polarizability α_i and induced dipole moment μ_i.[33] $\boldsymbol{E_i}$ is the electric field at $\boldsymbol{r_i}$ due to the light wave.[33] This interaction is expressed in the Equation 2 below

$$\mu_i = \alpha_i \left[E_i - \sum_{j=1}^{N} T_{ij} \mu_j \right] \qquad [3]$$

where $\boldsymbol{T_{ij}}$ is the dipole field tensor, which is a function of the positions of the two dipoles $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$.[33]

The matrix form of the system of equations represented by Equation 2 becomes

$$A\mu = E \qquad [4]$$

where $\boldsymbol{\mu}$ is a column vector of the moments $\boldsymbol{\mu_i}$, $\boldsymbol{E}$ is a column vector of the fields $\boldsymbol{E_i}$, and the square interaction matrix $\boldsymbol{A}$ contains the coefficients in Equation 2.[33-34] The solution to Equation 3 is

$$\mu = BE \qquad [5]$$

where $\boldsymbol{B} = \boldsymbol{A^{-1}}$.[33] Optical properties are determined by Equation 4 using the coefficients of the various field terms.[33] The optical absorption and dispersion phenomena are expressed most easily in terms of normal modes of the system of coupled dipole oscillators.[33, 35] Unit *i* has a number of dipole oscillators that are indexed by *is* with polarizability $\boldsymbol{\alpha_{is}}$ along a unit vector $\boldsymbol{u_{is}}$.[33, 35] Band shapes are assumed to be Lorentzian so that the dispersion of an isolated oscillator is represented by a Lorentzian function having wavenumber $\overline{\eta}_{is}$ with a half–peak bandwidth of Γ

$$\alpha_{is} = \frac{D_{is} u_{is} u_{is}}{\overline{\nu}_{is}^2 - \overline{\nu}^2 + i\Gamma\overline{\nu}} \qquad [6]$$

D_{is} represents a constant related to the dipole strength and $\bar{\eta}$ is the vacuum wavenumber of the light.[33] Equation 3 reduces to an eigenvalue problem where the eigenvalues of $\boldsymbol{A^o}$ (the $\boldsymbol{A}$ matrix at $\bar{\eta}$ = 0) are a set of squares of normal mode wavenumbers $\bar{\eta}_k^2$ and the normalized eigenvectors $\boldsymbol{t^{(k)}}$ are column vectors whose components are the relative amplitudes of the dipole moments of the oscillators.[33] Relative amplitudes of the electric dipole moment $\boldsymbol{\mu^{(k)}}$ and magnetic dipole moment $\boldsymbol{m^{(k)}}$ for the system in the *k*–th normal model are given by

$$\mu^{(k)} = \sum_{is} t_{is}^{(k)} u_{is} \qquad [7]$$

$$m^{(k)} = \sum_{is} t_{is}^{(k)} r_i \times u_{is} \qquad [8]$$

Dipole strength D_k and rotational strength R_k associated with the *k*–th normal mode are expressed as

$$D_k = \mu^{(k)} \bullet \mu^{(k)} \qquad [9]$$

$$R_k = \mu^{(k)} \bullet m^{(k)} \qquad [10]$$

If any of the natural wavenumbers $\bar{\eta}_{is}$ are far above the spectral region of interest, the corresponding oscillators are approximately nondispersive. The normal model problem can be simplified by partitioning the $\boldsymbol{A_o}$ matrix into blocks.[33, 35]

$$A^o = \begin{pmatrix} A_{11}^o & A_{12} \\ A_{21} & A_{22}^o \end{pmatrix} \qquad [11]$$

The A_{11}^o block contains the coefficients relating the dispersive oscillators to each other (i.e., the chromophoric part of the system), the A_{22}^o block contains the

nondispersive oscillators (i.e., the nonchromophoric part of the system), and the A_{12} and the A_{21} blocks contain the interactions between the two subsystems.[33, 35] The normal modes in the spectral region of interest (e.g., far UV for proteins) are those of the matrix

$$A_{11}^{o} - A_{12}\left(A_{22}^{o}\right)^{-1} A_{21} \qquad [12]$$

This means the order of the eigenvalue problem significantly is smaller than the full matrix *A*.[33] The advantage in computational efficiency is substantial in systems with only a few dispersive oscillators and many nondispersive oscillators.[33] For example, a small protein like lysozyme has 128 dispersive oscillators representing the amide groups in the backbone while all other atoms including the hydrogens are treated as nondispersive (1037 units).

Absorption molar extinction coefficient ε and circular dichroism $\Delta\varepsilon$ at each wavenumber are calculated as sums over the Lorentzian bands for all normal modes.[20a]

$$\varepsilon = \frac{8\pi^2 \bar{\nu}^2 N_A \Gamma}{6909 p} \sum_{k}^{q} \frac{D_k}{\left(\bar{\nu}_k^2 - \bar{\nu}\right)^2 + \Gamma^2 \bar{\nu}^2} \qquad [13]$$

$$\Delta\varepsilon = \frac{32\pi^3 \bar{\nu}^3 N_A \Gamma}{6909 p} \sum_{k}^{q} \frac{R_k}{\left(\bar{\nu}_k^2 - \bar{\nu}^2\right)^2 + \Gamma^2 \bar{\nu}^2} \qquad [14]$$

where N_A is Avogadro's number and p is the number of peptide residues, which is equal to q herein because there is one dispersive oscillator for each amide π–π* transition of which there is only one per residue.[20a] It is possible to have more dispersive oscillators per peptide (e.g., for the n–π* transition), but the current parameters used with DInaMo

are only coarse enough to produce normal modes in the correct region for the amide n–π* transition, but not necessarily the correct sign.

Four PDB files for three different proteins were used for this study namely: hen egg white lysozyme (2VB1),[36] horse heart myoglobin (3LR7),[37] sperm whale myoglobin (2JHO),[38] and a collagen triple helix model (1K6F).[39] High quality structures were necessary to predict CD for each of these proteins. Usually, files deposited in the PDB website from X–ray crystallographic studies do not contain hydrogen atoms. For this reason, hydrogen atoms were added to the PDB files since they are needed for CD calculations. They were added directly using a command from insight II™ that adds missing hydrogen atoms.

2.3. Structure Generation

The files downloaded from the PDB databank were not in a state to be used directly in insight II™. For this reason, there was need for structure generation. In this process, the downloaded PBD files were hauled into insight II any issues that needed fixing such as adding hydrogen atoms and correcting bond chemistry was adjusted with commands in the program.

The resulting protein structures were energy minimized with a strategy that composed of first minimizing using steepest descents and later using conjugate gradients. The aforementioned PDB files were selected for two reasons: (1) each was a high-resolution structure. (2) The structures chosen were the same species for which Synchrotron Radiation Circular Dichroism (SRCD) was available in the Protein Circular

Dichroism Data Bank (PCDDB).[31] Collagen was the only exception because there its SRCD was digitized from Wallace and Janes, 2001.[5]

2.4. Energy Minimization

Because the structures are obtained without hydrogen atoms and are rigid based on the nature of X–ray crystallography, they have to be energy minimized to obtain stable structures with the lowest possible energies. X-ray structures are known to have short bond lengths and the dipole interaction model is sensitive to them. This is vital because the code CDCALC, used for calculation the circular dichroism of the structures is very sensitive to structural changes and will output different spectra for different structural conformations. Each protein structure was minimized using the force field CVFF (Consistent–Valence Force Field)[40] within the Discover module of Insight®II (San Diego, CA). This force field is the original force field produced with the Discover program and contains parameters for amino acids, water, and a variety of functional groups.

The functional of the force field takes the mathematical form expressed in Equation 15.

$$V = \sum\left\{D_b\left[1 - e^{-\alpha(b-b_0)}\right]^2 - D_b\right\} + \frac{1}{2}\sum H_0(\theta - \theta_0)^2 + \frac{1}{2}\sum H_\theta(1 + s\cos n\theta) +$$

$$\frac{1}{2}\sum H_\chi \chi^2 + \sum\sum F_{bb'(b-b_0)(b'-b'_0)} + \sum\sum F_{\theta\theta'}(\theta - \theta')(\theta' - \theta'_0) +$$

$$+ \sum\sum F_{b_\theta}(b - b_0)(\theta - \theta_0) + \sum F_{\phi_{\theta\theta'}} cos\phi(\theta - \theta_0)(\theta' - \theta'_0)$$

$$+ \sum\sum F_{\chi\chi'\chi\chi'} + \sum \frac{A}{r^{12}} - \frac{B}{r^6} + \sum \frac{q_i q_j}{r} \qquad [15]$$

Where the first four terms represent diagonal terms, the fifth to the ninth terms represent off–diagonal terms or cross terms and the eleventh and twelfth terms represent non–bonded interactions. For our simulation, we limited the functional to the first four terms and did not include the cross terms. The first four terms (diagonal terms) can be described as follows; terms due to deformation of bond length (bond stretching), bond angles, torsion angles and out–of–plane interactions respectively. The fifth to the ninth terms (off–diagonal terms or cross terms) can be described as follows; coupling between stretch of adjacent bond, angles, bond and angles, torsion angles and out–of–plane interactions respectively. Terms eleven and twelve (non–bonded interactions) describe the van de Waals interaction (with a Lennard–Jones potential) and the Coulombic interactions respectively. The parameters D_b, H_0, $H_{\square}$, H_{χ}, F_{ij} are the force constants for the intramolecular deformations, and q_i and q_j are the partial charges carried by each atom of the molecule in question. V is the potential energy that defines the internal degrees of freedom and non–bonded distances. CVFF is a generalized force field that contains parameters for amino acids, water and other functional groups.[41]

A constant dielectric of 86.32 (for water at 4ºC) was used. This was used because the corresponding SRCD data with which the computed CD was compared was also obtained at the same temperature.[5, 31] In the minimization process, two different strategies were employed. These include a strategy of steepest descents followed by conjugate gradients. Generally, energy minimization is based on finding the local minimum in the energy profile of the protein molecule. It does not find the global minimum high-energy barriers exist on the potential energy surface. The method approaches the local minimum in a zig–zag path (step–wise fashion) where consecutive search directions are orthogonal

to each other.[42] In the process, the total energy of a conformation the protein is evaluated from the force field. Adjusting atomic coordinates of the atoms such that the energy of the molecule is finally lowered follows this. These adjustments occur in steps as seen in Figure 2. There are two algorithms for minimizing protein structures; steepest descents and conjugate gradient algorithms. With steepest descents, bad contacts that may occur in the starting structure (high energy conformation) are removed such that the steps are adjusted toward the negative end of the energy profile. It therefore improves the rather poor starting conformation to a better one, which can be treated with the conjugate gradient algorithm. Also this method usually does not converge. The step can be expressed as follows:

$$x_{\kappa+1} = x_{\kappa} \nabla V(x_{\kappa}) = x_{\kappa} - \lambda_{\kappa} g(x_{\kappa}) \qquad [16]$$

where $g(x_k)$ is the gradient at a given point. At the local minimum (that is, the point at which $V(x)$ is minimum, the directional derivative has to be zero.[43]

Therefore, the directional derivative is given by:

$$d/d\lambda_{\kappa} V(x_{\kappa+1}) = \nabla V_{x_{\kappa+1}} T \, d/d\lambda_{\kappa} x_{\kappa+1} = -\nabla V(x_{\kappa+1}^{T}) g(x_{\kappa}) \qquad [17]$$

A representation of the scheme is shown in Figure 2.

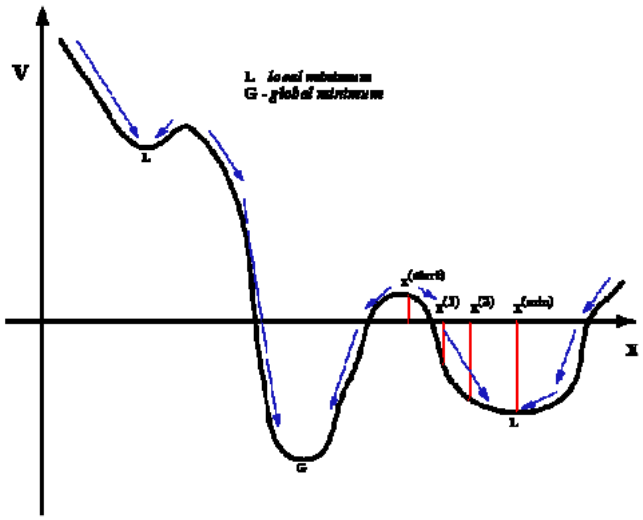


Figure 2. Representation of the energy landscape of a protein. This is a profile of potential energy (V) of the system with respect to atomic coordinate position (x).[2]

With conjugate gradient, a history of previous minimization is used to further refine the downward stepwise atomic adjustment. In this case, however, the adjustments are achieved by creation of a basis set of conjugate directions so that after a step, the adjustments are made so that the movement is conjugate to the previous one. With such an approach this algorithm increases the efficiency of minimization such that there is convergence when an appropriate number of conjugate steps of adjustments have been made. Equation 18 shows mathematically how conjugate the gradient adjustment is achieved.

$$h_{n+1} = g_{n+1} + \lambda_n h_n \qquad [18]$$

For each strategy, different numbers of steps were included for each protein structure. The number of steps used for steepest descents were as follows; 900,000 steps for lysozyme, 110,000 steps for both horse heart and sperm whale myoglobins, 1,000,000 steps for collagen. The number of steps used for minimizations employing conjugate gradients; 100,000 steps for lysozyme, 21,000 steps for both horse heart and sperm whale myoglobins, 100,000 steps for collagen. These many iterations were needed to fine–tune the structures so that they were good enough for use in circular dichroism calculations.

2.5. Molecular Dynamics

Molecular dynamics of the minimized structures were performed at 300 K using the Discover module of Insight®II. Molecular dynamics works by accessing different conformations of the protein over time. In this simulation, the atoms of the molecule move according to Newton's equation of motion defined as

$$F = ma \qquad [19]$$

where F is the force exerted by the particle of mass, m generating an acceleration, a. This net force exerted on any atom *i* by the rest of the atoms in the protein is expressed as a negative gradient of the potential energy of the protein in the form

$$F_i = -\nabla_i V \qquad [20]$$

where $\boldsymbol{V}$ is the potential energy of the system (protein) and is a function of all the positions of the atoms. Given that acceleration is second derivative of position $\boldsymbol{x}$, with respect to time, $\boldsymbol{t}$ the rate of change of the potential energy of the system with respect to position is given by

$$-\frac{\partial V}{\partial x_i} = m_i \frac{\partial^2_{x_i}}{\partial t^2} \quad [21]$$

This expression is integrated to provide the time series of conformations called trajectory or path followed by each atom of the protein during the simulation.

Cluster analysis was performed on the proteins using Discover module of Insight®II. This produced different frames (conformations) of the protein that were grouped together in clusters. Each cluster represented a specific root mean square (RMS) range. In principle, the RMS is a measure of how different the conformations are from one another. This measure considered each conformation and assessed the difference it has compared to the other conformations in the molecule. Therefore, a root mean square value other than zero means that the conformations are not identical. A few frames from each cluster were picked to represent all other conformations of the clusters. Molecular dynamics was done on each of these frames in order to predict their respective spectra. This was done to simulate the natural environment for which the protein exists so that there was continuous motion and other useful information such as the kinetic data of the protein could be obtained.

2.6. Mutations

Because the CD calculation treats only aliphatic amino acids (alanine, valine, proline, glycine, leucine, and isoleucine), all the proteins except collagen are mutated so that non–aliphatic amino acids were represented by alanine. The mutations are performed using the Brownian dynamics modeling package MACRODOX [44] and these are done after

minimization and molecular dynamics, but before using the structures for CD prediction. The mutations are done after minimization and molecular dynamics so as to correct for any minor displacements of hydrogen atoms during the mutation process.

2.7. Circular Dichroism Calculations

Cartesian coordinate files generated within Insight®II are used to calculate the $\pi - \pi^*$ amide transitions of the protein using the dipole interaction model (DInaMo)[45] With this method, coordinates of the non–chromophoric portions of the protein are treated directly, while the chromophoric amide portions are reduced to a single point at center of the N–C´ bond (o), shifted 0.1 Å towards the carbonyl carbon (x), or shifted 0.1 Å into the NC´O plane above the N–C´ bond (y). The Eulerian angles between the first amide chromophore and successive ones are calculated. From these, CDCALC generates normal modes and spectra for each protein. Three different dispersive parameters are tested: the original (O) parameters created for the dipole interaction model, the α–helical (H) parameters created for proteins, and the poly–L–Pro–II (J) parameters.[46] The output files generated by CDCALC are analyzed using the graphing software package OriginPro 7.5. (OriginLab Corporations, Northampton, MA). This involves locating the peaks at half–peak bandwidths of 4000 cm^{-1} and 6000 cm^{-1}, which represent the integral rotational strength of the combined oscillators. All data are fit to Lorentzian bands and the software did locate the peaks automatically. The theoretically predicted CD data are compared with SRCD data for each protein. The Root Mean Square Deviation (M), was used as a measure of accuracy to estimate how different the predicted CD data was from experimental (SRCD) data. The formula is expressed in mathematical form as

$$M = \sqrt{\sum \left[\text{Experiment al CD}(\lambda_i) - \text{Calculated CD}(\lambda_i)\right]^2} \quad [22]$$

A measure of the RM is relevant to provide a measure as to the performance of DInaMo in predicting CD of the proteins.

CHAPTER 3

THEORETICAL CIRCULAR DICHROISM PREDICTIONS OF HEN EGG–WHITE LYSOZYME WITH DInaMo

3.1. Introduction

Lysozyme, also known as muramidase or N–acetylmuramide glycanhydrolase, is a glycoside hydrolase enzyme.[47] This enzyme damages bacterial cell walls by catalyzing the hydrolysis of 1,4–β–linkages between N–acetylmuramic acid and N–acetyl–D–glucosamine residues in a peptidoglycan and between N–acetyl–D–glucosamine residues in chitodextrins.[47] It is present in large amounts in a number of secretions, some of which include tears, saliva, human milk, and mucus.[47] Hen egg–white also contains large amounts of this enzyme. It is also present in cytoplasmic granules of the polymorphonuclear neutrophils (PMN). Way back in 1909, the antibacterial activity of hen egg–white was first observed by Laschtschenko[48] but it wasn't until 1922 when Alexander Fleming named the enzyme in it lysozyme.[47] Lysozyme has been implicated in protection from gram positive bacteria like *Bacillus* and *Streptococcus* and protection of the conjunctiva of the eye.[49]

Lysozyme is one of the best characterized and most studied proteins. It is the second protein and the first ever enzyme for which a complete X–ray crystallographic structure was resolved.[50] It is also the first protein to be sequenced; it contains all 20 essential amino acids, and lysozme is the first enzyme for which a detailed mechanism of

action was proposed.[51] For these reasons, lysozyme is a very important protein and has been used extensively as a target protein in theoretical and experimental studies.

It is a compact globular protein consisting of a single polypeptide that has 129 amino acids. It has a mixture of the major secondary structures, with the four α–helices, two β–sheets, several turns, and other structures (Figure 3).

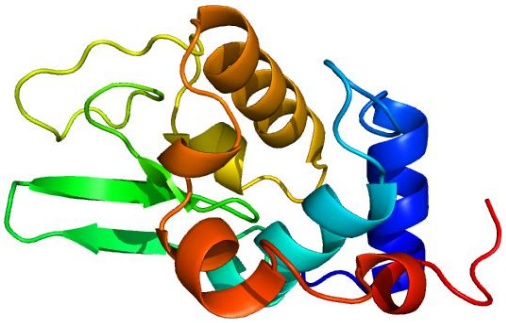


Figure 3. Lysozyme Secondary Structure. Lysozyme structure showing secondary structure elements: thick coils correspond to α–helices, the green tapes are β–sheets, and the thin ropes are turns and other structures.[52]

3.2. Previous Studies

Both theoretical and experimental studies have been done in an effort to better understand lysozyme. Levitt and Warshel reported a general method for studying enzymic reactions.[53] In their study, they focused on the dielectric, electrostatic and steric stabilization of the carbonium ion that exist in the reaction of lysozyme. This entailed studying in detail the different quantum mechanical and classical energy factors that can affect the reaction involving glycosidic bond cleavage by lysozyme. These involved quantum mechanical energies associated with bond cleavage and the charge redistribution that occurs on the substrate and the classical energies that result from steric and

electrostatic interaction between the enzyme and its substrate. A complete consideration of the enzyme–substrate complex together with surrounding solvent was necessary for their study. They found out that electrostatic stabilization is an important factor necessary for increasing the rate of the reaction step that leads to the formation of the carbocation intermediate. However, they also noted that steric factor such as strain that a result from binding of the substrate to the enzyme doesn't significantly contribute to this reaction rate. Pellicane studied the colloidal model of lysozyme in aqueous solutions using the Derjaguin–Landau–Verwey–Overbeek (DLVO) theory.[54] This involved calculating the structural functions of lysozyme at different pH values using Monte Carlo computer simulation and integral equation theories. He reported that at high pH values, DLVO theory predicts the formation of protein aggregates driven by competition of short–range attraction and long–range repulsion.

The pathways and intermediates involved in lysozyme folding have also been studied by Thomas Kiefhaber.[55] In his studies, he examined folding of hen egg white lysozyme by using interrupted refolding experiments. He observed that under strong native conditions, lysozyme can refold on parallel pathways, with the slow kinetic pathway involving the refolding of the major part (86 %) of the enzyme. He therefore suggested according to his results that the formation of the native state for the major part of the enzyme is retarded compared to the direct folding process meaning that partially structured intermediates formed in the refolding process are seemingly kinetically trapped, therefore slowing the refolding process to reach the native structure.

Solvation contributions to the thermodynamic stability of mutants of T4 lysozyme have also been studied theoretically by Ahluwalia and Deep.[56] Because lysozyme is a

benchmark protein used for a lot of theoretical and experimental studies, it was therefore chosen as one of the proteins for circular dichroism studies.

3.3 Methods

Energy minimization of lysozyme took 900,000 steps using the steepest descents algorithm followed by 100,000 steps using the conjugate gradient algorithm. By using steepest descents, it was possible to locate the local minima. The minimized structure was slightly refined by the conjugate gradient algorithm. These together helped in correcting any distorted bond lengths and angles that existed in the protein. Implicit solvent treatment was achieved by inclusion of a constant dielectric of 86.32 for water at 4 ^{o}C. This was used so that the conditions of the results matched experimental conditions used for obtaining the SRCD spectra with which the results were to be compared. Molecular dynamics of the minimized structures were performed at 300 K using the Discover module of Insight®II (Accelrys, San Diego). All non–alphatic amino acids were mutated to alanine because the code COREUL only recognizes the aliphatic amino acids at the moment. Mutation was done briefly after minimization and molecular dynamics so as to correct for any displacements in hydrogen atoms, which occur during molecular dynamics.

3.4. Results and Discussions

CDCALC and CAPPS did well with reproducing the positive $\pi - \pi^*$ band at 191 nm using Ho parameters with a bandwidth of 6000 cm^{-1} (Figure 4). Also, the location and intensities of the SRCD negative band at 207 nm was successfully reproduced with CAPPS using Hy parameters (Figure 5). With CDCALC, the intensity of the 207 nm band was reproduced best with the α–helical Ho parameters at 4000 cm^{-1} bandwidth. The poly–L–proline parameters consistently red–shifted predicted CD for both bands (Table 1 and Figure 4).

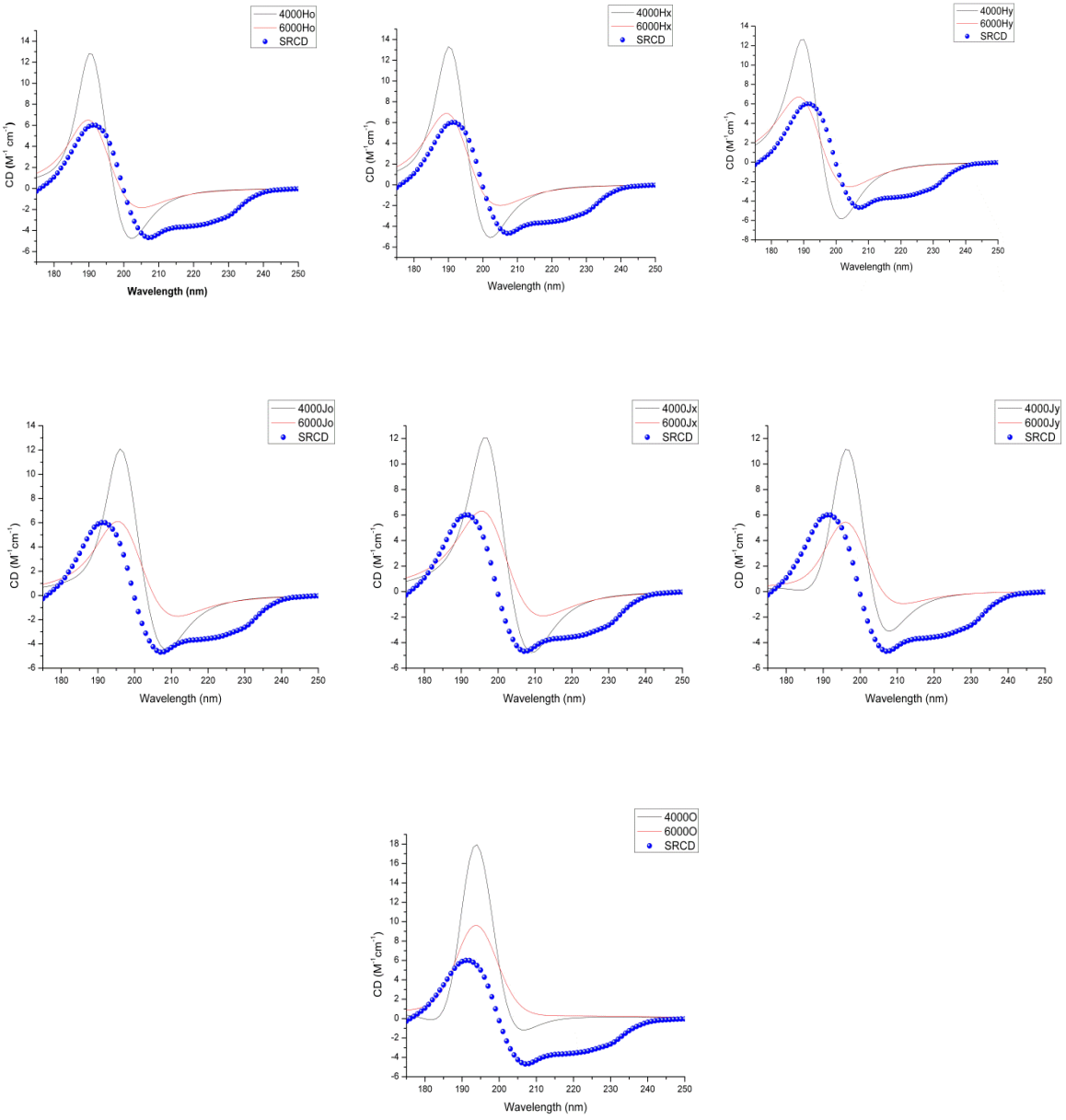


Figure 4. Lysozyme Predicted CD Using CDCALC.[52] Spectra are for the minimized lysozyme at bandwidths of 4000 and 6000 cm^{-1}. The blue dots are the experimental SRCD (CD0000045000).[31]

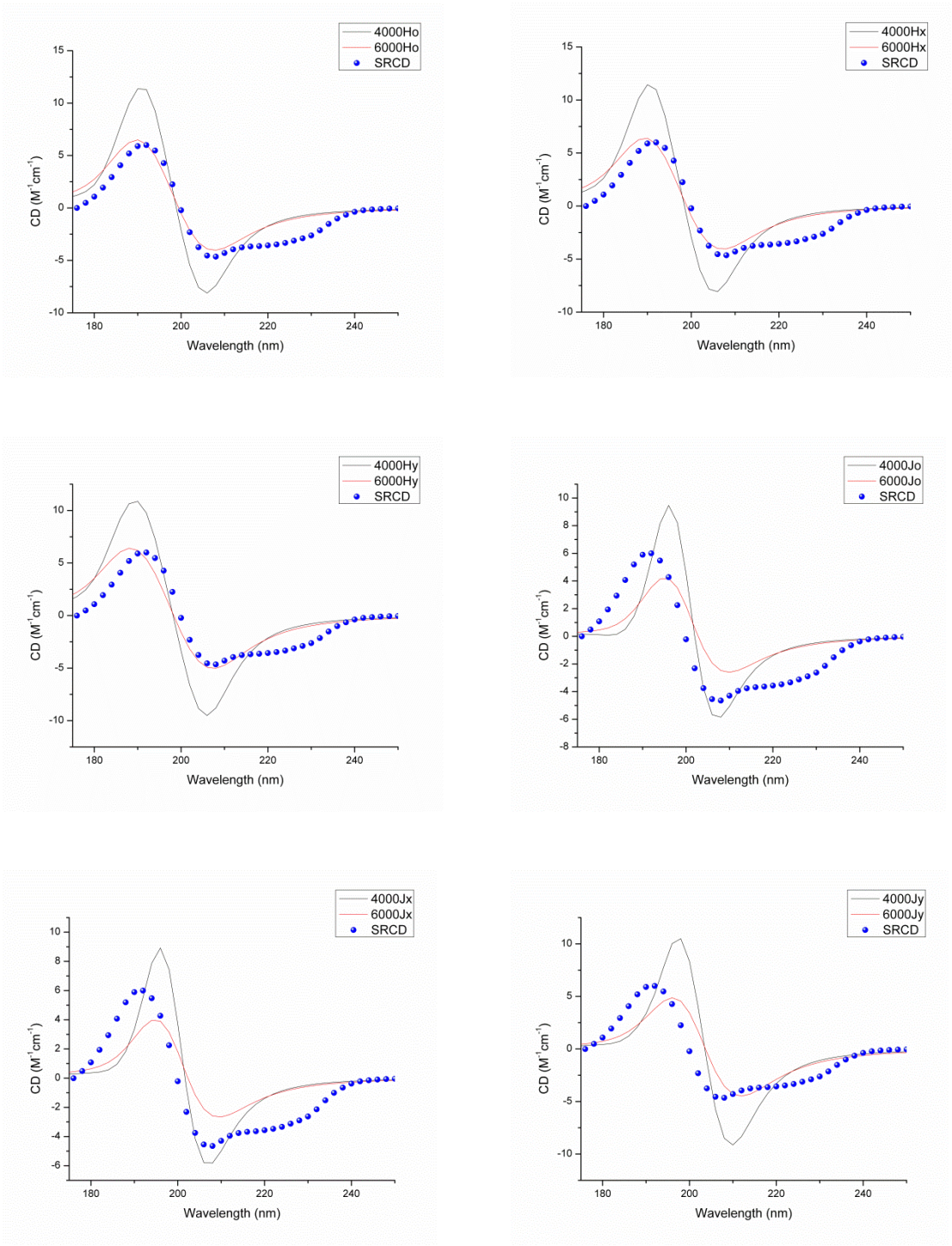


Figure 5. Lysozyme Predicted CD Spectra Using CAPPS.[52] The blue dots are the experimental SRCD (CD0000045000).[31]

Table 1. CD Analysis of Lysozyme.[52] The DInaMo calculations are for the minimized or rebuilt structure using CDCALC or CAPPS.

CD Method	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
[a]SRCD	191	6.01	13.20	207	–4.68	29.25	0.000
[b]4000Ho	190	12.79	8.80	202	–4.75	10.33	22.755
[b]6000Ho	190	6.51	13.19	205	–1.83	13.85	16.618
[b]4000Hx	190	13.30	9.15	202	–5.08	10.50	24.403
[b]6000Hx	189	6.87	13.83	205	–2.00	13.98	16.982
[b]4000Hy	190	12.64	9.89	202	–5.83	11.19	26.214
[b]6000Hy	188	6.70	14.64	204	–2.54	14.68	17.618
[b]4000Jo	196	12.09	9.27	209	–4.45	11.02	24.642
[b]6000Jo	195	6.10	14.03	212	–1.71	14.78	19.350
[b]4000Jx	197	12.06	10.02	209	–4.78	11.37	25.467
[b]6000Jx	196	6.30	15.07	212	–1.90	15.10	19.428
[b]4000Jy	196	11.15	8.71	208	–3.11	9.78	26.255
[b]6000Jy	196	5.45	12.33	212	–0.93	8.04	22.259
[b]4000O	194	17.94	9.48	207	–1.17	7.31	38.229
[b]6000O	194	9.63	13.53	not observed	not observed	not observed	28.783
[c]4000Ho	190	11.36	11.76	206	–8.11	12.16	14.474
[c]6000Ho	190	6.49	14.85	208	–4.03	16.57	7.243
[c]4000Hx	190	11.44	11.48	206	–8.07	12.27	14.825
[c]6000Hx	190	6.40	15.03	208	–4.04	16.59	7.784
[c]4000Hy	190	10.87	12.82	206	–9.52	12.77	16.680
[c]6000Hy	188	6.41	15.79	208	–5.01	17.40	8.288
[c]4000Jo	196	9.47	8.56	208	–5.85	10.90	13.835
[c]6000Jo	196	4.19	11.51	210	–2.60	15.46	11.387
[c]4000Jx	196	8.92	8.58	208	–5.82	11.11	12.481
[c]6000Jx	194	3.96	12.00	210	–2.66	15.69	10.888
[c]4000Jy	198	10.49	9.42	210	–9.14	11.99	18.510
[c]6000Jy	196	4.88	12.69	212	–4.48	17.00	9.885
[d]4000Hy	190	11.20	11.67	204	–7.01	11.87	20.946
[d]6000Hy	190	6.23	12.35	206	–3.41	17.51	11.594
[e]Woody	190	5.13	15.24	212	–2.10	15.60	9.673

[a]SRCD[31] Note: Origin was not able to separate the bands at 208 and 222 nm, which represent the π–π* and n–π* transitions respectively.
[b]CDCALC results from DInaMo using PDB structure 2VB1.[57]
[c]CAPPS results from DInaMo (Figure 5) using PDB structure 2VB1.[57]
[d]CAPPS results done by Bode and Applequist 1998 using PDB structure 1LSE.[58]
[e]Matrix method results by Sreerama and Woody 2004 using PDB structure 7LYZ.[59]

The matrix method of Sreerama and Woody,[60] which has been reported in the literature had similarity with our CDCALC results when compared to experiment by visual inspection of the peak locations and intensities (Appendix 1 or Figure 30) with the matrix method providing a reasonable approximation for the n – π* band. However, CDCALC did slightly better for locating the positive π – π* band at 192 nm. The smallest root square deviation (M) between experiment and calculated CD were obtained for CDCALC and CAPPS to be 16.618 $M^{-1}cm^{-1}$ and 7.234 $M^{-1}cm^{-1}$ respectively and both were with the α–helical parameters Ho (Table 1). However, the M values obtained using CAPPS were smaller and comparable to that obtained using the matrix method (9.673 M^{-1} cm^{-1}).[60]

A root mean square deviation comparing the backbone of the minimized structure to the original was done using PyMOL. It showed that the α–helices though preserved after minimization had small changes while very large changes occurred in the undefined structures. The value of the RMS from PyMOL was 1.934 (Figure 6)

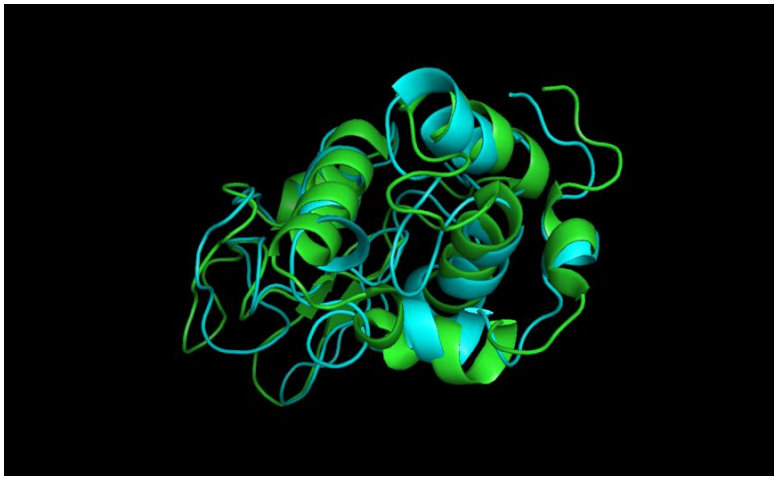


Figure 6. Structural Difference in Minimized and Original Lysozyme structures, RMS deviation (PyMOL) = 1.934. Green structure is original and blue structure is minimized[52]

MD cluster analysis (Figure 7) was used to examine how sensitive the dipole interaction model was to different structures used with the CDCALC. It was observed that the predicted CD spectra resembled experiment the most using Hy and Jx parameters at 6000 cm^{-1}, (Tables 2 and 3). The Hy predictions are slightly blue shifted and the Jx predictions are slightly red shifted when compared with SRCD.

Table 2. Lysozyme CD Analysis for MD Clusters Using the α–Helical Hy Parameters at a Bandwidth of 6000 cm^{-1} Running CDCALC[52]

Snap–Shot[a]	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
L5	187	5.84	14.74	204	–2.82	16.14	18.197
L7	186	6.06	15.99	205	–2.14	15.79	19.188
L27	187	7.24	16.64	205	–3.67	16.59	19.359
L30	188	8.60	14.43	205	–3.30	14.98	18.420
L42	186	7.73	17.25	204	–3.79	15.80	21.807
L53	187	8.14	16.13	204	–4.34	15.84	19.856
L54	186	7.87	16.92	204	–4.05	15.73	19.856
L55	187	7.73	16.07	204	–4.59	16.07	20.550
L57	187	8.13	17.07	205	–3.91	16.35	20.565
L58	187	8.22	16.86	205	–3.76	15.83	20.762
L63	186	7.46	16.93	204	–4.03	16.00	20.953
SRCD[b]	191	6.01	13.20	207	–4.68	29.25	0.000

[a]The number refers to the frame number of the snapshot in the MD trajectory.
[b]SRCD.[31]

Table 3. Lysozyme CD Analysis for MD Clusters Using the Poly–L–Proline Jx Parameters at a Bandwidth of 6000 cm^{-1} Running CDCALC[52]

Snap–shot[a]	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
L5	194	5.72	14.92	212	–2.24	16.96	16.436
L7	193	5.51	16.92	213	–1.47	16.93	18.404
L27	194	6.91	17.54	213	–3.05	17.69	16.019
L30	195	8.50	14.92	213	–2.57	15.22	22.318
L42	194	7.27	18.70	212	–2.98	16.60	17.404
L53	194	7.77	16.41	212	–3.53	16.62	17.253
L54	194	7.37	18.19	212	–3.27	16.39	16.703
L55	194	7.35	16.82	212	–3.73	16.82	15.572
L57	194	7.73	17.69	212	–3.03	17.01	18.122
L58	194	7.60	17.95	213	–2.83	16.32	19.517
L63	193	7.16	17.54	212	–3.30	16.80	15.365
SRCD[b]	191	6.01	13.20	207	–4.68	29.25	0.000

[a]The number refers to the frame number of the snapshot in the MD trajectory.
[b]SRCD.[31]

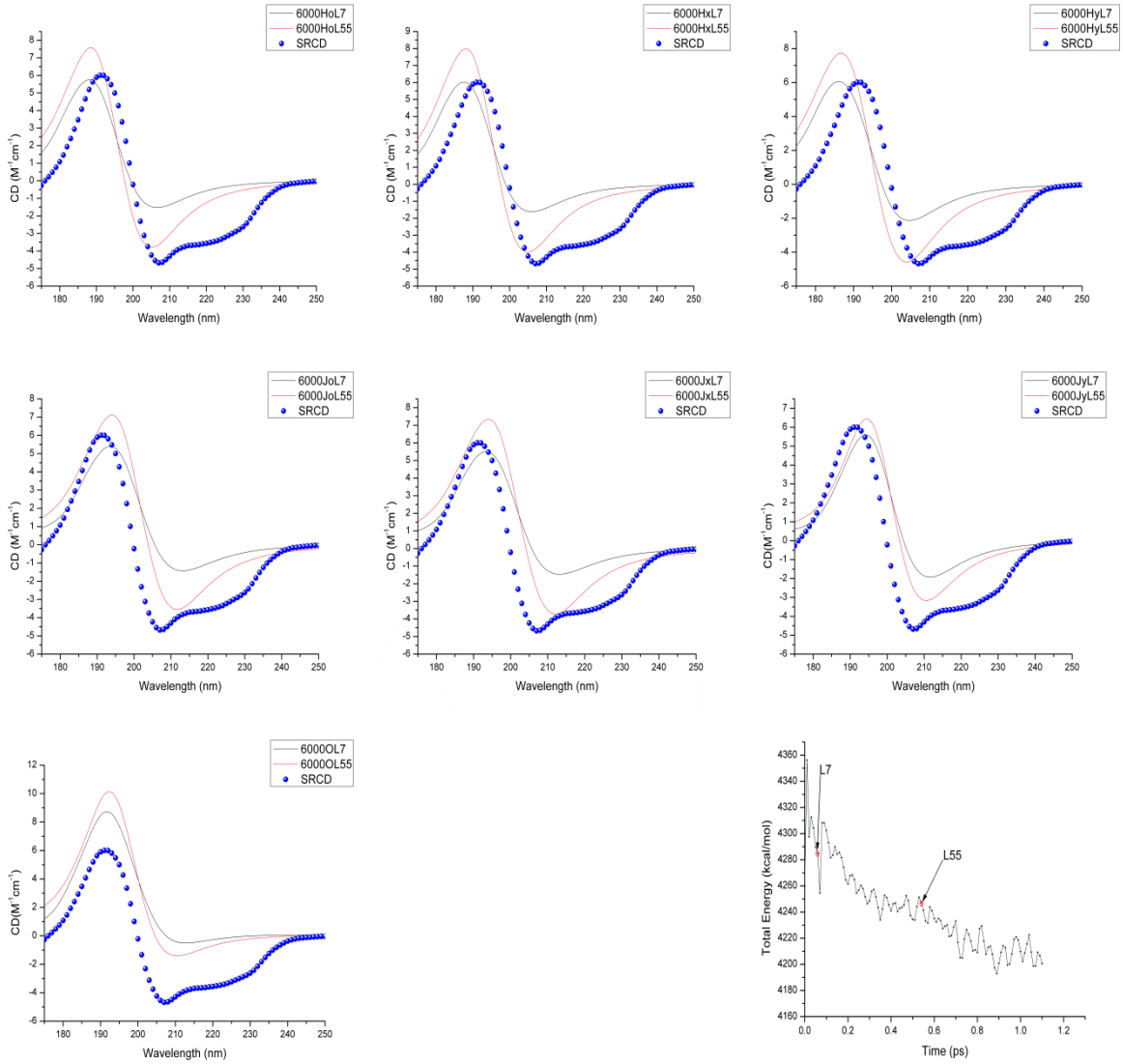


Figure 7. Lysozyme CD and MD.[52] CD spectra for two MD snapshots predicted by CDCALC using α–helical parameter (Hy) at bandwidth of 6000 cm^{-1}. The blue dots are the experimental SRCD (CD0000045000).[31] The figure in the lower right corner is the CD trajectory with the two representative snapshots labeled.

Generally, it was noticed that root square deviation from SRCD did change as the structures underwent small conformational changes, predominantly in the side chains. Calculations with the α–helical Hy parameters were typically higher than that of the energy minimized structure and varied over a range of 3.61 in the root square deviation

(Table 2). A greater variation was observed for the Jx parameters with root square deviations between 15.365 $M^{-1}cm^{-1}$ and 22.318 $M^{-1}cm^{-1}$ (Table 3) and these values were occasionally better than the energy minimized structure**.**

Also, the root mean square deviation comparing the backbones was done using PyMOL, to compare the minimized structure to the MD structures. Structure with the largest deviation was L55, with an RMS of 2.828. In a similar manner, small changes were observed to have occurred in the α–helices while substantially larger changes occurred in the undefined structures (Figure 8)

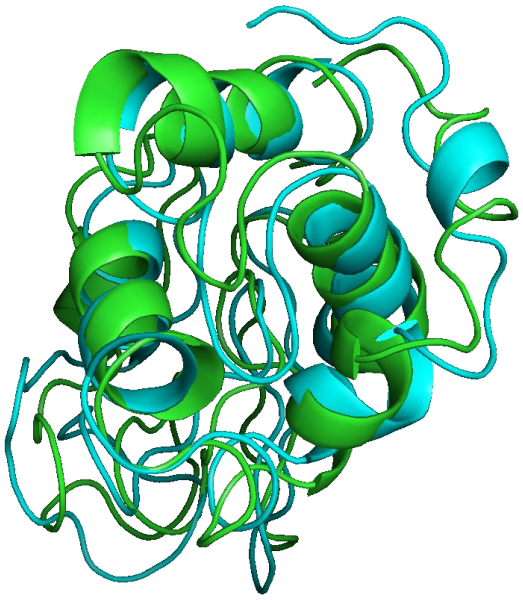


Figure 8. Structural Difference in Minimized and MD Lysozyme structure (L55), RMS deviation (PyMOL) = 2.828. Green structure is cluster and blue structure is minimized.[52]

Lysozyme being a benchmark protein was a great candidate for testing the DInaMo. The results for both CDCALC and CAPPS resembled SRCD at the positive band at 191 nm. From the results reported, we noticed that minimization of lysozyme did provide for prediction of the positive $\pi - \pi^*$ band at 191 nm using the 6000 cm^{-1} and the

negative $\pi - \pi^*$ band at 207 nm using the 4000 cm^{-1} bandwidth. These findings are consistent with the findings of Applequist findings using CAPPS. Previous studies done on lysozyme by Applequist using CAPPS[46] gave results similar to our CAPPS calculations and both had similar features when compared with SRCD. Also, studies by Sreerama and Woody[60] using the matrix method gave results comparable to our CAPPS results. CDCALC results obtained after molecular dynamics gave better spectra than those before molecular dynamics, which showed that molecular dynamics is relevant for better prediction of SRCD.

3.5. Conclusions

The use of DInaMo as a means of theoretically predicting the CD of lysozyme was successful. The band at 191 nm was best predicted with the Hy and Jx parameters when molecular dynamics was used to fine tune the protein structure. This approach gave good correlation with SRCD using just one band unlike the case of minimization only that gave correlation of the two bands using both 4000cm^{-1} and 6000 cm^{-1} bandwidths.

CHAPTER 4

THEORETICAL PREDICTION OF THE CIRCULAR DICHROISM OF HORSE HEART MYOGLOBIN AND SPERM WHALE MYOGLOBIN USING DInaMo

4.1. Introduction

Myoglobin is an oxygen–binding carrier protein abundantly present in the muscle tissue of almost all mammals.[61] This protein is a single–chain globular protein composed of 153 amino acids.[62] It also contains a heme ring otherwise called an iron–containing porphyrin or prosthetic group at its center surrounded by the remaining fold also known as the apoprotein fold. It is predominantly helical with eight alpha helices around a hydrophobic core (Figure 9).

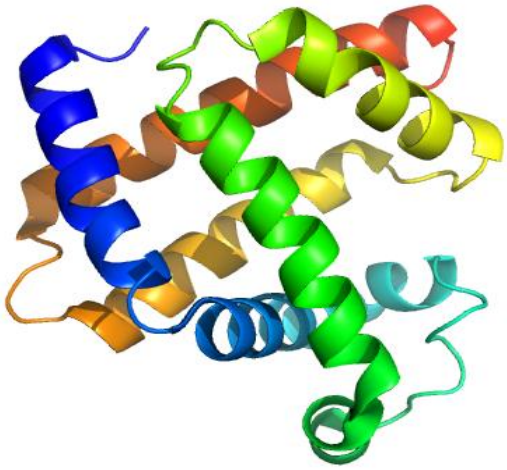


Figure 9: Cartoon Representation of Myoglobin Secondary Structure: the coils correspond to α–helices and the ropes are turns structures.[52]

It has a molecular weight of 17,699 Daltons (with heme), and is the primary oxygen–carrying pigment of muscle tissues.[63] It was the first protein for which a three–dimensional structure X–ray structure was determined.[64]

4.2. Background Studies

Theoretical and experimental studies have both been reported on myoglobin. Back in 1964, Holzwarth and Dotty[64] reported CD for sperm whale metmyoglobin. They compared the data with that for the α–helical peptide copoly–L–Glu–Lys–Ala and found the three peptide rotatory bands that characterize helical proteins present in both CD plots.[64]

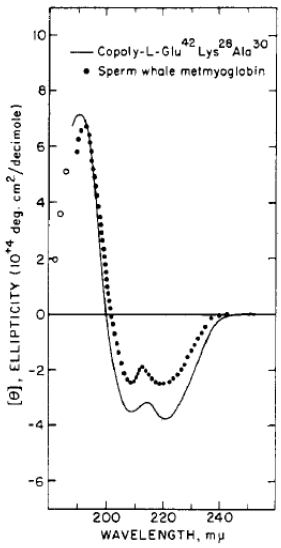


Figure 10: Circular dichroism of copoly–L–Glu–Lys–Ala at pH 3.1 in 0.1M NaF, and the circular dichroism of spermwhale metmyoglobin at pH 7.0 in 0.02M phosphate buffer[64]

Based on these findings, they concluded that the electronic structure and conformation of both synthetic and natural polymers were identical. Also, Lees and Wallace compared CD obtained from a conventional CD instrument with to that obtained using synchrotron radiation and found both to be similar for the wavelength region spanning from 175 – 300 nm except for the lower wavelength region when using the conventional CD instrument.[18] Orry e*t al.,* also reported conservation in the integrity of horse myoglobin even after extensive exposure to synchrotron radiation.[65] Much of the results reported in this chapter appeared in Forlemu *et al.*,[52] and Nori helped with calculations using CAPPS.

4.3. Methods

Energy minimization of each myoglobin structure took 110,000 steps using the steepest descents algorithm followed by 21,000 steps using the conjugate gradient algorithm. In each case, the steepest descents strategy located a local minimum that was slightly refined by the conjugate gradient algorithm. A constant dielectric of 86.32 for water at 4 ^{o}C, was used so that the conditions of the results matched experimental conditions used for the SRCD spectra. Molecular dynamics of the minimized structures were performed at 300 K using the Discover module of Insight®II (Accelrys, San Diego). All non–alphatic amino acids were mutated to alanine because the code COREUL only recognizes the aliphatic amino acids at the moment. Mutation was done briefly after minimization and molecular dynamics so as to correct for any displacements in hydrogen atoms, which occur during molecular dynamics.

4.4. Results and Discussion

Most of the secondary structure, especially the α–helices, of both horse heart and sperm whale myoglobin were retained after minimization. However, the coil structures were modified. (Figures 11 and 12)

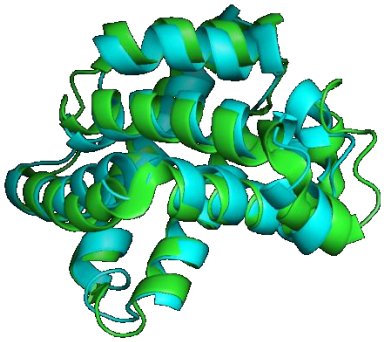


Figure 11: Horse Heart Myoglobin structural difference between original and minimized structures, RMS deviation (PyMOL) = 1.883. Green structure is original and blue structure is minimized.[52]

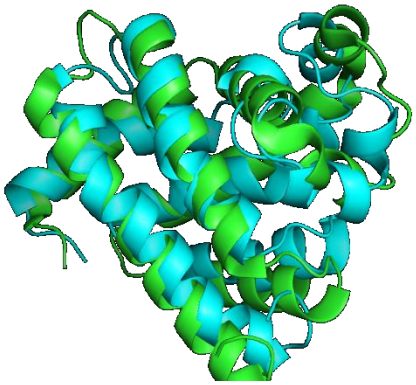


Figure 12: Sperm Whale Myoglobin structural difference between original and minimized structures, RMS deviation (PyMOL) = 3.246. Green structure is original and blue structure is minimized.[52]

The circular dichroism spectra computed for the amide $\pi - \pi^*$ transitions for horse heart myoglobin using CDCALC is shown in Figure 13. They were similar to results obtained for sperm whale myoglobin (Figure 14, Table 4). In the same like, CAPPS results obtained for the two species also produced similar results (Figure 15, Figure 16, and Table 5).

Table 4: Horse Heart Myoglobin CD Analysis for Single Minimized for Rebuilt Structure[52]

CD Method	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth (nm)	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth (nm)	M (Root Mean Square Deviation) (M^{-1} cm^{-1})
[a]SRCD	192	16.75	14.58	209	–7.51	7.78	0.000
[b]4000Ho	189	30.16	9.36	203	–17.40	11.81	64.64
[b]6000Ho	189	15.49	12.68	205	–8.46	15.90	37.70
[b]4000Hx	189	30.51	9.65	203	–17.82	11.96	66.73
[b]6000Hx	188	15.93	13.17	205	–8.76	16.02	38.58
[b]4000Hy	188	26.00	10.10	203	–18.22	12.22	67.69
[b]6000Hy	188	13.81	13.74	205	–9.31	16.45	44.81
[b]4000Jo	195	28.36	9.90	210	–16.34	12.55	45.83
[b]6000Jo	194	14.53	13.44	212	–7.93	16.91	22.01
[b]4000Jx	195	27.10	10.51	211	–16.09	12.86	43.96
[b]6000Jx	194	14.30	14.35	213	–7.94	17.23	21.89
[b]4000Jy	195	32.65	8.65	208	–18.76	11.65	58.63
[b]6000Jy	195	14.66	11.48	210	–8.69	16.23	26.52
[b]4000O	193	46.39	9.44	206	–16.53	11.11	82.45
[b]6000O	192	23.89	12.45	209	–6.69	14.78	28.78
[c]4000Ho	192	21.50	10.11	206	–16.33	12.78	29.60
[c]6000Ho	190	11.25	14.16	208	–8.52	17.20	22.65
[c]4000Hx	190	21.05	10.55	206	–16.40	12.81	31.55
[c]6000Hx	190	11.23	14.70	208	–8.55	17.17	24.14
[c]4000Hy	190	20.23	10.99	206	–18.46	13.22	33.95
[c]6000Hy	190	10.77	15.27	208	–10.01	17.80	25.34
[c]4000Jo	196	18.75	7.88	208	–13.09	11.62	30.12
[c]6000Jo	196	7.75	10.97	210	–6.25	16.33	28.93
[c]4000Jx	196	17.74	7.95	208	–12.96	11.61	28.95
[c]6000Jx	194	7.39	11.22	210	–6.24	16.38	28.94
[c]4000Jy	198	21.92	8.35	210	–18.68	12.52	38.72
[c]6000Jy	196	9.23	11.24	212	–9.48	17.45	26.52

[a]SRCD for horse heart myoglobin (CD000047000). There was a 3rd peak at 221 nm with CD intensity of –8.08[66] with a half–peak bandwidth of 19.17 nm.

[b]CDCALC results from DInaMo for horse heart myoglobin from 3LR7[37] that was minimized via molecular mechanics.

[c]CAPPS results from DInaMo for horse heart myoglobin from 3LR7[37] that contained rebuilt secondary structures only.

Table 5: Sperm Whale Myoglobin CD Analysis for Single Minimized for Rebuilt Structure[52]

CD Method	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth (nm)	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth (nm)	M (Root Mean Square Deviation) (M^{-1} cm^{-1})
[a]SRCD	193	17.33	14.58	210	–7.77	7.48	0.00
[b]4000Ho	187	18.48	10.93	202	–11.90	12.47	64.45
[b]6000Ho	186	10.27	15.09	204	–6.07	16.29	54.39
[b]4000Hx	187	19.38	11.25	202	–12.30	12.68	67.06
[b]6000Hx	186	10.94	15.39	204	–6.34	16.44	55.13
[b]4000Hy	186	18.31	11.22	202	–12.82	13.36	73.06
[b]6000Hy	185	10.30	15.21	203	–7.00	17.04	60.06
[b]4000Jo	193	17.41	11.55	203	–11.16	13.27	27.20
[b]6000Jo	192	9.65	15.99	210	–5.71	17.31	34.56
[b]4000Jx	193	17.74	12.07	209	–11.14	13.73	25.14
[b]6000Jx	192	10.06	16.49	211	–5.80	17.70	32.70
[b]4000Jy	194	19.04	8.98	207	–12.45	11.79	34.60
[b]6000Jy	193	9.11	12.99	209	–5.93	16.23	37.80
[b]4000O	191	31.24	9.42	205	–10.97	11.80	47.83
[b]6000O	191	16.49	13.29	208	–4.60	15.20	30.78
[c]4000Ho	190	20.29	11.04	206	–16.91	13.12	36.03
[c]6000Ho	188	10.88	14.82	208	–9.02	17.61	29.04
[c]4000Hx	190	20.10	11.39	206	–17.02	13.23	39.03
[c]6000Hx	188	11.11	15.01	206	–9.27	17.51	30.90
[c]4000Hy	188	19.18	11.89	206	–18.97	13.81	42.12
[c]6000Hy	188	10.74	15.30	206	–10.61	18.47	32.95
[c]4000Jo	196	17.42	8.24	208	–13.49	11.80	28.84
[c]6000Jo	194	7.54	11.38	210	–6.61	16.57	30.04
[c]4000Jx	196	16.45	7.48	208	–13.51	11.86	28.34
[c]6000Jx	194	7.28	11.80	210	–6.68	16.60	30.15
[c]4000Jy	196	19.71	8.93	210	–19.55	12.63	34.51
[c]6000Jy	196	8.87	11.48	213	–10.13	17.69	25.76
[d]4000Hy	189	17.29	13.00	206	–18.30	13.39	55.31
[d]6000Hy	188	9.49	16.25	207	–10.35	18.19	44.31
[e]SrWo	184	16.38	12.16	215	–4.70	17.52	36.59
[f]Hirst	191	9.52	16.90	213	–7.61	21.51	26.14

[a]SRCD for sperm whale myoglobin (CD000048000). There was a 3rd peak at 223 nm with CD intensity of –8.66[31] with a half–peak bandwidth of 19.44 nm.

[b]CDCALC results from DInaMo for sperm whale myoglobin starting from 2JHO that was minimized via molecular mechanics.

[c]CAPPS results from DInaMo for sperm whale myoglobin starting from 2JHO that contained rebuilt secondary structures only.

[d]CAPPS results published by Bode and Applequist 1998[46] for sperm whale myoglobin PDB structure 1VXA.

[e]Matrix method results published by Sreerama and Woody 2004[60] for sperm whale myoglobin PDB 1MBN.

[f]Matrix method results published by Hirst et al. 2003[66] for sperm whale myoglobin PDB 4MBN.

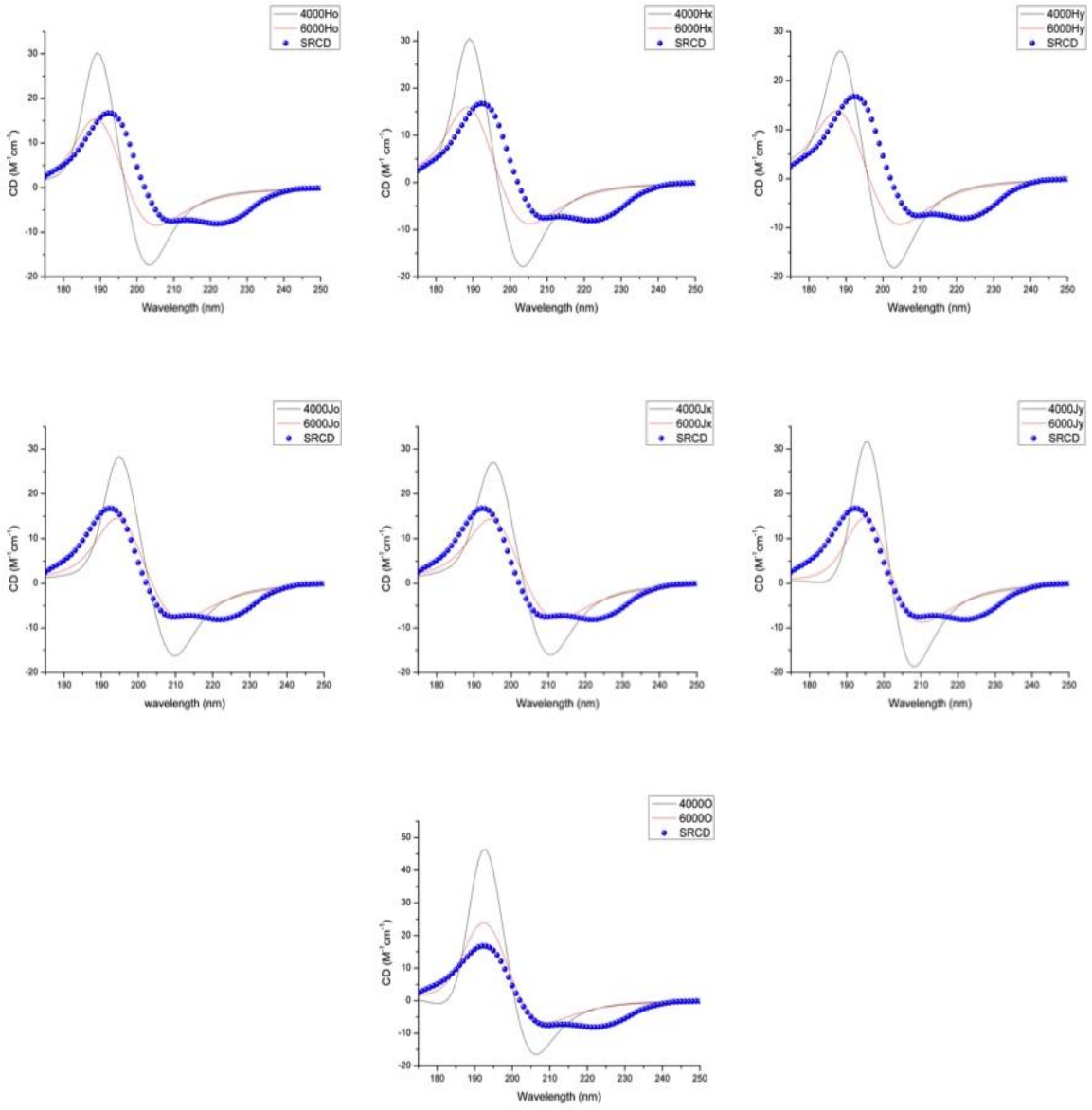


Figure 13: Horse Heart Myoglobin Predicted CD Using CDCALC: Spectra are for the minimized myoglobin at bandwidths of 4000 and 6000 cm^{-1}.[52] The blue dots are the experimental SRCD (CD0000047000).[31]

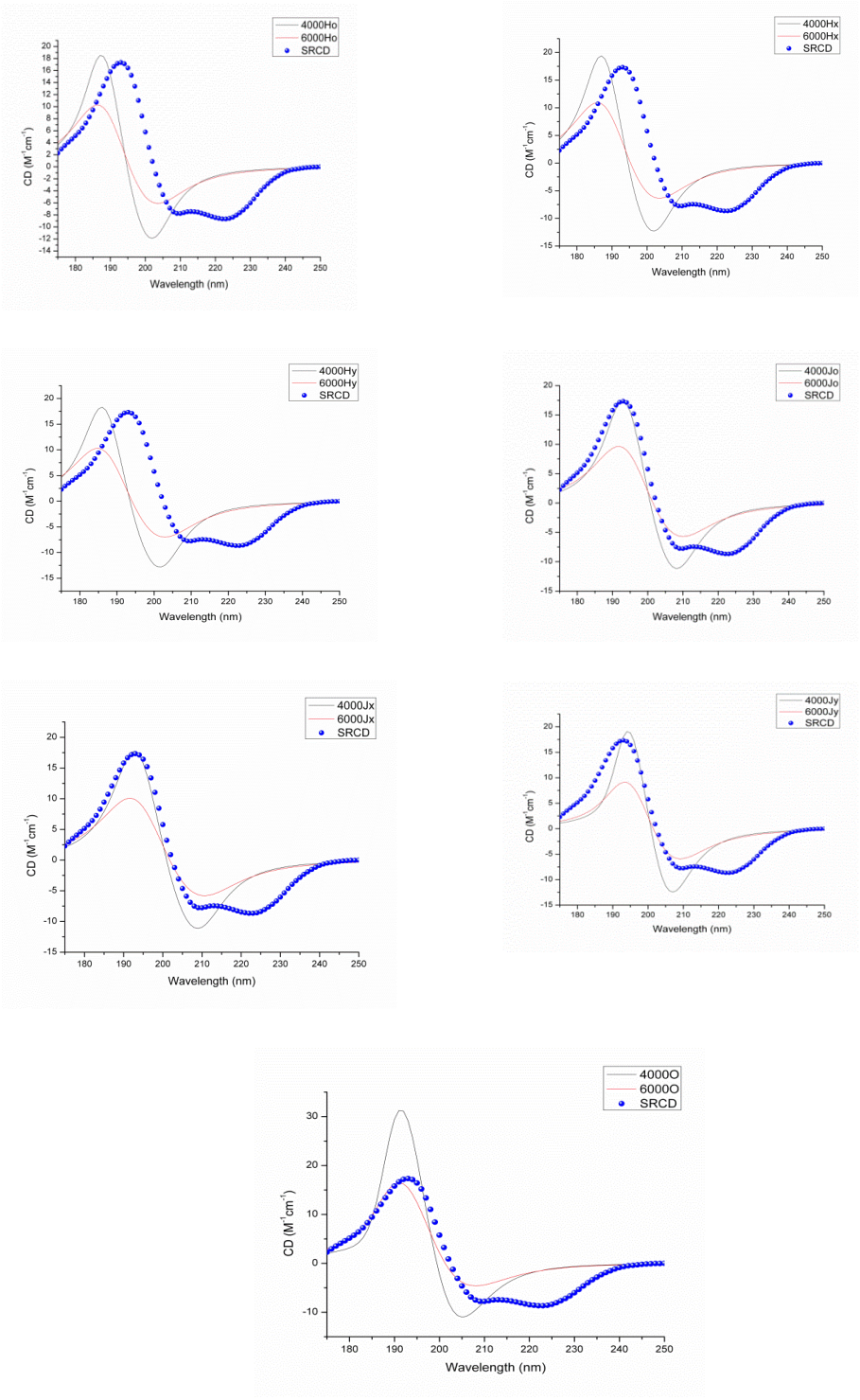


Figure 14: Sperm Whale Myoglobin Predicted CD Spectra Using CDCALC.[52] The blue dots are the experimental SRCD (CD0000047000).[31]

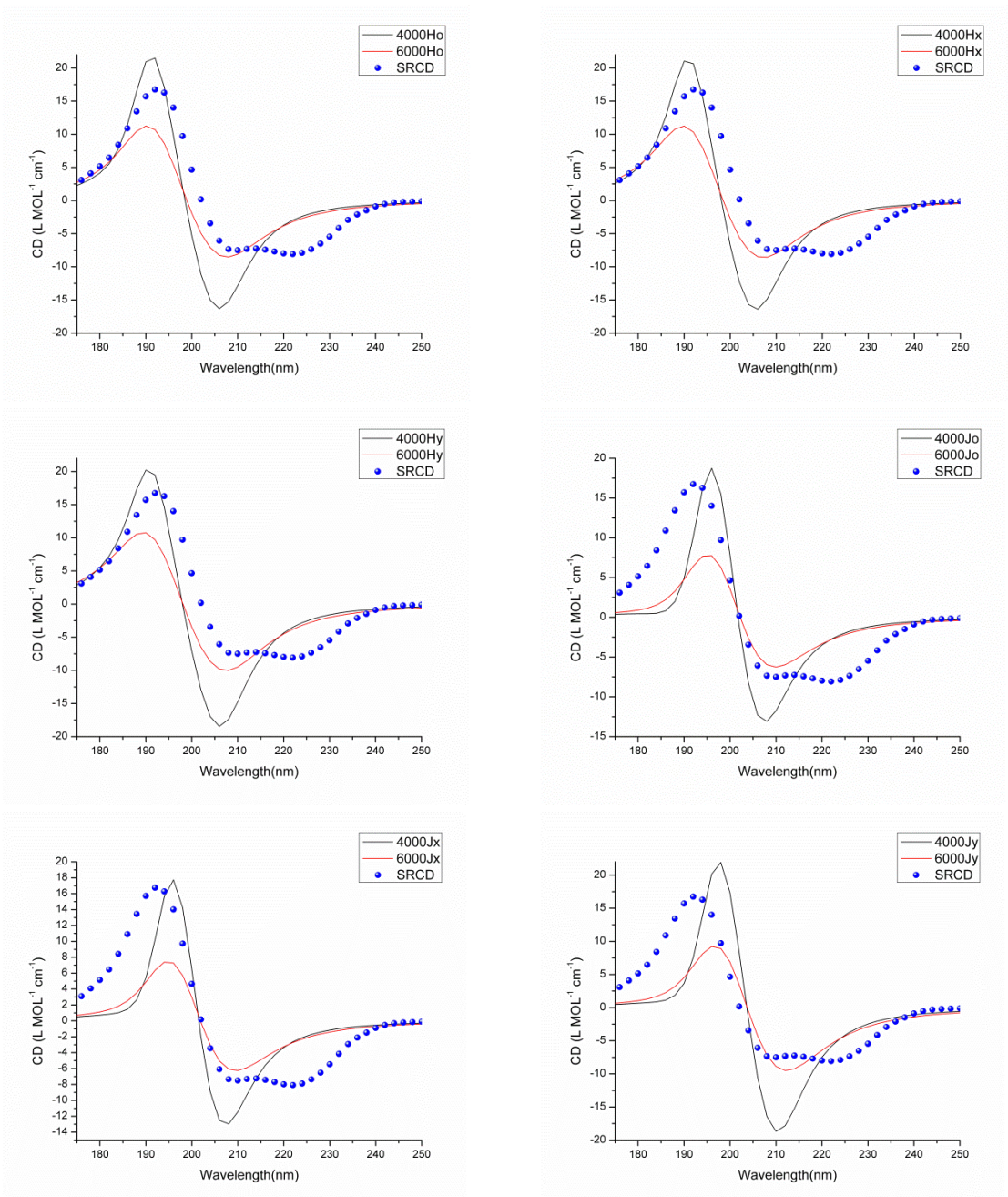


Figure 15: Horse Heart Myoglobin Predicted CD Spectra Using CAPPS.[52] The blue dots are the experimental SRCD (CD0000047000).[31]

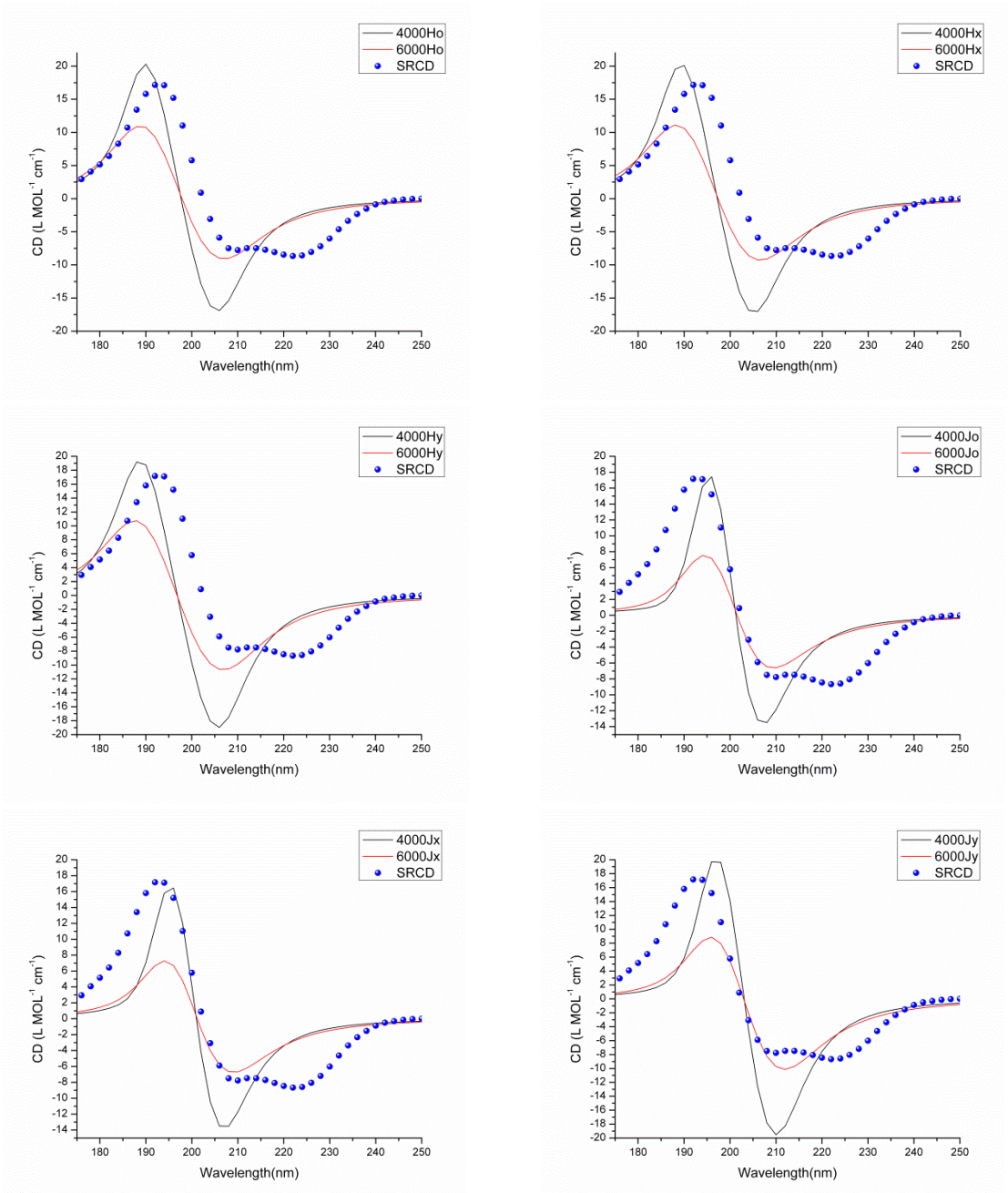


Figure 16: Sperm Whale Myoglobin Predicted CD Spectra Using CAPPS.[52] The blue dots are the experimental SRCD (CD0000047000).[31]

In terms of the positions of the peaks, the intensity of the absorption and the half–peak bandwidth for the π–π* transitions at 192 nm for horse heart myoglobin (Figure 13), CDCALC produced good intensity and half–peak bandwidth using the Hy

parameters with a 6000 cm^{-1} bandwidth However, the maximum was blue shifted by 4 nm (Table 4). The maximum was better located with Jx 6000 cm^{-1} nm, but it is red shifted by 2 nm. CAPPS for the same band in horse heart myoglobin (Figure 15), did not fully reach the intensity using the Ho 6000 cm^{-1} parameters, but successfully located the band closer to that of the experiment at 192 nm and did as well as CDCALC with the half–peak bandwidth. As for the 209 nm SRCD peak for horse heart myoglobin, the Hy 6000 cm^{-1} parameters agreed best with experiment for CDCALC (Figure 14), whereas the Ho 6000 cm^{-1} agreed best for CAPPS (Figure 15). In either case, the logical α–helical parameters successfully reproduced the second band. The third SRCD peak, which usually occurs at ~222 nm and assigned to amide n – π* was not predicted. It has however been predicted by the matrix methods.[60, 66]

Visual inspection of the predicted CD for sperm whale myoglobin (Figures 14 and 16) showed that both CDCALC and CAPPS prefer the 4000 cm^{-1} bandwidth for the peak at 193, but it was the CDCALC predictions obtained using either the Jo or Jx parameters that located the peak best. For the band at 210 nm, all J parameters yielded better intensities with the bandwidth of 6000 cm^{-1}. Helical parameters for both methods were blue shifted.

Comparing predicted CD to SRCD and using helical parameters, the best root mean square deviation (M) for horse heart myoglobin was found using CAPPS to be 24.14 M^{-1} cm^{-1} with 6000 cm^{-1} bandwidth (Table 4). With CDCALC, best root mean square deviation was also surprisingly obtained to be 21.89 M^{-1} cm^{-1} for horse heart myoglobin using the poly–L–proline Jx parameters with a bandwidth of 6000 cm^{-1}.

Contrary to expectation, the best helical parameter results for horse heart were obtained using Ho 6000 cm^{-1} to be 37.70 M^{-1} cm^{-1} and this was using CDCALC. This M value was significantly larger than that obtained with CAPPS (22.65 M^{-1} cm^{-1}) using the same parameters.

For sperm whale, the best root mean square deviation (M) between SRCD and calculated CD was found to be 25.14 M^{-1} cm^{-1} with CDCALC and using the poly–L–proline parameters Jx with a bandwidth of 4000 cm^{-1} (Table 5). Using CAPPS, the poly–L–proline parameters Jy yielded the best M value of 25.76 M^{-1} cm^{-1} with a bandwidth of 6000 cm^{-1}. This is better than the previously reported M value obtained using CAPPS calculation on sperm whale myoglobin in the literature (44.31 M^{-1} cm^{-1}) which was for the α–helical parameters Hy with the 6000 cm^{-1} bandwidth[46] using the PDB structure 1VXA.[67] The matrix method of Hirst, Coletta and Gilbert that uses higher level quantum mechanical parameters for the amide chromophore and the refined sperm whale myoglobin structure 4MBN does as well as CDCALC yielding an M value of 26.14 M^{-1} cm^{-1}.[66]

Molecular Dynamics simulation and cluster analysis for horse heart myoglobin produced small variations in the backbone structures (Figure 17). A root mean square deviation comparing the backbones was done using PyMOL, to compare the minimized structure to the MD structures for horse heart myoglobin. Structure with the largest deviation was M13, with an RMS of 1.804 Å. Small shifts occurred in backbone with no loss in the α–helices.

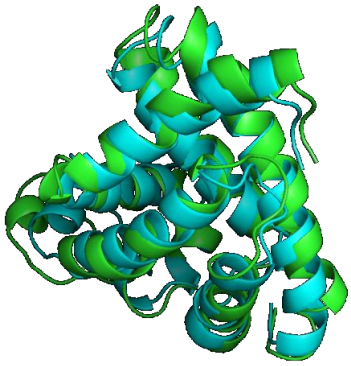


Figure 17: Horse Heart Myoglobin Structural Difference in Minimized and MD Structures, RMS deviation (PyMOL) = 1.804. Green structure is cluster and blue structure is minimized.[52]

A root mean square deviation comparing the backbones was done using PyMOL to compare the minimized structure to the MD structures for sperm whale myoglobin. Structure with the largest deviation was M103, with an RMS of 1.682 Å (Figure 18). Small shift occurred in backbone with no loss in the α–helices.

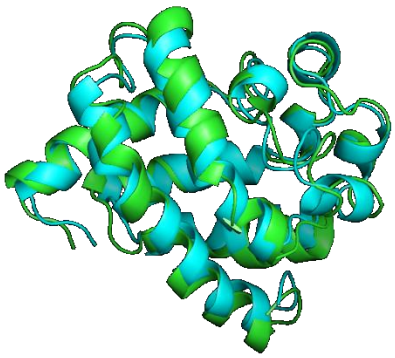


Figure 18: Sperm Whale Myoglobin Structural Difference in Minimized and MD Structures, RMS deviation (PyMOL) = 1.682. Green structure is cluster and blue structure is minimized.[52]

The best results for horse heart myoglobin using CDCALC were obtained using Jy parameters with a bandwidth of 6000 cm^{-1} (Table 4). The best root square deviation in predicted CD occurred using the Jx. Visually, they compared well to experiment

(Appendix 2 Figure 31 and Appendix 4 Table 13) with very little difference between the two MD structures presented (M28 ad M36). The predicted spectra obtained were consistently blue shifted for both structures with the most shifts from using the α–helical Hx parameters.

The MD simulation and cluster analysis for sperm whale myoglobin (Appendix 3 Figure 32 and Appendix 5 Table 14) showed small variations between the two structures for all parameters. The lowest M value occurred for Jy 6000 cm^{-1}, but several J parameters came close using the 4000 cm^{-1}. Visually, the plot obtained using Jy parameters at 4000 cm^{-1} was most similar to experiment for structures M101 and M105. With both minimization and MD, the 4000 cm^{-1} did a decent job at attempting to reproduce experiment.

4.5. Conclusions

The results indicate that minimization did well with generating decent structures for CD prediction. However, the method of choice for CD calculation determined which bandwidth gave the closest result to experiment. CDCALC seems to do a better job than CAPPS at these predictions as seen in its ability to predict experiment for both bands with just one wavelength (6000 cm^{-1}). Molecular dynamics proves to be necessary to fine-tune the structures such that they give even better predictions than for minimized structures only. It can therefore be concluded that minimization proceeded by adequate molecular dynamics can provide good enough structures that predict experimental CD. It is necessary for more work to be done to obtain the best strategies for minimization and molecular dynamics.

CHAPTER 5

THEORETICAL CIRCULAR DICHROISM OF THE TRIPLE HELIX MODEL OF COLLAGEN USING CDCALC OF DInaMo

5.1. Introduction

Collagen is a group of naturally occurring proteins found in nature, exclusively in flesh and connective tissues of animals.[68] It is the main component of connective tissue and makes the most in percentage of mammalian protein (about 25 – 35% of whole body protein). It forms elongated fibrils in fibrous tissue such as the ligament, tendon and even in the skin and is also found in cartilage, bone, blood vessels, intervertebral disc and the cornea of the eye. The molecular structure of collagen was first determined in the mid–1930s and since then a lot of effort has been put into studying the conformation of the monomer.[69] Its packing structure has mostly been defined to be fibrillar even though some studies report it to be hexagonal or quasi–hexagonal.[70] Several studies conflict on the structure, some reporting it to be sheet–like while others report it to be microfibrillar.[71]

Collagen is made up of three polypeptide strands each possessing the poly–L–proline II geometry. These three left–handed helices are twisted together into a right–handed coiled coil, called a triple helix or "super helix" (Figure 19).[52, 72] The strands measure up to 300 nm long with a diameter of 1.5 nm and the strands are also referred to as α–chains. This triple helix arrangement generates a cooperative quaternary structure stabilized by numerous hydrogen bonds. Collagen has an unusual amino acid

composition, which is distinctive in that it contains a regular arrangement of amino acids for all three chains. It contains large amounts of glycine and proline in motifs of X–Y–Gly where X is any amino acid (including proline) and Y is often proline or hydroxyproline.[72] Therefore, glycine makes up a third while proline or hydroxyproline makes up a sixth of the sequence. The relatively high content of proline or hydroxyproline, which has geometrically strained carboxy and amino groups, together with the high content of glycine makes collagen makes it possible for collagen to spontaneously form the left–handed helix with use intra–chain hydrogen bonding.[72] The location of glycine at third position of the monomer is critical for stability because it is the smallest amino acids and is usually put in the interior of the helix during assembly to allow for space for the Proline rings that point outwards.[72] This arrangement accounts for the stability of the triple helix.

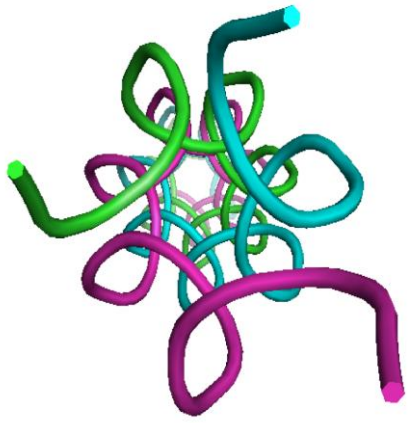


Figure 19. Triple Helix Structure of Collagen. Each poly–L–proline strand is colored differently.[52]

5.2. Methods

Energy minimization of collagen took 1,000,000 steps using the steepest descents algorithm followed by 100,000 steps using the conjugate gradient

algorithm. This many steps were required for collagen unlike the fewer steps for the other proteins because of its rigid structure. The minimized structure was slightly refined by the conjugate gradient algorithm. These together helped in correcting distortions in bond lengths and angles that existed in the protein. Implicit solvent treatment was achieved by inclusion of a constant dielectric of 86.32 for water at 4 ^{o}C. This was used so that the conditions of the results matched experimental conditions used for obtaining the SRCD spectra with which the results were to be compared. Molecular dynamics of the minimized structures were performed at 300 K using the Discover module of Insight®II (Accelrys, San Diego). Mutation was not done for collagen because the structure used was a triple helix model with an amino acid make up containing two prolines and a glycine. The structures were prepared for CD calculation using COREUL and CDCALC used to obtain CD data for all Applequist parameters excluding the general parameters. The output files generated by CDCALC were analyzed using the graphing software package OriginPro 7.5. (OriginLab Corporations, Northampton, MA).

5.3. Results and Discussion:

CAPPS did not recognize the secondary structure of poly–L–proline, which make it incompatible with collagen. Therefore, the results obtained were those for CDCALC only. The CDCALC results did produce the intense negative band for the full trimer of collagen, but it was red shifted compared to the SRCD at 200 nm with the best agreement obtained with the 4000 cm^{-1} bandwidth (Figure 20 and Table 6). The smallest root square deviation (M) between experiment and calculated CD using CDCALC was 58.070 using the original parameters (Table 6). Interestingly, the next best predictions occurred with

the α–Helical parameters Ho, Hx and Hy with only very small differences among them. These H parameters were less red–shifted than the poly–L–proline J parameter predictions. Visually, comparing the location and intensities of the peaks of CDCALC with SRCD, the original parameters at 6000 cm–1 bandwidth produced the best fit most notably with the original parameter predictions resembling the shape of the negative band better. Molecular dynamics of collagen using the same procedure as the other three proteins did not provide enough structural variation to be detectable with the CD calculations.

Table 6. Collagen CD Analysis for Single Minimized Structure Running CDCALC[52]

[a]CD Method	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
4000 Ho	192	9.83	5.96	204	–16.00	12.45	85.812
6000 Ho	191	2.59	8.08	205	–8.87	17.33	82.117
4000 Hx	192	9.83	5.96	204	–16.00	12.45	85.813
6000 Hx	191	2.59	8.08	205	–8.87	17.33	82.117
4000 Hy	192	9.83	5.96	204	–16.00	12.45	85.813
6000 Hy	191	2.59	8.08	205	–8.87	17.33	82.117
4000 Jo	198	9.25	6.31	210	–15.05	13.21	107.908
6000 Jo	197	2.41	8.65	212	–8.30	18.48	96.133
4000 Jx	198	9.25	6.31	210	–15.05	13.21	107.908
6000 Jx	197	2.41	8.65	212	–8.30	18.48	96.133
4000 Jy	198	9.25	6.31	210	–15.05	13.21	107.908
6000 Jy	197	2.41	8.65	212	–8.30	18.48	96.133
4000 O	186	11.95	8.07	205	–27.61	17.95	71.317
6000 O	184	3.93	10.25	204	–17.30	21.28	58.070
SRCD[b]				200	–12.11	21.47	0.000

[a]The 4000 and 6000 refer to the bandwidth in wavenumbers (cm^{-1}).
[b]SRCD[31] had no positive band between 180–190 nm. It did have a small positive band in at 223 nm, with an intensity of 0.09.

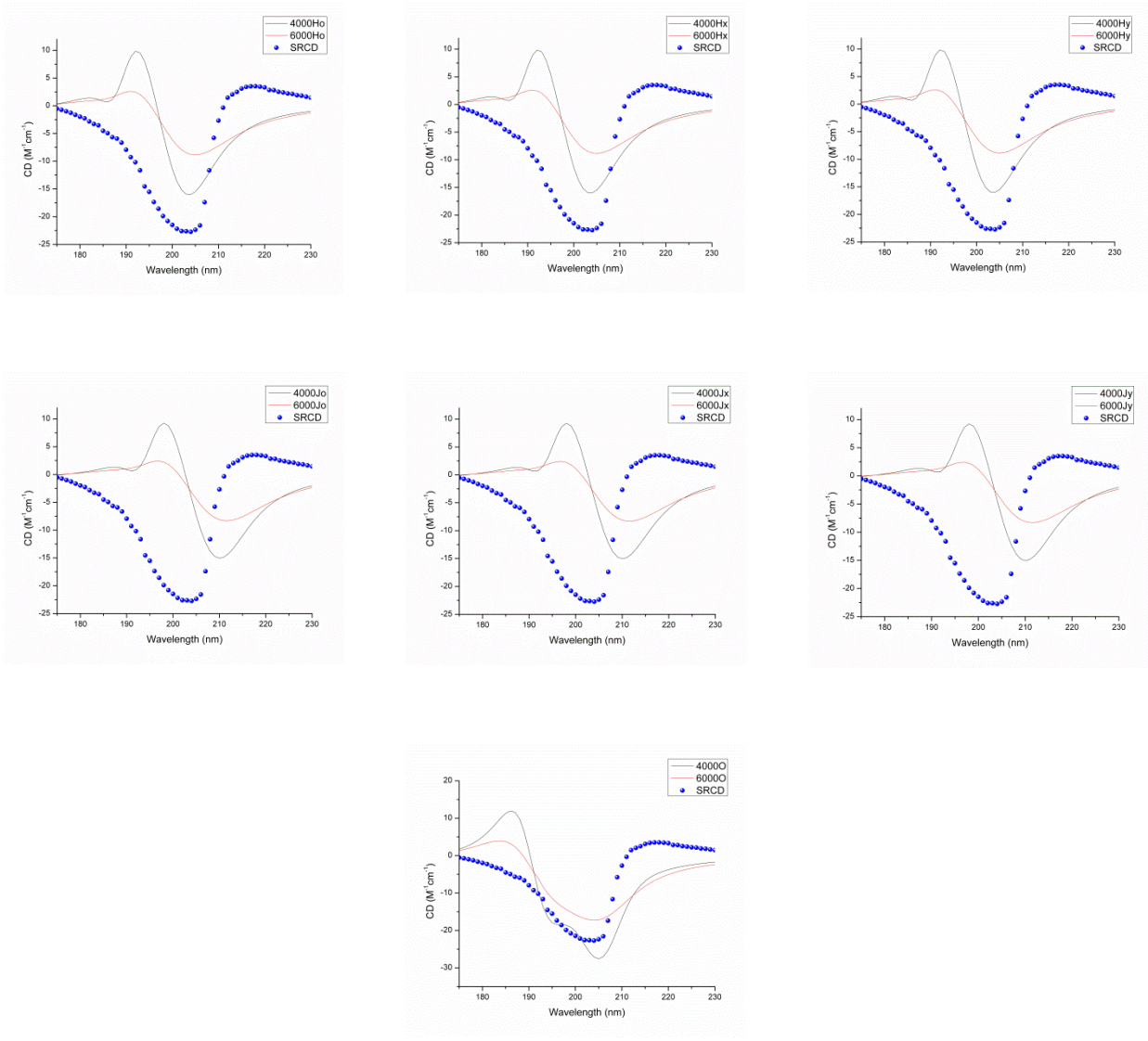


Figure 20. Collagen CD Spectra Using CDCALC.[52] Predictions of the minimized collagen using a bandwidth of 4000 cm^{-1} (black curve) and 6000 cm^{-1} (red curve). The blue dots are the experimental SRCD.[31]

The results were also compared with experimental CD results of poly–L–Pro II reported by Jenness and Sprecher[73] and it was found that the trimer did quite well at predicting the large negative band at ~200 nm using a bandwidth of 6000 cm^{-1} with the

best match obtained using the original (O) parameters (Figure 21). The best match in terms of morphology was obtained using the Jy parameters for both 4000 cm^{-1} and 6000 cm^{-1} but the intensities of the peak were smaller than experiment (Figure 21).

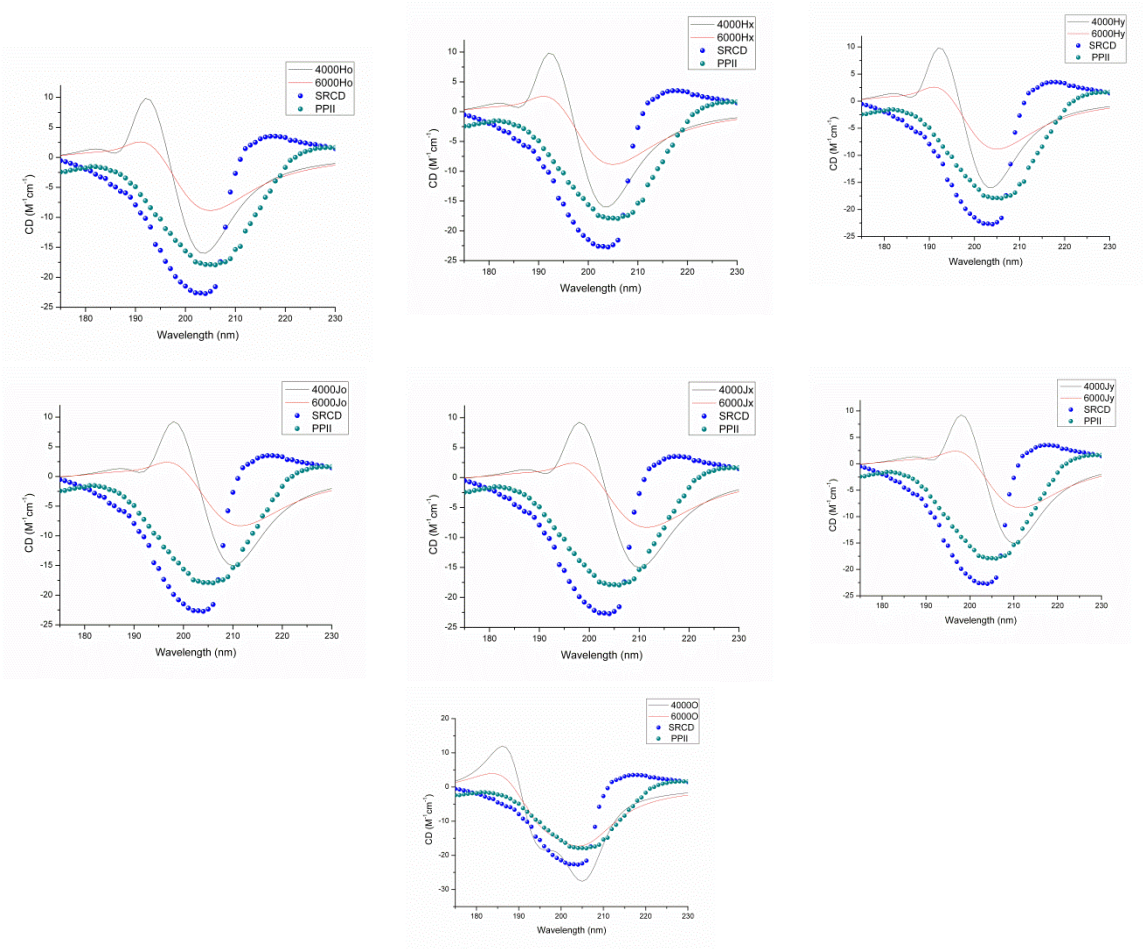


Figure 21. Comparing Collagen CD Spectra obtained using CDCALC with Poly–Proline II and SRCD.[52] Black plot is for 4000 cm^{-1} while red plot is for 6000 cm^{-1}. The blue and green dots are Poly–L–Proline II CD[73] and the experimental SRCD[5] respectively.

The CD prediction for collagen was very challenging because of its rigid structure. It was not possible to treat the hexamer as one unit, therefore the first trimer (containing the first three chains) was used as reported in Figure 21. Previous work had

been done on the monomer by Forlemu *et al.*,[2] and they found that the monomers predicted CD as intense as the poly–L–Proline II experiment as well as SRCD (Figure 22). Also the matrix methods[74] have also been used to fairly predict the CD of poly–L–proline II. Even though red shifted, the original parameters have been able to predict the large negative band at 200 nm of SRCD.

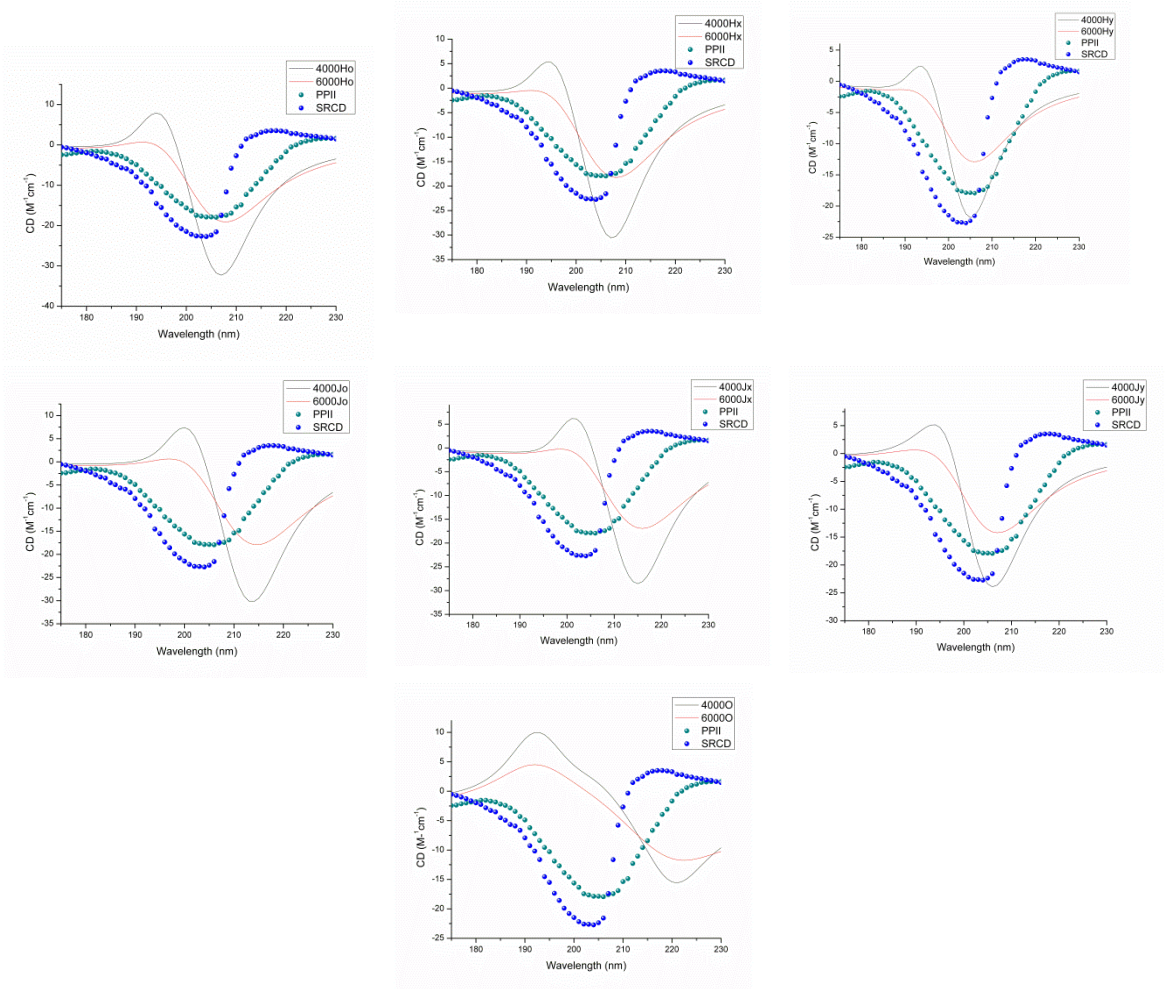


Figure 22. CD Spectra of Collagen Monomer Using CDCALC. Black plot is for 4000 cm^{-1} while red plot is for 6000 cm^{-1}.[52] The blue and green dots are Poly–L–Proline II CD[73] and the experimental SRCD[5] respectively

5.4. Conclusion

Collagen is a very rigid structure and required more minimization than other proteins like lysozyme. The prediction of the spectra obtained gives hope that even more minimization and good molecular dynamics can provide for better prediction than has been done thus far. Considering that the trimer did well with predicting SRCD, more work with the complete hexamer might provide even greater results in the future.

CHAPTER 6

EXPERIMENTAL CIRCULAR DICHROISM AND SECONDARY STRUCTURE ANALYSES OF HORSE HEART MYOGLOBIN AND PEA LECTIN AS AN EXPERIMENT FOR THE UPPER DIVISION UNDERGRADUATE LABORATORY

6.1. Introduction

This chapter is based on experimental circular dichroism experiments conducted as a project for developing a physical chemistry laboratory exercise for senior undergraduate level physical chemistry students. It involves the use of two proteins (myoglobin and pea lectin) as test proteins for the CD project development. These proteins were selected because myoglobin is made mostly of α–helices while pea lectin is predominantly composed of β–sheets. The exercise is designed to teach students how to successfully measure CD, calculate the concentration and do secondary structure analysis of the proteins.

6.2. Characteristics of CD Spectra

Generally, secondary structure features of proteins provide characteristic bands in three different wavelength regions of the electromagnetic spectrum. These bands are obtained from electronic transitions and are based on the nature and energy of the transitions occurring at different wavelength regions.[75] These regions include; the far UV region (below 250 nm), where contributions to CD come from amides in the proteins;

near UV region (250 – 300 nm), where contributions to CD are from aromatic side chains; and the near UV–Vis region (300 – 700 nm), where contributions to CD come from extrinsic chromophores in the proteins.[75] Conventional CD measured in the UV wavelength range (~190 – 250 nm) is well suited for characterizing secondary structural elements such as α–helices and β–sheets, but provide no information on tertiary structures because the transitions that are picked up by a conventional CD instrument are limited to far UV region.[18] Typically, for an α–helix, the negative band at ~222 nm and both the positive and negative bands at ~190 nm and ~208 nm respectively, represent the $\pi - \pi^*$ transitions occurring between the ground and excited state molecular orbitals in the protein. In the case of β–sheets, there is a negative band at ~215 nm, which corresponds to the $n - \pi^*$ transition, and two bands (positive and negative) at ~198 nm and ~175 nm respectively corresponding to $\pi - \pi^*$ transitions.[20a] For polyproline II–type structures, there is a positive band at ~220 nm corresponding to $\pi - \pi^*$ parallel transition and a negative band at ~200 nm for $\pi - \pi^*$ perpendicular transition.[76] With the use of SRCD, additional features are observed in the VUV region. For helices, there is a band at ~175 nm corresponding to $n - \pi^*$ transitions.[18]

6.3. Experimental Circular Dichroism Procedure

6.3.1. Preparation of Protein Solutions

The spectrum of horse heart myoglobin was done in millipore water while that of pea lectin was done in phosphate buffer saline. Before using deionized millipore water it was deoxygenated by sparging with nitrogen. Also, distilled water was used and before use it was also degassed. Phosphate buffer saline is a water-based solution containing

sodium chloride and sodium phosphate. The phosphate group in the buffer helps in maintaining the solution at a constant pH when used for the experiment. These conditions were chosen because these proteins were soluble in the solvents respectively. For myoglobin, to obtain a solution of 0.1 mg/mL concentration, 0.0053 g (5.3 mg) of lyophilized horse heart myoglobin powder was weighed out and added to 5 mL of deionized water. An additional 300 μg of the lyophilized powder was added to the solution to make an approximate 0.1–mg/mL solution. For pea lectin, 1 mg of the protein was weighed out and 10 mL of phosphate buffer saline added to make 0.1 mg/mL of solution. These solutions were each vortexed using a VWR Genie 2 vortexer and filtered with a 0.22 μm Millex–GV filter using a 10 mL syringe. In order to accurately determine the concentration of the protein non–destructively, the absorbance at 280 nm was used.[19] The buffers used were produced as described below.

For myoglobin, good spectra were obtained using Millipore water while for pea lectin, 0.01M phosphate buffered saline was used as buffer. Phosphate buffered saline was purchased from Sigma Aldrich and came as powder in foil pouches. The buffer solution was prepared by dissolving a pouch full powder in one liter of deionized water to produce a 0.01M concentration of the solution with pH 7.4. According to the manufacturer, the composition of the prepared buffer should include NaCl at 0.138 M and KCl at 0.0027 M. The buffers were chosen because they did not absorb in the region of interest (far UV) and were not optically active.

6.3.2. Preparation of Denaturant and Denaturation of Proteins

The proteins were denatured because their concentration measurements could only be obtained using their molar extinction coefficients for absorbance at 280 nm. The extinction coefficients, which are usually obtained for the linear sequence of the proteins, do not take into account the secondary structures of the proteins. Therefore, denaturation renders the proteins in single chains making their absorbance values suitable for concentration calculations.[11] The denaturant used for these experiments was 6 M guanidine HCl. It was prepared to a pH of 6.4 by dissolving 0.0577 g (57.7 mg) of guanidine HCl powder in 75 mL of water. The solution was stirred using a magnetic stirrer. Titration using 1 M NaOH and 1 M HCl solutions was employed to obtain the desired pH value of 6.4 as recommended by the manufacturer. The process of denaturing the protein involved pipetting 5 mL each of the prepared protein solution and the denaturant solution into a clean bottle and shaking the solution. This produced a one-to-one ratio from both solutions.

6.3.3. Measurement of Absorbance

Measurement of the absorbance was done in three runs for each protein. A Shimadzu Spectrophotometer was loaded with two 1 cm standard near IR to UV spectrophotometric quartz cuvettes from SHIMADZU® both containing the buffer and the instrument was set to measure at 280 nm. After measurement of the baseline absorbance, one of the cuvettes was emptied and the buffer replaced with the protein solution and the absorbance measured again at 280 nm. The concentration was calculated as shown in section 6.3.6.

6.3.4. Preparation of the Calibration Solution

The calibration solution was prepared by weighing out 0.0065 g (6.5 mg) of 1S–(+)–10–camphorsulphonic acid (CSA) into 100 mL of deionized water and vortexing in order to dissolve the powder. This produced a 0.06 % w/v solution. The spectrum of CSA is shown below with the peak appearing at 290 nm.

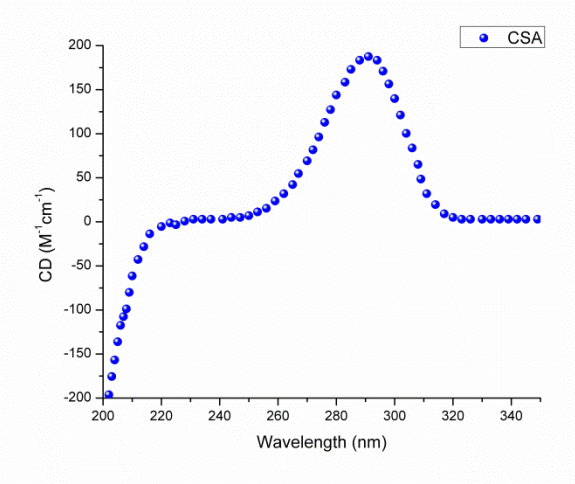


Figure 23. Camphorsulphonic Acid (CSA) spectrum.

6.3.5. Measurement of Circular Dichroism Spectra

The CD Spectra of interest are those of the far ultraviolet (UV) region of the electromagnetic spectrum. Far UV CD spectra were recorded on a model J–810–150S spectropolarimeter (JASCO) equipped with a JASCO Peltier temperature controller (model No 423S). Measurements were done in a 1–mm path length type 21 quartz cuvette from WILMAD® at room temperature. Spectra were recorded from 164 to 240 nm with a step size of 1 nm, a data pitch of 0.1 nm, and a scanning speed of 20 nm/min.

The scanning mode was set to continuous and a response time of 2 seconds was used. Scanning speeds were varied to experiment its effect if any on the spectra of the protein. Trial scans were done using water only and averages of 3 scans were recorded. All spectra were corrected by subtracting the background spectrum of the water from that of the average. The results were obtained as mean residue ellipticity (MRE).

6.3.6. Determination of Protein Concentration

The concentrations of the proteins were calculated using the absorbance measurements obtained from the spectrophotometer and their respective theoretical extinction coefficients obtained from literature. Knowing the sequence of the protein, the theoretical extinction coefficient at 280 nm, ε280, was estimated using the Equation [23][77]

$$\varepsilon_{280}\ (M^{-1}cm^{-1}) = (5500 \text{ X nTrp}) + (1490 \text{ X nTyr}) + (125 \text{ X nS-S}) \quad [23]$$

where the numbers 5500, 1490 and 125 are the molar extinction coefficients for tryptophan (Trp), tyrosine (Tyr), and cystine (the disulfide bond, S–S), respectively, and nTrp, and nS–S are the number of tryptophan, tryrosine and cystine (disulfide) bonds present in the protein respectively. Beers Law, shown in Equation [24], was used to calculate the protein concentration from the extinction coefficients and the absorbance measurements.

$$A_{280\ nm} = (\varepsilon_{280\ nm})(l)(c) \quad [24]$$

where (c) is the protein concentration in M, A_{280nm} is the measured absorbance at 280 nm, ε_{280} is the calculated extinction coefficient at 280 nm, and *l* is the cuvette path length in cm. The concentration of the protein in mg/mL was later obtained by multiplying the concentration in M by the molecular weight of the protein.

For horse heart myoglobin, the extinction coefficient was calculated using the following information; two tryptophan residues, two tyrosine residues, and no disulfide bonds.[78]

$\varepsilon 280 = (5500 \times 2) + (1490 \times 2) + (125 \times 0) = 13980\ M^{-1}cm^{-1}$.

The absorbance at 280 nm (A_{280nm}) was obtained to be 0.0735. With the cell path length of 1 cm, the concentration of the protein, [Protein], was be calculated to be

[Protein] (mg/mL) = 0.0735 /(13980 $M^{-1}cm^{-1}$)(1 cm) X 16951.49 Da = 0.0891 mg/mL.

For pea lectin, the extinction coefficient was calculated using the following information; five tryptophan residues, ten tyrosine residues, and no disulfide bonds.

$\varepsilon 280 = (5500 \times 5) + (1490 \times 10) + (125 \times 0) = 42400\ M^{-1}cm^{-1}$.

The absorbance at 280 nm (A_{280nm}) was obtained to be 0.0978. With the cell path length of 1cm, the concentration of the protein, [Protein], was be calculated to be

[Protein] (mg/mL) = 0.0978 /(42400 $M^{-1}cm^{-1}$)(1 cm) X 30269.7 Da = 0.0698 mg/mL.

This method of determining the protein concentration was used because it has the advantage of being simple, fast, and reliable.

6.4. Secondary Structure Analysis

Secondary structural analysis was done using the DichroWeb server. This server, which can be located at http://dichroweb.cryst.bbk.ac.uk/html/home.shtml was used to analyze the secondary structural elements of the horse heart myoglobin and pea lectin in order to confirm the presence and percentages of α–helices, β–sheets, turn structures and other unordered structures which may be present in the proteins. Inputs used were data obtained from the JASCO spectropolarimeter as well as data deposited into the PCDDB for the proteins and analysis was done in that order. After logging into the webpage with an ID and password, CD data was filled in and the algorithm SELCON 3 used. SELCON 3 was chosen because it gave the best results compared to the other algorithms. Also, most of previous secondary structural analyses that have been done have had to use SELCON 3. The analysis was performed using both conventional CD data obtained from the spectrophotometer and SRCD data downloaded from the PCDDB. In the case of the analysis of input from the PCDDB, protein codes CD0000047000[20b] for horse heart myoglobin CD0000053000[20b] for pea lectin. The data obtained from the PCDDB is not formatted for dichroweb and so needs formatting as described in Appendices 6 and 7.

6.5. Results

The students successfully obtained CD spectra for natured and denatured myoglobin and also for natured pea lectin, but not for the denatured sample of pea lectin. The reported spectra were obtained from plotting data obtained from reference datasets and from the experimental data obtained from the CD measurements. The reconstructed

data were plots from the reference datasets. These are necessary because they have been found to very well describe the secondary structural elements of the reference proteins, which are used for comparing with the test protein of interest. The spectrum for denatured pea lectin was measured later. Figures 23 and 24 show the spectra of the natured and denatured myoglobin respectively as produced by DichroWeb after analysis.

6.5.1. Conventional CD Analyses

Conventional CD produced quite good spectra as seen in results produced by natured and denatured horse heart myoglobin (Figures 23 and 24 respectively), and natured and denatured pea lectin (Figures 25 and 26 respectively).

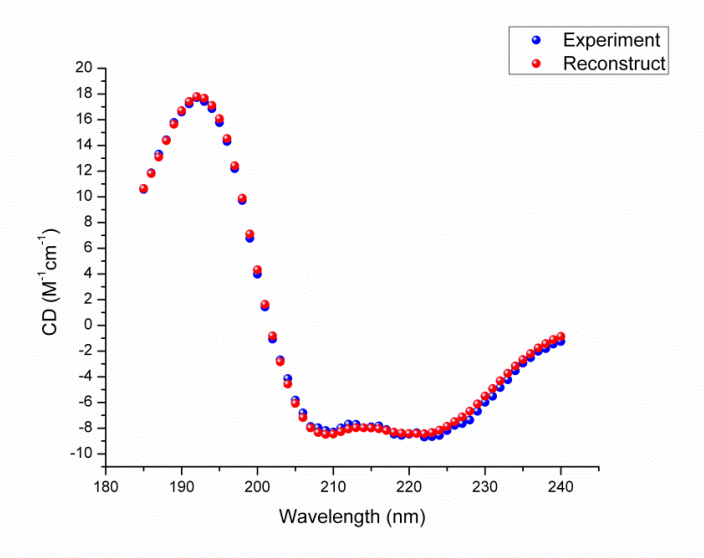


Figure 24. Conventional CD of Natured Horse Heart Myoglobin. The blue dots represent experimental data while the red dots represent the reconstructed data from reference dataset 3 of DichroWeb. Reference set 3 is X-ray data obtained from 29 proteins (both helical and sheet). In DichroWeb, they are optimized for the 185 – 240 nm wavelength range.

Secondary structure analysis of natured horse heart myoglobin (see Table 7) confirmed the protein to be predominantly α-helical. Using reference dataset 3, DichroWeb correctly guessed the protein to be myoglobin and predicted 5.816 helices for every 100 residues and the average length of the helices was calculated to be 14.208 residues. There were with no β-sheets present and only a small amount of turns or other structures.

Table 7. Secondary Structural Predictions for Natured Horse Heart Myoglobin Using Conventional CD. NRMSD is 0.033

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Guess	0.582	0.222	0.000	0.000	0.052	0.144	1
SVD	0.593	0.314	0.001	0.026	0.190	0.302	1.426
Convergent	0.594	0.233	-0.004	-0.003	0.042	0.149	1.011
Stage2	0.594	0.233	-0.004	-0.003	0.042	0.149	1.010
Final	0.594	0.233	-0.004	-0.003	0.042	0.149	1.010

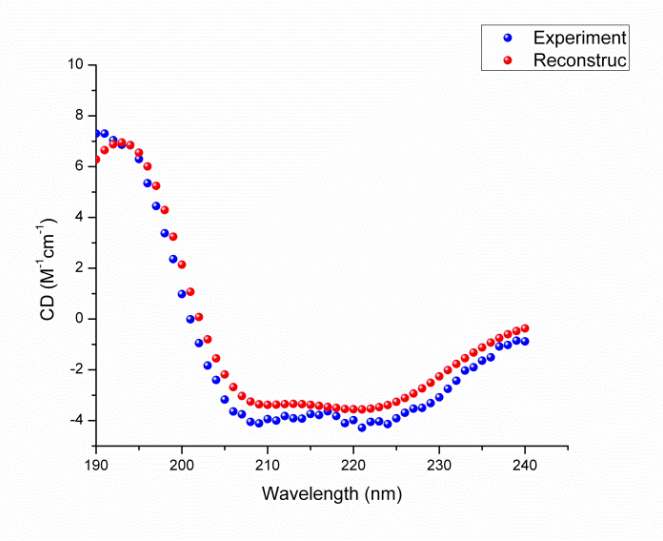


Figure 25. Conventional CD for Denatured Horse Heart Myoglobin. The blue dots represent experimental data while the red dots represent the reconstructed data from reference dataset 3 of DichroWeb.

In the case of denatured horse heart myoglobin and using the same method and reference dataset, DichroWeb showed some helices at reduced amounts of 3.643 helices per 100 residues with average length of 10.272. However, because of denaturation the analysis showed β-strands structure at a rate of 3.219 strands per 100 residues with average helix length reported as 4.809. There was also an increase in unordered structures. The predicted percentage values are found on Table 8.

Table 8. Secondary Structural Analysis of Denatured Horse Heart Myoglobin Using Conventional CD. NRMSD is 0.765

Result	**Helix1**	**Helix2**	**Strand1**	**Strand2**	**Turns**	**Unordered**	**Total**
Guess	0.150	0.147	0.041	0.068	0.235	0.359	1
SVD	0.167	0.157	0.043	0.076	0.246	0.374	1.063
Convergent	0.153	0.147	0.043	0.068	0.235	0.357	1.003
Stage2	0.152	0.147	0.042	0.068	0.234	0.357	1.001
Final	0.228	0.146	0.090	0.064	0.184	0.248	0.960

Even though present in the denatured protein, the helix signals were less intense than were in the natured protein and the percentage structural value of the sum of Helix 1 and Helix 2 dropped from 0.827 to 0.374 (see Figure 24 and Table 8). This is why there was a decrease reported in the number of helices per 100 residues (3.643 from 5.816) and the average length of helices changed from to 14.208 to 3.790. However, denaturation did tend to give the protein mixture some β-strand character as seen in the increase of the percentage structural value of the sum of strands 1 and 2 from -0.007 to 0.154. The negative value in the percentage just means that there was none evidence of the structure on question. Similarly, the percentages increased for the turns and unordered structures (0.042 and 0.142 to 0.184 and 0.248) respectively.

Pea lectin produced a spectrum, which depicted the presence of β–sheets. The intensities were good enough such that the broad negative peak spanning 220 to 300 nm was recorded (Figure 25). This matched the spectrum of native pea lectin obtained by Khan and Naseem.[79] Analysis of the spectrum for a wavelength range below 195 and above 240 nm was not possible with DichroWeb because of the wavelength limitation of the reference dataset 6, which gave the best result (All datasets are found online at the DichroWeb server site). Also, the spectra obtained by Khan and Naseem only went as low as 200 nm. For pea lectin, secondary structural analysis of the natured state of the protein showed that the protein is made of mostly of β-sheets. There were 7.945 strands per 100 residues and only 2.147 helices per 100 residues. The average length of the strands was 5.720 while that of the helices was 4.313. DichroWeb guessed γ-crystallin to be the closest match to the pea lectin, which is reasonable because γ-crystallin is exclusively made up of β-sheets. These are shown on Table 9.

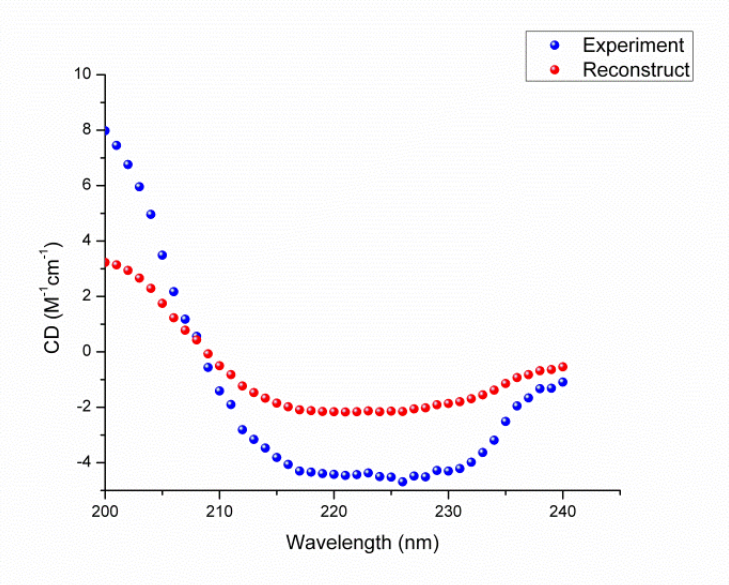


Figure 26. Conventional CD of Natured Pea Lectin. The blue dots represent experimental data while the red dots represent the reconstructed data from reference dataset 3 of DichroWeb.

Table 9. Secondary Structural Analysis of Natured Pea Lectin Using Conventional CD. NRMSD is 0.691

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Guess	0.006	0.086	0.299	0.161	0.109	0.339	1
SVD	0.042	0.081	0.181	0.088	-0.004	0.355	0.743
Convergent	0.006	0.086	0.297	0.160	0.107	0.338	0.994
Stage2	0.007	0.086	0.296	0.159	0.106	0.338	0.991
Final	0.007	0.086	0.296	0.159	0.106	0.338	0.991

After denaturation, the spectrum obtained for pea lectin showed an immense loss of the secondary structural elements (Figure 26). The negative band at 222 nm is very less intense even though present.

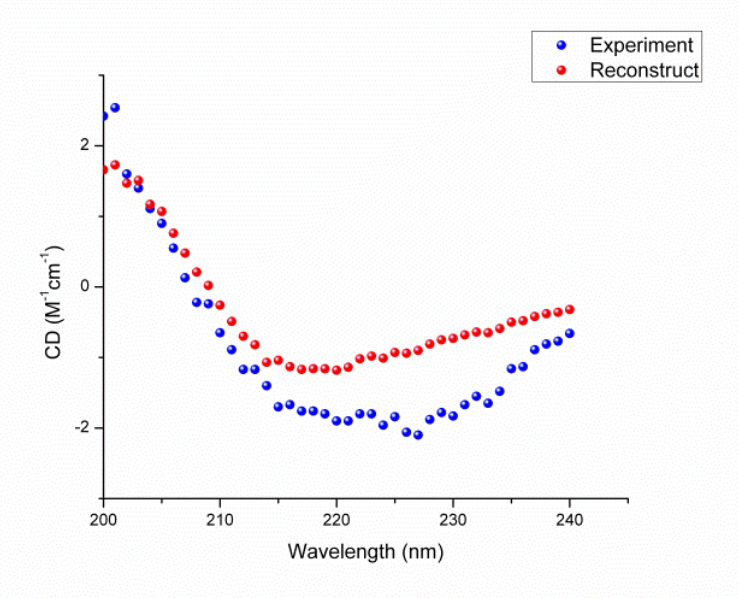


Figure 27. Conventional CD of Denatured Pea Lectin. The blue dots represent experimental data while the red dots represent the reconstructed data from reference dataset 3 of DichroWeb.

Secondary structural elements of pea lectin still appeared to be detected by DichroWeb after denaturation. The number of strands counted per 100 residues was 9.618 while the number of helices was 0.702 per 100 residues (see Table 10). The average length of helices was 4.139 while that for strands was 5.720. The NRMSD was

observed to be 0.742 compared to 0.691 for the natured form of the protein. However, the spectrum of the natured protein was more intense than the reconstructed one. The increase in strand component and decrease in helix component may be due to reorganization with denaturation conferring strand character with loss of helix character.

Table 10. Secondary Structural Analysis of Denatured Pea Lectin Using Conventional CD. NRMSD is 0.742

Result	**Helix1**	**Helix2**	**Strand1**	**Strand2**	**Turns**	**Unordered**	**Total**
Guess	0.000	0.028	0.294	0.196	0.229	0.252	0.999
SVD	0.050	0.026	0.137	0.086	0.085	0.396	0.780
Convergent	0.000	0.028	0.289	0.192	0.225	0.258	0.993
Stage2	0.001	0.028	0.289	0.192	0.225	0.258	0.993
Final	0.001	0.028	0.289	0.192	0.225	0.258	0.993

6.5.2. SRCD Analyses

Analysis of myoglobin data obtained from the PCDDB (see Figure 27 and Table 11) confirmed the protein to be predominantly α-helical. Using reference dataset 3, DichroWeb correctly guessed the protein to be myoglobin and predicted 5.860 helices for every 100 residues and the average length of the helices was calculated to be 13.827 residues. Similarly, no β-sheets were present and only a small amount of turns or other structures.

Table 11. Secondary Structural Predictions for Natured Horse Heart Myoglobin Using SRCD from the PCDDB (CD0000047000).[31] NRMSD is 0.063

Result	**Helix1**	**Helix2**	**Strand1**	**Strand2**	**Turns**	**Unordered**	**Total**
Guess	0.582	0.222	0.000	0.000	0.052	0.144	1
SVD	0.536	0.278	0.039	0.043	0.188	0.300	1.384
Convergent	0.576	0.234	0.001	0.000	0.062	0.168	1.041
Stage2	0.576	0.234	0.001	0.000	0.062	0.168	1.041
Final	0.576	0.234	0.001	0.000	0.062	0.168	1.041

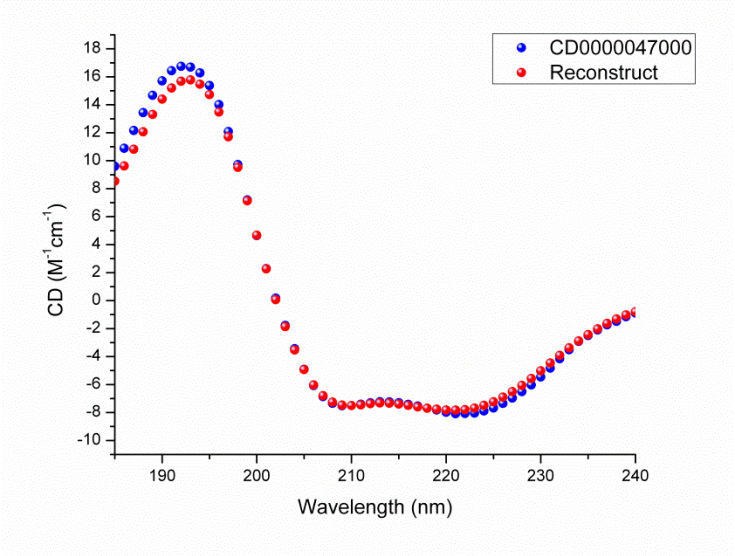


Figure 28. SRCD of Natured Horse Heart Myoglobin from PCDDB (CD0000047000)[31]. The blue dots represent crystal structural data from PCDDB while the red dots represent the reconstructed data from reference dataset 3 of DichroWeb.

For pea lectin data obtained from the PCDDB (see Figure 28 and Table 12) also confirmed the protein to be predominantly β-sheets. Using reference dataset 6, DichroWeb guessed the protein to be concanavalin A, another predominantly β-sheet protein, and predicted 6.867 sheets for every 100 residues and an average length of the β-sheet to be 5.458 residues.

Table 12. Secondary Structural Predictions for Natured Pea Lectin Using SRCD from PCDDB (CD0000053000).[31] NRMSD is 0.404

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Guess	0.000	0.038	0.329	0.135	0.236	0.262	1
SVD	0.069	0.055	0.119	0.062	0.091	-0.032	.0364
Convergent	0.025	0.044	0.240	0.131	0.227	0.324	0.991
Stage2	0.022	0.042	0.240	0.132	0.227	0.325	0.990
Final	0.024	0.042	0.237	0.137	0.222	0.323	0.985

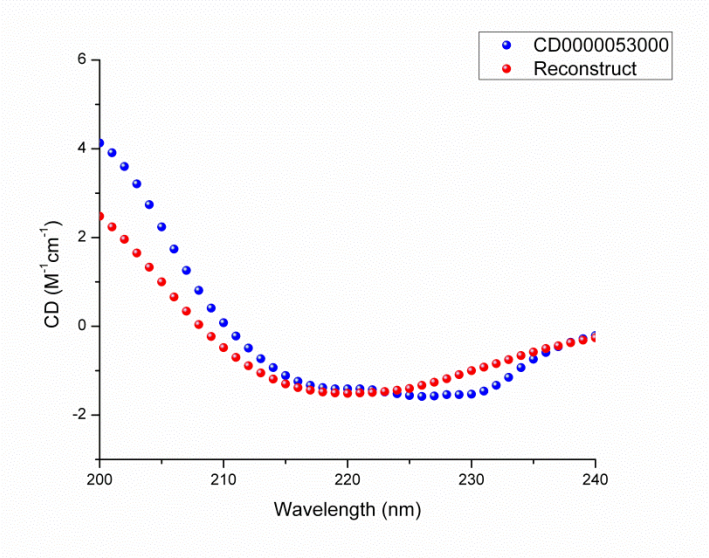


Figure 29. SRCD of Natured Pea Lectin from PCDDB (CD0000053000)[31]. The blue dots represent crystal structural data from PCDDB while the red dots represent the reconstructed data from reference dataset 6 of DichroWeb.

6.6 Discussion

The experiments were designed such that they were easy to perform and the data was really saved into text files after each measurement. The students obtained great spectra for myoglobin, but not the best spectra for pea lectin. The not so good spectra for pea lectin might be because our protein solution for CD was less concentrated than SRCD. We made a 0.1 mg/mL solution but according the PCDDB, the pea lectin solution used for obtaining the SRCD spectra was 14.5 mg/mL. Also, our solution was made using 0.1 M phosphate buffer saline as stipulated by our protein supplier (Sigma Aldrich®) while the SRCD spectrum was obtained for a pea lectin (from Calbiochem®) solution made using water as buffer. Also our pea lectin was a fluorescein isothiocyanate

(FITC) conjugated at 2 – 4 mole FITC per mole protein (Lowry). The pea lectin used for cCD was designed for sedimentation equilibrium studies. These could have affected the results (See Appendix 9). Overall, by doing the exercise, the students learned the experimental skills involved with measuring the circular dichroism of proteins, in this case for the mostly α-helical protein myoglobin and the β-sheet protein pea lectin. However, the spectra for the denatured proteins still showed secondary structural elements, even though with very less intense bands. The process of denaturation basically involved pipetting 5 mL aliquots each of the prepared protein and the denaturant into a clean bottle and shaking the solution to produce a one-to-one ratio from both solutions. This requires that in future the proteins be denatured for a longer time or with a more concentrated form of the denaturant. Overnight denaturation may be a great way of ensuring full denaturation. In addition to trying overnight denaturation with guanidine HCl, urea which does not absorb at 280 nm can also be used to assess which of both denaturants leads to a greater loss of the secondary structural elements.

The values expressed in Tables 7 – 12 describe the percentage content of secondary structure analyzed by DichroWeb. Basically, the server goes through a series of steps from guessing the percentages of secondary structural element in the test protein to obtaining the best values of the elements based on comparison with reference structures from the database. It solves singular values of the secondary structural elements by using the input values from the experimental data and the reference proteins, which it enters into matrix form to obtain singular values most representative of both data. All stages involved provide information used at the final stage to get overall percentages of the secondary structural elements.[3]

A comparison of the results obtained from conventional CD with those obtained from SRCD show that both methods are reliable methods for obtaining protein secondary structure. In the case of horse heart myoglobin, conventional CD and SRCD provided NRMSD that were very close (0.033 and 0.063 respectively) while for pea lectin, the NRMSD were different (0.691 and 0.404 respectively). The number of helix and strands per 100 residues and the average helix and strand lengths are summarized in Appendix 8 Table 15. It was noticed that the conventional CD spectrum for pea lectin was more intense than the reconstruct from DichroWeb, thus explaining the difference observed with the NRMSD values, while that horse heat myoglobin only showed a very slight shift from the reconstruct obtained from DichroWeb. This also matches the similarity in the NRMSD values obtained. Because SRCD provides more information by measuring CD at lower wavelengths that cannot be done by conventional CD instruments and because the signal-to-noise ratio is high, thus enabling small sample sizes to be used, SRCD still remains a better option for measuring and predicting protein secondary structure. However, for the physical chemistry laboratory exercise, both methods are reliable at producing spectra representative of the secondary structure of proteins. The students faced fewer problems with obtaining the spectra from the conventional CD instrument and from the PCDDB than with formatting these files for use in DichroWeb. This problem was more evident with the extraction of information from the PCDDB for analysis. They had use Microsoft Excel as a means of formatting their data and then transferred these data to ASCII text format, which is accepted by DichroWeb.[3] In future, a sample file with data obtained from the PCDDB should be provided to the students with directions on formatting before the laboratory exercise as shown in Appendices 6 and 7.

APPENDICES

APPENDIX 1

CD for Lysozyme Comparing DInaMo, Experiment and Reference Calculations.

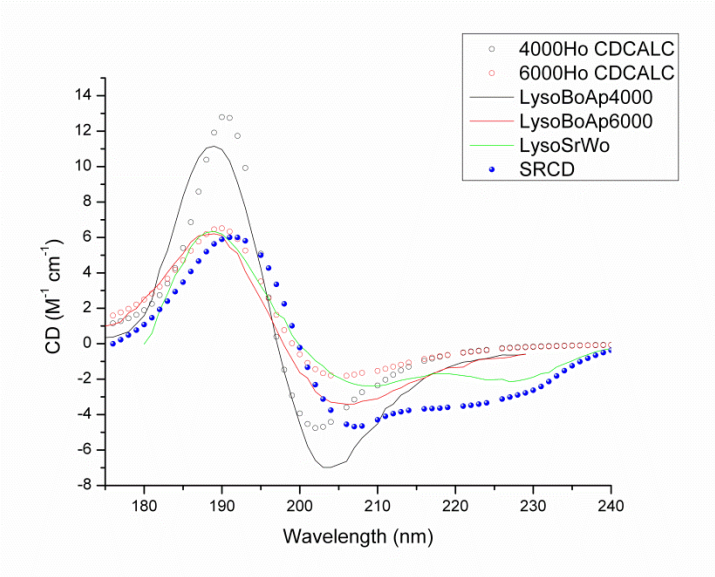


Figure 30: CD for Lysozyme Comparing DInaMo, Experiment and Reference Calculations.[52] The black and red hollow circles represent the CDCALC results using the a–helical Ho parameters. The black and red lines represent CAPPS results by Bode and Applequist[46] using the a–helical Hy parameters. The green line represents the results obtained from the matrix method by Sreerama and Woody.[60] The blue dots are the experimental SRCD (CD0000045000)[31]

APPENDIX 2
Horse Heart Myoglobin MD and CD

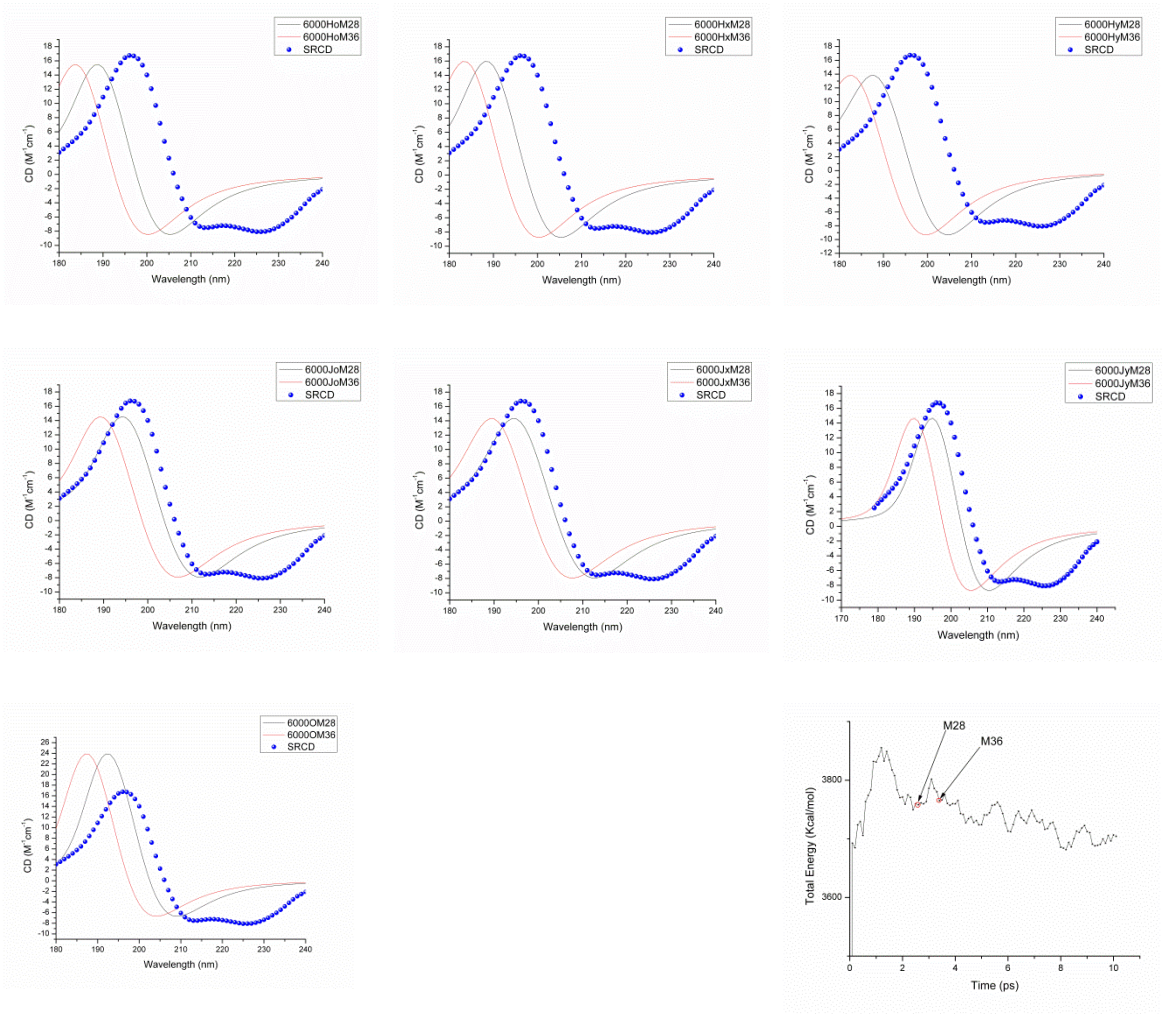


Figure 31. Horse Heart Myoglobin MD and CD.[52] MD cluster analysis comparing two MD snapshots obtained by CDCALC with the SRCD data at bandwidth of 6000 cm^{-1}. The blue dots are the experimental SRCD (CD0000047000).[31] The figure in the lower right corner is the CD trajectory with the two representative snapshots labeled.

APPENDIX 3
Sperm Whale Myoglobin MD and CD

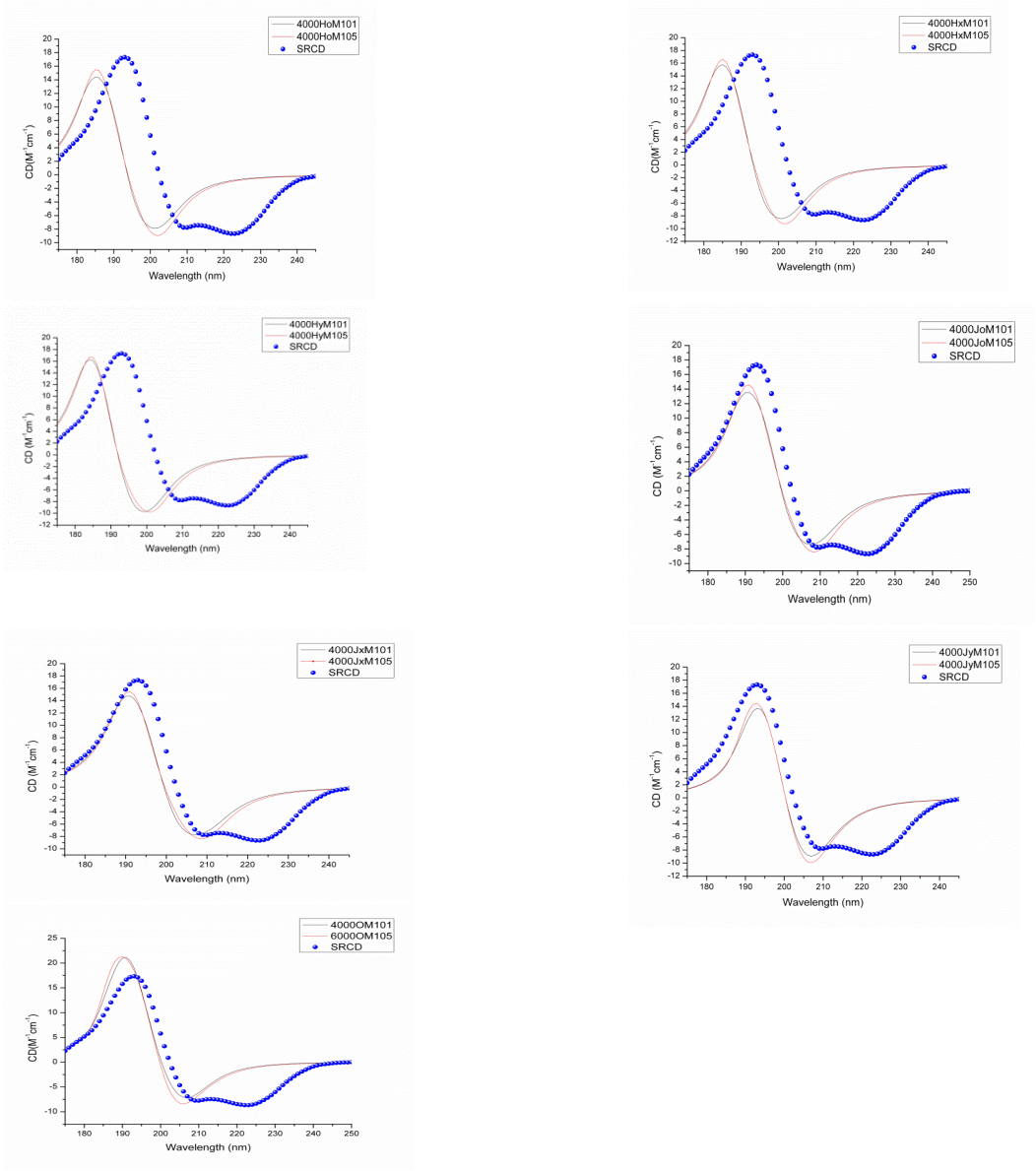


Figure 32. Sperm Whale Myoglobin MD and CD.[52] MD cluster analysis comparing two MD snapshots obtained by CDCALC with the SRCD data at bandwidth of 4000 cm^{-1}. The blue dots are the experimental SRCD (CD0000047000).[31]

APPENDIX 4

Horse Heart Myoglobin CD Analysis for MD Clusters Running CDCALC Poly–L–Proline Jx Parameters and α–Helical Ho Parameters at 4000 cm^{-1} Bandwidth

Table 13. Horse Heart Myoglobin CD Analysis for MD Clusters Running CDCALC Poly–L–Proline Jx Parameters and α–Helical Ho Parameters at 4000 cm^{-1} Bandwidth.[52]

Snap–shot[a]	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
Hx M4	187	22.67	10.98	203	–13.81	13.07	59.620
Hx M13	184	16.07	13.43	201	–12.54	12.90	68.525
Hx M21	185	16.30	13.45	202	–11.35	12.83	60.715
Hx M28	187	16.08	12.77	203	–10.68	12.62	53.143
Hx M36	188	17.14	12.76	202	–11.42	12.47	59.427
Hx M68	186	15.16	13.49	202	–9.24	12.39	58.521
Hx M76	188	16.92	11.79	202	–9.26	13.24	55.628
HxM97	186	15.39	11.97	201	–8.51	12.58	59.811
Jx M4	193	20.52	11.82	210	–12.48	14.09	23.073
Jx M13	190	15.18	13.88	208	–11.34	13.95	27.256
Jx M21	191	15.28	14.01	209	–10.26	13.88	22.063
Jx M28	192	14.65	13.68	210	–9.56	13.63	21.422
Jx M36	193	15.59	13.81	209	–10.34	13.47	21.887
Jx M68	192	14.21	14.19	209	–8.47	13.40	24.297
Jx M76	194	15.50	12.69	209	–8.45	14.37	33.869
JxM97	192	14.43	12.57	208	–7.80	13.72	26.721
SRCD[b]	192	16.75	14.58	209	–7.51	7.78	0.000

[a]For SRCD, there was a 3rd peak at 221 nm with CD intensity of –8.07831[5] and a half–peak bandwidth of 19.17303.

APPENDIX 5

Sperm Whale Myoglobin CD Analysis for MD Clusters Running CDCALC Poly–L–Proline Jy Parameters and α–Helical Hx Parameters at 6000 cm^{-1} Bandwidth.

Table 14. Sperm Whale Myoglobin CD Analysis for MD Clusters Running CDCALC Poly–L–Proline Jy Parameters and α–Helical Hx Parameters at 6000 cm^{-1} Bandwidth.[52]

Snap–shot[a]	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	Peak Wavelength (nm)	Δε (M^{-1} cm^{-1})	Half–peak Bandwidth	M (Root Square Deviation)
Hx M8	185	11.35	16.15	205	–6.09	17.68	52.651
Hx M17	185	10.74	20.35	203	–6.62	16.56	57.206
Hx M24	185	9.94	16.66	204	–5.60	16.94	55.606
Hx M28	185	9.95	21.73	204	–5.56	16.73	55.314
Hx M71	185	10.07	16.63	203	–5.30	16.56	57.410
Hx M101	184	9.48	27.09	203	–4.54	17.06	59.993
Hx M103	184	9.86	16.26	202	–4.73	15.73	61.455
HxM105	184	9.62	18.46	203	–5.00	17.06	59.339
Jy M8	193	9.26	15.02	211	–6.19	17.06	34.441
Jy M17	192	8.89	15.08	209	–6.29	17.03	36.466
Jy M24	193	8.25	14.23	209	–5.36	16.93	39.552
Jy M28	193	8.29	14.89	210	–5.26	17.01	38.683
Jy M71	192	8.53	14.80	209	–5.35	16.47	39.132
Jy M101	192	7.40	15.28	209	–4.43	16.82	43.166
Jy M103	192	7.69	14.52	209	–4.24	16.34	43.423
JyM105	192	7.74	14.66	209	–4.90	16.65	42.061
SRCD[b]	193	17.33	14.59	210	–7.77	7.48	0.000

[a]The 4000 and 6000 refer to the bandwidth in wavenumbers (cm^{-1}).

[b]For SRCD, there was a 3rd peak at 221 nm with CD intensity of –8.07831[5] and a half–peak bandwidth of 19.17303.

APPENDIX 6
Procedure for Obtaining and Formatting Data from the PCDDB

The DichroWeb server, which can be accessed by typing in the web address http://dichroweb.cryst.bbk.ac.uk/html/home.shtml is free and available to all. Before analyzing data obtained from the PCDDB, the following procedure can be used to obtain and format the data into a form that is easily recognized by DichroWeb. An example showing the use of these steps for myoglobin is outlined

The procedure for analyzing the secondary structure of myoglobin using data from the PCDDB is as follows;

1) Type in http://pcddb.cryst.bbk.ac.uk/home.php in a web browser.
2) The homepage already has the plot for horse heart myoglobin displayed with the code CD0000047000.
3) This code is written as a hyperlink. Click on it and it leads you to records uploaded on this protein. A number of tablets ranging from sample on the left to validation on the right provide different options that lead to available information and these can be explored.
4) Click on the tablet by name 'spectra'. This gives a list of different data ranging from the raw spectrum in millidegrees to the final processed spectrum in delta epsilon. For each, the data file can be downloaded as well as the spectrum plot.
5) Click on 'file' and copy the entire data onto your mouse. This data contains two unformatted rows - wavelength (in nanometers) and the circular dichroism (in delta epsilon).
6) Open a new excel spreadsheet and past the copied data in. The unformatted data will be formatted in this sheet as follows
    a. Click on the tab entitled 'data' and then on the 'text to columns' icon. This pops up a new smaller window. Leave the file type as 'delimited' and click on next. Check the space box and click finish. This separates the the columns.
    b. Still in the 'data' tab, click on the icon entitled 'sort'. This pops up a smaller window. In this new window, click on sort. It automatically sorts the data from the smallest to the largest wavelengths.
    c. Change the CD values from scientific notation to decimal point notation by highlighting the column and selecting format in the menu bar of the program. Select cells from the pop-up window and then 'General' for category. This changes the notation as specified above.
    d. Copy the columns into a notepad and save the file as a text file (.txt).

The file is now ready for use in DichroWeb as shown in Appendix 7. The right column shows the formatted file while the left column shows the non-formatted file.

APPENDIX 7
Sample of a Section of both Non-formatted/Formatted Files for Horse Heart Myoglobin (CD0000047000)

Non-formatted from PCDDB

```
250.0 -1.00454E-01
249.0 -9.12808E-02
248.0 -1.60280E-01
247.0 -1.56482E-01
246.0 -2.03289E-01
245.0 -2.25575E-01
244.0 -3.01567E-01
243.0 -3.41532E-01
242.0 -5.28093E-01
241.0 -7.16273E-01
240.0 -8.89179E-01
239.0 -1.14587E+00
238.0 -1.47871E+00
237.0 -1.73180E+00
236.0 -2.10257E+00
235.0 -2.49887E+00
234.0 -2.92532E+00
233.0 -3.50806E+00
232.0 -4.15028E+00
231.0 -4.82662E+00
230.0 -5.46095E+00
229.0 -6.03244E+00
228.0 -6.51179E+00
227.0 -6.97442E+00
226.0 -7.34638E+00
225.0 -7.66687E+00
224.0 -7.89120E+00
223.0 -8.03551E+00
222.0 -8.07427E+00
221.0 -8.07831E+00
220.0 -7.98063E+00
219.0 -7.81885E+00
218.0 -7.68787E+00
217.0 -7.53598E+00
216.0 -7.41567E+00
215.0 -7.29687E+00
214.0 -7.23668E+00
213.0 -7.22589E+00
212.0 -7.30626E+00
211.0 -7.40676E+00
210.0 -7.49723E+00
209.0 -7.51279E+00
208.0 -7.33970E+00
207.0 -6.86258E+00
206.0 -6.07124E+00
205.0 -4.91438E+00
204.0 -3.43745E+00
203.0 -1.77333E+00
202.0  1.77188E-01
201.0  2.28884E+00
200.0  4.65002E+00
```

Formatted (ready for analysis)

```
200      4.650020
201      2.288840
202      0.177188
203     -1.773330
204     -3.437450
205     -4.914380
206     -6.071240
207     -6.862580
208     -7.339700
209     -7.512790
210     -7.497230
211     -7.406760
212     -7.306260
213     -7.225890
214     -7.236680
215     -7.296870
216     -7.415670
217     -7.535980
218     -7.687870
219     -7.818850
220     -7.980630
221     -8.078310
222     -8.074270
223     -8.035510
224     -7.891200
225     -7.666870
226     -7.346380
227     -6.974420
228     -6.511790
229     -6.032440
230     -5.460950
231     -4.826620
232     -4.150280
233     -3.508060
234     -2.925320
235     -2.498870
236     -2.102570
237     -1.731800
238     -1.478710
239     -1.145870
240     -0.889179
241     -0.716273
242     -0.528093
243     -0.341532
244     -0.301567
245     -0.225575
246     -0.203289
247     -0.156482
248     -0.160280
249     -0.091281
250     -0.100454
```

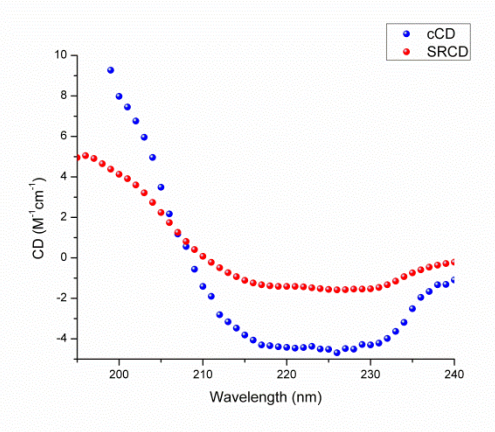
APPENDIX 8

Number of Secondary Structural Elements per 100 Residues and Average Lengths of Secondary Structural Elements for Conventional CD.

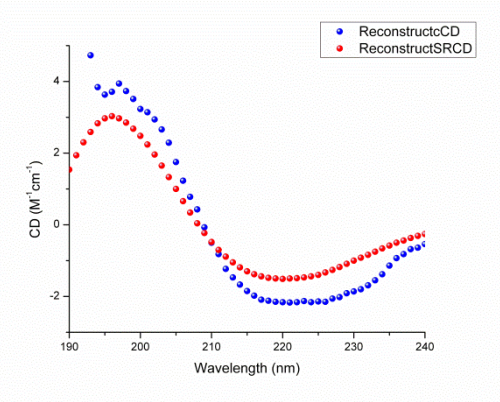
Table 15. Number of Secondary Structural Elements per 100 Residues and Average Lengths of Secondary Structural Elements.

Result	Number of Strands per 100 Residues	Average Length of Strand	Number of Helices per 100 Residues	Average Length of Helices
Natured Horse Heart Myoglobin	0	0	5.816	14.202
Denatured Horse Heart Myoglobin	3.219	4.809	3.643	10.272
Natured Pea Lectin	7.945	5.720	2.147	4.313
Denatured Pea Lectin	9.618	5.720	0.702	4.139

APPENDIX 9
Comparing Conventional CD Versus SRCD for Pea Lectin



(a)



(b)

Figure 33(a,b). Comparing Conventional Pea Lectin Vs SRCD for Pea Lectin

REFERENCES

1. Kelly, S. M.; Price, N. C., The application of circular dichroism to studies of protein folding and unfolding. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1997,** *1338* (2), 161-185.

2. Forlemu, N. Y. Predicting Functional Protein Complexes in the Glycolytic Pathway: Computer Simulations of Compartmentation and Channeling in Glycolysis. University of North Dakota, Grand Forks, ND, August 2009.

3. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004,** *32* (suppl 2), W668-W673.

4. Wallace, B.; Lees, J.; Orry, A.; Lobley, A.; Janes, R. W., Analyses of circular dichroism spectra of membrane proteins. *Protein Sci.* **2003,** *12* (4), 875-884.

5. Wallace, B.; Janes, R. W., Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. *Curr. Opin. Chem. Biol.* **2001,** *5* (5), 567-571.

6. Marassi, F. M.; Opella, S. J., A solid-state NMR index of helical membrane protein structure and topology. *J. Magn. Reson.* **2000,** *144* (1), 150-155.

7. Feng, J. A.; Kao, J.; Marshall, G. R., A Second Look at Mini-Protein Stability: Analysis of FSD-1 Using Circular Dichroism, Differential Scanning Calorimetry, and Simulations. *Biophys. J.* **2009,** *97* (10), 2803-2810.

8. Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K., Crystal Structure of a Ten-Amino Acid Protein. *J. Am. Chem. Soc.* **2008,** *130* (46), 15327-15331.

9. Nath, S.; Meuvis, J.; Hendrix, J.; Carl, S. A.; Engelborghs, Y., Early Aggregation Steps in α-Synuclein as Measured by FCS and FRET: Evidence for a Contagious Conformational Change. *Biophys. J. 98* (7), 1302-1311.

10. (a) Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **2007,** *1* (6), 2527-2535; (b) Greenfield, N. J., Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nat. Protoc.* **2007,** *1* (6), 2891-2899.

11. Greenfield, N. J., Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat. Protoc.* **2007,** *1* (6), 2733-2741.

12. Stjernschantz, E.; Oostenbrink, C., Improved Ligand-Protein Binding Affinity Predictions Using Multiple Binding Modes. *Biophys. J. 98* (11), 2682-2691.

13. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.

14. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614.

15. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How Fast-Folding Proteins Fold. *Science 334* (6055), 517-520.

16. Meersman, F.; Atilgan, C.; Miles, A. J.; Bader, R.; Shang, W.; Matagne, A.; Wallace, B. A.; Koch, M. H. J., Consistent Picture of the Reversible Thermal Unfolding of Hen Egg-White Lysozyme from Experiment and Molecular Dynamics. *Biophys. J. 99* (7), 2255-2263.

17. Finke, J. M.; Jennings, P. A.; Lee, J. C.; Onuchic, J. N.; Winkler, J. R., Equilibrium unfolding of the poly(glutamic acid)20 helix. *Biopolymers* **2007,** *86* (3), 193-211.

18. Lees, J.; Wallace, B., Synchrotron radiation circular dichroism and conventional circular dichroism spectroscopy: A comparison. *Spectrosc.-Int. J.* **2002,** *16* (4), 121-126.

19. Wallace, B. A.; Janes, R. W., *Modern techniques for circular dichroism and synchrotron radiation circular dichroism spectroscopy: 1*. IOS Press: 2009.

20. (a) Bode, K. A.; Applequist, J., Improved Theoretical π-π^* Absorption and Circular Dichroic Spectra of Helical Polypeptides Using New Polarizabilities of Atoms and NC'O Chromophores. *J. Phys. Chem.* **1996,** *100* (45), 17825-17834; (b) Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A., A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **2006,** *22* (16), 1955-1962.

21. Wallace, B. A.; Lees, J. G.; Orry, A. J. W.; Lobley, A.; Janes, R. W., Analyses of circular dichroism spectra of membrane proteins. *Protein Sci.* **2003,** *12* (4), 875-884.

22. Whitmore, L.; Wallace, B. A., Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **2008,** *89* (5), 392-400.

23. Clarke, D. T.; Bowler, M. A.; Fell, B. D.; Flaherty, J. V.; Grant, A. F.; Jones, G. R.; Martin-fernandez, M. L.; Shaw, D. A.; Todd, B.; Wallace, B. A.; Towns-andrews, E., A high aperture beamline for vacuum ultraviolet circular dichroism on the srs. *Synchrotron Radiation News* **2000,** *13* (2), 21-27.

24. Teeters, C. L.; Lodish, H. F.; Ciechanover, A.; Wallace, B. A., Transferrin and Apotransferrin: pH-dependent Conformational Changes Associated with Receptor-mediated Uptakea. *Ann. N. Y. Acad. Sci.* **1986,** *463* (1), 403-407.

25. Lobley, A.; Whitmore, L.; Wallace, B. A., DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **2002,** *18* (1), 211-212.

26. Sreerama, N.; Woody, R. W., Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Anal. Biochem.* **2000,** *287* (2), 252-260.

27. Mao, D.; Wachter, E.; Wallace, B. A., Folding of the mitochondrial proton adenosine triphosphatase proteolipid channel in phospholipid vesicles. *Biochemistry* **1982,** *21* (20), 4960-4968.

28. (a) Sreerama, N.; Woody, R. W., A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. *Anal. Biochem.* **1993,** *209* (1), 32-44; (b) Sreerama, N.; Venyaminov, S. Y. U.; Woody, R. W., Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **1999,** *8* (2), 370-380.

29. (a) Compton, L. A.; Johnson Jr, W. C., Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **1986,** *155* (1), 155-167; (b) Manavalan, P.; Johnson Jr, W. C., Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **1987,** *167* (1), 76-85.

30. Andrade, M. A.; Chacón, P.; Merelo, J. J.; Morán, F., Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **1993,** *6* (4), 383-390.

31. Whitmore, L.; Woollett, B.; Miles, A. J.; Klose, D. P.; Janes, R. W.; Wallace, B. A., PCDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata. *Nucleic Acids Res. 39* (suppl 1), D480-D486.

32. Applequist, J.; Carl, J. R.; Fung, K. K., Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972,** *94* (9), 2952-2960.

33. Applequist, J. In *In Calculation of Electronic Circular Dichroic Spectra by a Dipole Interaction Model, Chirality and Circular Dichroism: Determination and Analyitical Applications*, 5th International Conference on Circular Dichroism, Colorado State University, Fort Collins, CO, Colorado State University, Fort Collins, CO, 1993; pp 152 - 157.

34. Applequist, J., Cavity model for optical properties of solutions of chiral molecules. *J. Phys Chem.* **1990,** *94* (17), 6564-6573.

35. (a) Applequist, J.; Sundberg, K. R.; Olson, M. L.; Weiss, L. C., A normal mode treatment of optical properties of a classical coupled dipole oscillator system with Lorentzian band shapes. *J. Chem. Phys.* **1979,** *70*, 1240; (b) Applequist, J.; Sundberg, K. R.; Olson, M. L.; Weiss, L. C., Erratum: A normal mode treatment of optical properties of a classical coupled dipole oscillator system with Lorentzian band shapes. *J. Chem. Phys.* **1979,** *71*, 2330.

36. Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z., Triclinic lysozyme at 0.65 A resolution. *Acta Crystallogr., Sect. D* **2007,** *63* (12), 1254-1268.

37. Yi, J.; Orville, A. M.; Skinner, J. M.; Skinner, M. J.; Richter-Addo, G. B., Synchrotron X-ray-Induced Photoreduction of Ferric Myoglobin Nitrite Crystals Gives the Ferrous Derivative with Retention of the O-Bonded Nitrite Ligand. *Biochemistry 49* (29), 5969-5971.

38. Arcovito, A.; Benfatto, M.; Cianci, M.; Hasnain, S. S.; Nienhaus, K.; Nienhaus, G. U.; Savino, C.; Strange, R. W.; Vallone, B.; Della Longa, S., X-ray structure analysis of a metalloprotein with enhanced active-site resolution using in situ x-ray absorption near edge structure spectroscopy. *Proc. Natl. Acad. Sci.* **2007,** *104* (15), 6211-6216.

39. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)10]3. *Protein sci.* **2002,** *11* (2), 262-270.

40. Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T., Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins: Structure, Function, and Bioinformatics* **1988,** *4* (1), 31-47.

41. Pappalardo, M.; Milardi, D.; Grasso, D. M.; La Rosa, C., Free energy perturbation and molecular dynamics calculations of copper binding to azurin. *J. Comput. Chem.* **2003,** *24* (6), 779-785.

42. Angelis, P. L. D.; Toraldo, G., On the Identification Property of a Projected Gradient Method. *SIAM J. Numer. Anal.* **1993,** *30* (5), 1483-1497.

43. Charalambous, C.; Conn, A., An efficient method to solve the minimax problem directly. *SIAM J. Numer. Anal.* **1978**, 162-187.

44. (a) Northrup, S. H.; Thomasson, K. A.; Miller, C. M.; Barker, P. D.; Eltis, L. D.; Guillemette, J. G.; Mauk, A. G.; Inglis, S. C., Effects of charged amino acid mutations on the bimolecular kinetics of reduction of yeast iso-1-ferricytochrome c by bovine ferrocytochrome b5. *Biochemistry* **1993,** *32* (26), 6613-6623; (b) Castro, G.; Boswell, C. A.; Northrup, S. H., Dynamics of protein-protein docking: cytochrome c and cytochrome c peroxidase revisited. *J. Biomol. Struct. Dyn.* **1998,** *16* (2), 413.

45. (a) Applequist, J., A full polarizability treatment of the π–π^* absorption and circular dichroic spectra of alpha-helical polypeptides. *J. Chem. Phys.* **1979,** *71* (11), 4332-4338; (b) Applequist, J., Erratum: A full polarizability treatment of the π–π * absorption and circular dichroic spectra of alpha-helical polypeptides [J. Chem. Phys. [bold 7][bold 1], 4332 (1979)]. *J. Chem. Phys.* **1980,** *73* (7), 3521-3521.

46. Bode, K. A.; Applequist, J., Globular protein ultraviolet circular dichroic spectra. Calculation from crystal structures via the dipole interaction model. *J. Am. Chem. Soc.* **1998,** *120* (42), 10938-10946.

47. Mestecky, J. L., M.E.; Strober, W.; Bienenstock, J.; McGhee, J.R.; Mayer, L., *Mucosal Immunol.* 3rd ed. ed.; Elsevier Academic Press: MA, 2005.

48. Laschtschenko, P., Uber die keimtötende und entwicklungshemmende Wirkung von Hühnereiweiß. *Med. Microbiol. Immun.* **1909,** *64* (1), 419-427.

49. Nester, E. W. A., D.G.; Roberts, C.E.; Nester, M.T., *Microbiology: A human perspective*. McGraw Hill: 2007.

50. Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R., Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2Å Resolution. *Nature* **1965,** *206* (4986), 757-761.

51. Vernon, C. A., The Mechanisms of Hydrolysis of Glycosides and Their Revelance to Enzyme-Catalysed Reactions. *Proc. R. Soc. Lond., Ser. B Biol. Sci.* **1967,** *167* (1009), 389-401.

52. Forlemu, N. Y. S., B.A.; Pothuganti, S.; Nori, R.; Mbote, Y.E.B; Thomasson, K.A., Introducing DInaMo: A package for calculating protein circular dichroism using classical electromagnetic theory. *Int. J. Mol. Sci.* **2011**, Manuscript in preparation.

53. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227-249.

54. Pellicane, G., Colloidal Model of Lysozyme Aqueous Solutions: A Computer Simulation and Theoretical Study. *J. Phys. Chem. B 116* (7), 2114-2120.

55. Kiefhaber, T., Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci.* **1995,** *92* (20), 9029-9033.

56. Deep, S.; Ahluwalia, J. C., Theoretical studies on solvation contribution to the thermodynamic stability of mutants of lysozyme T4. *Protein Eng.* **2003,** *16* (6), 415-422.

57. Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z., Triclinic lysozyme at 0.65 A resolution. *Acta Crystallogr. Sect. D* **2007,** *63* (12), 1254-1268.

58. Kurinov, I. V.; Harrison, R. W., The influence of temperature on lysozyme crystals. Structure and dynamics of protein and water. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995,** *51* (1), 98-109.

59. Herzberg, O.; Sussman, J. L., Protein model building by the use of a constrained–restrained least-squares procedure. *J. Appl. Crystallogr.* **1983,** *16* (1), 144-150.

60. Sreerama, N.; Woody, R. W., Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* **2004,** *383*, 318-351.

61. Ordway, G. A.; Garry, D. J., Myoglobin: an essential hemoprotein in striated muscle. *J. Exp. Biol.* **2004,** *207* (20), 3441-3446.

62. Kendrew, J., The Three-Dimensional Structure of Proteins. *Sci. Amer.* **1961,** *205*.

63. Hendgen-Cotta, U. B.; Kelm, M.; Rassaf, T., A highlight of myoglobin diversity: the nitrite reductase activity during myocardial ischemia-reperfusion. *Nitric Oxide 22* (2), 75-82.

64. Holzwarth, G.; Doty, P., The Ultraviolet Circular Dichroism of Polypeptides1. *J. Am. Chem. Soc.* **1965,** *87* (2), 218-228.

65. Orry, A.; Janes, R. W.; Sarra, R.; Hanlon, M.; Wallace, B., Synchrotron radiation circular dichroism spectroscopy: vacuum ultraviolet irradiation does not damage protein integrity. *J. Synchrotron Radiat.* **2001,** *8* (3), 1027-1029.

66. Hirst, J. D.; Colella, K.; Gilbert, A. T. B., Electronic circular dichroism of proteins from first-principles calculations. *J. Phys. Chem. B* **2003,** *107* (42), 11813-11819.

67. Yang, F.; Phillips Jr, G. N., Crystal structures of CO-, deoxy-and met-myoglobins at various pH values. *J. Mol. Biol.* **1996,** *256* (4), 762-774.

68. Di Lullo, G. A.; Sweeney, S. M.; Körkkö, J.; Ala-Kokko, L.; San Antonio, J. D., Mapping the Ligand-binding Sites and Disease-associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. *J. Biol. Chem.* **2002,** *277* (6), 4223-4231.

69. Clark, G. L.; Parker, E. A.; Schaad, J. A.; Warren, W. J., New Measurements of Previously Unkwown Large Interplanar Spacings in Natrural Materials. *J. Am. Chem. Soc.* **1935,** *57* (8), 1509-1509.

70. (a) Hulmes, D. J. S.; Miller, A., Quasi-hexagonal molecular packing in collagen fibrils. *Nature* **1979,** *282* (5741), 878-880; (b) Jesior, J. C.; Miller, A.; Berthet-Colominas, C., Crystalline three-dimensional packing is a general characteristic of type I collagen fibrils. *FEBS Letters* **1980,** *113* (2), 238-240.

71. Fraser, R. D. B.; MacRae, T. P.; Miller, A., Molecular packing in type I collagen fibrils. *J. Mol. Biol.* **1987,** *193* (1), 115-125.

72. Szpak, P., Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis. *J. Archaeol. Sci. 38* (12), 3358-3372.

73. Jenness, D. D.; Sprecher, C.; Johnson, W. C., Circular dichroism of collagen, gelatin, and poly(proline) II in the vacuum ultraviolet. *Biopolymers* **1976,** *15* (3), 513-521.

74. (a) Woody, R. W., Circular Dichroism Spectrum of Peptides in the Poly(Pro)II Conformation. *J. Am. Chem. Soc.* **2009,** *131* (23), 8234-8245; (b) Liu, Z.; Chen, K.; Ng, A.; Shi, Z.; Woody, R. W.; Kallenbach, N. R., Solvent Dependence of PII Conformation in Model Alanine Peptides. *J. Am. Chem. Soc.* **2004,** *126* (46), 15141-15150.

75. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* **2005,** *1751* (2), 119-139.

76. Thomasson, K. A.; Applequist, J., Effects of proline ring conformation on theoretical π-π* absorption and CD spectra of helical poly(L-proline) forms I and II. *Biopolymers* **1991,** *31* (5), 529-535.

77. Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T., How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **1995,** *4* (11), 2411-2423.

78. Protoparam Physicochemical Parameters of A Protein Sequence. http://www.expasy.ch/tools/protparam.html.

79. Naseem, F.; Khan, R. H., Characterization of a common intermediate of pea lectin in the folding pathway induced by TFE and HFIP. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2005,** *1723* (1–3), 192-200.