

ANALYTICAL APPLICATIONS MEASURING  
THE CHIRALITY OF PROTEINS AND  
PEPTIDES USING CIRCULAR  
DICHROISM

DENNIS WILLIAM PROVINCE

Bachelor of Arts

University of Colorado

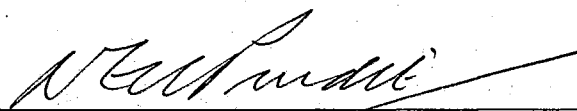
Boulder, Colorado

1993

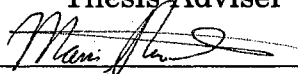
Submitted the Faculty of the  
Graduate College of the  
Oklahoma State University  
in partial fulfillment of  
the requirement for  
the Degree of  
DOCTOR OF PHILOSOPHY  
May 1999


ANALYTICAL APPLICATIONS MEASURING  
THE CHIRALITY OF PROTEINS AND  
PEPTIDES USING CIRCULAR  
DICHROISM

Thesis Approved:

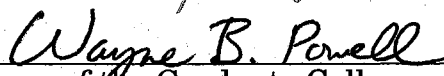


Thesis Adviser









Dean of the Graduate College

## PREFACE

Analytical analysis of chiral molecules becomes more important because of the importance of biomolecules. Circular Dichroism (CD) is a powerful tool to study these molecules. Most scientists would agree that the three most important classes of biomolecules are the proteins, the nucleic acids, and the carbohydrates. All of these molecules exhibit chirality in their most basic building block form. Proteins, peptides, and oligopeptides were the focus of this study. But because of the prolific nature of these molecules and the millions upon millions of possible combinations of protein sequence the work here is only a small offering in a field that is essentially untapped.

The credit for any work presented here should be given to my advisor Dr. Purdie. For by he alone has any of this come to pass. The vision of the potential methodology and their possible applications are his solely.

The inspiration to apply to graduate school came from Dr. Sievers and Dr. Barkley. These mentors from the University of Colorado allowed me a glimpse at the life of a graduate student and instilled me with the confidence to succeed. Dr. Siever's research group provided me a peer group to learn from in an encouraging and positive way.

A large amount of this work would not have been possible if not for generous contributions in the form of work material. Chiron now Torrey Pines Institute for Molecular Studies provided a number of small

oligopeptides as well as the neural peptides that were used in this study. Thanks mainly to the effort of Sylvie Blondelle. A large amount of gratitude must be given to the FDA and especially to the Division of Testing and Applied Analytical Development for supplying the insulin samples that were used in this study. Many thanks to Moheb Nasr and Tom Layloff for those samples and for their technical assistance.

On a personal note, I begin by thanking my Father in Heaven for blessing me with the talents, opportunities, and family support that I have. Also I would like to thank my family for their encouragement of my academic pursuits both financially and emotionally. My wife has provided me with the energy and the confidence to finish my work here by her loving spirit and belief in my abilities. To her I dedicate this work.

## TABLE OF CONTENTS

Chapter	Page
I. CIRCULAR DICHROISM, ITS USES IN ANALYSIS. . . . .	1
II. BACKGROUND	
Historical. . . . .	6
Theoretical. . . . .	10
III. INSTRUMENTATION. . . . .	17
IV. ANALYSIS OF PROTEIN . . . . .	22
Importance of Chirality in Industry. . . . .	22
Chromatographic Methods Used to Analyze Proteins . . . . .	24
V. ANALYZING STRUCTURALLY SIMILAR PEPTIDES . . . . .	28
Introduction . . . . .	28
Experimental . . . . .	29
Results and Discussion . . . . .	32
Summary . . . . .	41
VI. CHIRAL PROPERTIES OF INSULINS . . . . .	43
Introduction . . . . .	43
Experimental . . . . .	45
Results and Discussion. . . . .	46
Summary . . . . .	59

VII. QUANTITATIVE VALIDATION OF THE CHIRAL PROPERTIES OF PEPTIDES .....	60
Introduction .....	60
Experimental .....	63
Results and Discussion.....	65
Summary .....	82
VIII. TRIPEPTIDE DISCRIMINATION USING CIRCULAR DICHROISM DETECTION.....	84
Introduction .....	84
Experimental .....	86
Results and Discussion.....	87
Summary .....	99
LITERATURE CITED .....	101
APPENDIX A .....	108

## LIST OF TABLES

Table	Page
5.1 Primary sequence of the synthetically engineered analogs . . . . .	30
5.2 Some naturally occurring Dynorphins . . . . .	31
5.3 Primary sequence of the six residue stereoisomers. . . . .	31
5.4 Data reduction by Principal Component Analysis of a Spinning Plot® presentation of CD data for the host and mixed copper complexes of DSLET and DTLET . . . . .	34
5.5 PCA factors selected to characterize the Dynorphins . . . . .	38
5.6 PCA factors selected to characterize the hexapeptides . . . . .	40
6.1 Data reduction by Principal Component Analysis of a Spinning Plot® presentation of CD data for the host and mixed copper complexes of human insulin . . . . .	54
6.2 Comparison of mean values for PC2 eigenvectors for D-Histidine/Insulin complexes among insulin types, manufacturers, and lot numbers. . . . .	56
6.3 Bovine chain A and chain B PC2 values for separate mixtures, equimolar mixtures, and the spectral average of the two . . . . .	59
7.1 Ratios and concentrations of the dipeptides used in the chemical purity study where GA was the host and the other nine peptides the impurities . . . . .	65
7.2 Determination of enantiomeric purities for prepared binary mixtures of GA and G(D)A . . . . .	77
7.3 Principal component values calculated for the GA versus AG system. . . . .	80
7.4 Corresponding values for GA plotted against spectral data for the dipeptides in 8.0mM solution . . . . .	80

8.1	Determination of enantiomeric purities for prepared binary mixtures of GHG/GhG, LGG/IGG, and YGG/yGG . . . . .	93
8.2	Principal components calculated for the GGA vs GGH system . . . . .	97
8.3	Comparative PC22 values for all the L-tripeptides against GGA . . . . .	97



## LIST OF FIGURES

Figures	Page
2.1 Polarization of light by reflection from glass . . . . .	7
2.2 Examples of rotatory dispersion . . . . .	10
2.3 Electronic component of plane polarized light . . . . .	11
2.4 Vector representation of right- and left- circularly polarized light . . . . .	12
2.5 Effect of plane polarized light . . . . .	13
2.6 Ellipse created when the components of plane polarized light are unequally absorbed . . . . .	14
3.1 Optical system of the J-500A recording spectrophotometer . . . . .	17
3.2 Signal intensity received at the photomultiplier tube . . . . .	18
4.1 R,Z isomer of the Japanese beetle pheromone . . . . .	23
5.1 CD data for the + ephedrine, DTLET, DSLET, DADLE, and DAGO complexes . . . . .	32
5.2 3-D Spin Plots® for the experimental neural peptides DSLET and DTLET. . . . .	33
5.3 CD spectral data for the - ephedrine stock, Dyn A (1-9), Dyn A (1-11), Dyn A (1-13), Dyn A (1-13)NH <sub>2</sub> . . . . .	35
5.4 CD spectra plot of the stereoisomers of HO-RFMWMR-NH <sub>2</sub> . . . . .	36
5.5 Spectral CD data for the - ephedrine, DTLET, DSLET, DADLE, and DAGO complexes . . . . .	37
5.6 CD spectra for + ephedrine, DTLET, DSLET, DADLE, and DAGO at 0.94mM concentration of each. . . . .	39
6.1 Human insulin . . . . .	43

6.2	Visible range CD spectra for the Cu(II) complex with D-Histidine, and the mixed Cu(II) complexes of D-Histidine with bovine insulin, human insulin, porcine insulin, and human lispro insulin . . . . .	48
6.3	Correlation of the full spectral data for the mean of all Lilly human insulin mixed complex spectra (x-axis) against the mean of all Novo human mixed complex spectra (y-axis) . . . . .	50
6.4	Correlation of the full spectral data for the mean of all Lilly human insulin mixed complex spectra (x-axis) against the means for the corresponding data for porcine, human lispro, and bovine insulins . . . . .	51
6.5	Spinning Plots® for the spectral data for human lispro and bovine insulins . . . . .	53
6.6	CD spectral data for mixed complexes of D-Histidine with intact bovine insulin, bovine insulin chain A, bovine insulin chain B, an equimolar mixture of chain A and chain B, and the average value of the sum of the spectra for chain A and chain B . . . . .	58
7.1	Covalent structures of the amino acid monomers used . . . . .	63
7.2	Visible CD spectra for the Cu(II) complexes of the nine chiral dipeptides . . . . .	67
7.3	The chiral ligands can effect the ground state, the excited state of both . . . . .	68
7.4	Bonding scheme for the dipeptide glycylglycine (GG) to the copper ion . . . . .	69
7.5	Dipeptide AA added to Cu <sup>2+</sup> at different concentrations. . . . .	70
7.6	CD spectra of the GA complexed with AG added as an impurity at the levels of 1, 3, 5, and 10%. . . . .	71
7.7	Plot of ellipticity vs %AG added to GA as an impurity. . . . .	72
7.8	Correlation plots of ellipticity for the Cu(II)-GA complex versus the ellipticities for the analogous complexes with G(D)A, GG, AG, GA, and AA . . . . .	73
7.9	Correlation plots of ellipticity for the Cu(II)-GA complex versus the ellipticities for the analogous complexes with YG, YA, GA, GY, YY, and AY. . . . .	74

7.10	CD spectra of the GA with G(D)A added as an impurity. . . . .	75
7.11	Correlation plots of ellipticity for the Cu(II)-GA complex versus the ellipticities for the enantiomeric mixtures of percent composition . . . . .	76
7.12	Spin Plots® for the presentation of wavelength (x-coordinate), spectral data for the GA complex (y-coordinate), and spectral data for the AG and the GA complexes . . . . .	79
7.13	Linear correlation plots of PC2 values versus the percent “chemical impurities” for all nine dipeptides added to GA. . . . .	82
8.1	Bonding scheme of the tripeptide glycylglycylglycine (GGG) to the copper ion . . . . .	84
8.2	Visible CD spectra for the Cu(II) complexes of GGA, GGH, GGI, GGL, GGP, GHG, LGG, and YGG . . . . .	88
8.3	Correlation plots of ellipticity for the Cu(II)-GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of GGA, GGH, GGI, GGL, and GGP . . . . .	90
8.4	Correlation plots of ellipticity for the Cu(II)-GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of GGA, GHG, LGG, and YGG. . . . .	91
8.5	Correlation plots of ellipticity for the Cu(II)-GGA complex versus the ellipticities for the 5% chemical mixtures with GGH, GGI, GGL, and GGP. . . . .	95
8.6	Spinning Plots® for the presentation of wavelength (x-coordinate), spectral data for the GGA complex (y-coordinate), and spectral data for the GGA, GGH, GHG, and LGG complexes . . . . .	96
8.7	Plots of the percent chemical impurity versus the eigenfactor PC22 for the Cu(II) complexes of GGH GGP, LGG, and YGG . . . . .	99
A.1	Graph of two variables where P1 and P2 are principal components of the data . . . . .	109

## CHAPTER ONE

### CIRCULAR DICHROISM, USES IN ANALYSIS

The field of Circular Dichroism (CD) includes studies of the interaction of circularly polarized light with optically active material. From the beginning experiments have focused on materials that are naturally CD active and scientists have speculated on how ellipticities and molecular structure can be related. CD has been used to extract information about chiral organic molecules (1, 2, 4-8) as well as to determine relative and even absolute conformation (19-21, 33-35). These studies are limited and tend to provide insight only into very select groups of molecules. The power of CD is not in its ability to determine absolute configuration but in its sensitivity to changes in and around the chiral centers and those electrons associated with the chiral centers.

As the light sources and the instrumentation used in making CD measurements improved smaller concentrations could be used to achieve measurable signals. The availability of chromatography (26, 28) allowed interfering components to be separated out before analysis. Because the requirements of CD are twofold, many analyses have been performed on mixtures without chromatographic separation (1-3, 24). To induce CD activity, chirality and/or chromophores have been added to the analyte(40-44). Chiral molecules possess enantiomers that have all the same physical characteristics of the analyte, with the exclusion of the rotation of plane polarized light, making determination of enantiomeric purity difficult without the use of CD(36, 70, 73-74).

Chiral molecules such as carbohydrates, nucleic acids, and especially proteins have become very popular analytes of late because of their biological significance. These biological molecules are being used as models and precursors to pharmaceutical drugs. The ability to inhibit, change, or enhance cell function specifically by administering a designed or modeled pharmaceutical is the goal of medicine. The designing or regulating of drug molecules requires a very specific structural composition which includes the specific chirality. Analytical techniques that can differentiate between very similar biological molecules including enantiomers are needed to simplify the task of ensuring structural integrity. The sensitivity of the CD signal to the chirality of these molecules makes it an excellent candidate for providing a method of analysis in the elucidation of these molecules.

New opportunities are emerging that can benefit from the application of CD detection. These opportunities are related, at least in part, to some of the following. An aging population in this country has an increased interest in accessing drugs that target elderly needs. The pharmaceutical industry is maturing in that there are fewer companies now than in the past and these companies are able to bring a larger number of products through the screening process than in the past (50). There has been a boom in the biotechnology industry related to an expanding pharmaceutical marketplace as well as an increase in technological advances (32, 54). The awareness of proper nutrition and the need for a balance of vitamins has led to an explosion in the dietary

supplement industry. The Food and Drug Administration is moving towards good manufacturing practices for the dietary supplements industry (13), which is at a higher level than that industry must now maintain. There has been a recent increase of self regulation among the supplement industry to standardize methods of analysis where there previously were none (51). Increasingly competitive markets demand that companies maximize their efficiencies by increasing the number of internal standards that include in shop quality control and batch monitoring (32).

This general rise in (i) the demand for pharmaceuticals and (ii) the increase in the regulation of these same molecules and their manufacturing, has led to an increased need for new regulation of analytical procedures. CD can be specifically suited to provide quality control (QC) information for protein based molecules, and may have the potential to provide information for other types of drug substances as well. Analysis of a manufactured drug molecule demands that the methods employed be quick, accurate, simple, and inexpensive. A good number of pharmaceutical molecules of interest now are analogs of biomolecules and inherently chiral (113). For many of the popular analytical methods used in the past the objective was to determine molecules present within a complex matrix. For the QC analysis of molecules that are synthesized in the processes of large scale manufacturing, the analytes are in their purest attainable form. No separations are needed and the quantities can be, for analysis sake, unlimited. Any analytical method would seem to be suitable

in an environment like this.

The landscape of pharmaceuticals has changed. Combinatorial chemistry has led to the increased use of proteins and peptides as drug molecules (22). These sequences are sometimes conceived of from huge libraries of possible sequence combinations. They may be selected by a process that utilizes the target molecule as the stationary phase in a High Performance Liquid Chromatography (HPLC) column. Those sequences that are retained the longest are the object of studies to determine if their interactions with the target changes the biological function (22). Some peptide drugs are comprised of both D- and L- stereocenters and can come in a variety of lengths and sequences.

Methods that are able to sequence even complex enzymes have been available for a number of years (25). Many molecules of interest are man made analogs of enzymes that are already being manufactured. Some of the most important work in terms of improving human life has been in the manufacture of those proteins for human subjects that are unable to produce their own. For these enzymes to work as substitutes for the natural forms, they must be just as specific. Even the absence of just one amino acid in a fifty amino acid protein can drastically change the overall function of the molecule (93). The binding properties of these synthetic molecules must be able to regulate, inhibit, or promote cell function in a totally specific pathway or they may do more harm than good.

With these problems in mind a CD method has been developed that can provide answers to questions that occur when QC is the issue. The

method has shown a sensitivity to changes in concentration. With a new way of treating the data, the presence of chiral as well as achiral types of impurities can be detected. With each change in the stereocenters and/or the sequence of the amino acids, there are dramatic changes in the overall CD spectra for peptides with up to fifty residues. For instance the sensitivity of the spectral differences for a series of insulin molecules with only one sequence inversion was found to be sufficient to differentiate among three out of four analogs.

Literally hundreds of molecules are potential analytes for this method providing manufacturing, R & D, and regulating agencies with QC information on polypeptide drug forms much faster and easier.



## CHAPTER TWO

### I. HISTORICAL BACKGROUND

The first recorded CD experiment on a solution was described by Cotton in 1896 (5). However the science behind this technique goes back even farther. Most accounts agree that Christian Huygens was the first on record for noting the phenomenon that is termed polarization of light. In 1690 Huygens found that when objects were viewed through calcite crystals (calcium carbonate in hexagonal form) double images were observed. These crystal plates, sometimes called Iceland spars, separated the incoming light into two beams which were assigned the names ordinary ray and the extraordinary ray (5).

The real beginning of scientific inquiry into this phenomenon began in 1816 when E.T. Malus, through his investigations, concluded that “the light acquires in these circumstances properties which are independent of its inclination to the surface...but are unique relative to the sides of the vertical ray...I will describe as polarisation the modification which gives to the light its properties” (5). Malus, when he wrote this, had been observing the effects of reflecting sunlight off an untinned mirror on to another mirror that was parallel to the first. He found that if the second mirror was rotated in the plane by  $90^\circ$  relative to the first that “it will not reflect a single molecule of light” (5) (Figure 2.1). These observations led Fresnel to conclude that vibrations of light were transverse and not longitudinal. Fresnel stated in his *Mémoire sur la double Réfraction* that “the vibrations of a polarized beam must be perpendicular to what is called its plane of

polarization (5).”

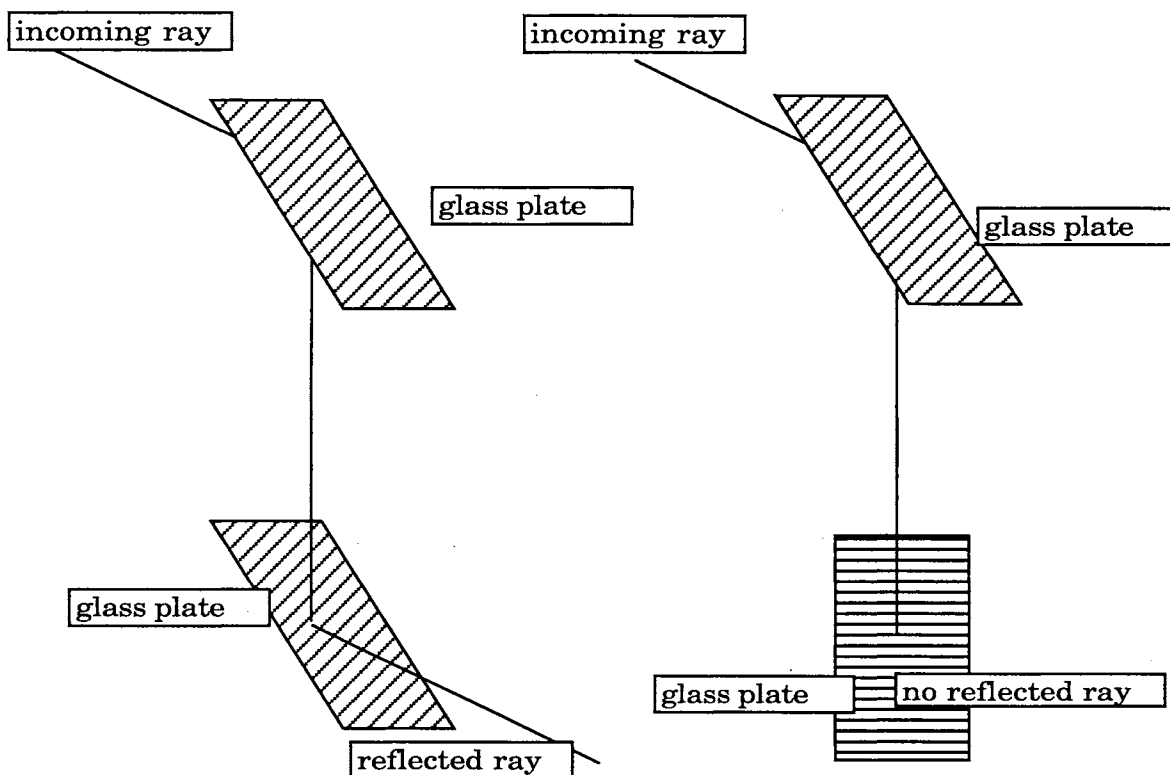


Figure 2.1: Polarization of light by reflection from glass (adapted from reference 5)

Biot (1815) demonstrated that optical activity was not limited to crystalline materials, but was also a property of solutions of certain natural products. The measurement of the rotation of polarized light at a specific wavelength by particular solutions was the beginning of the science of polarimetry.

Once started it was not long before Biot and Fresnel independently observed that as the wavelength of light incident upon the sample decreases, the angle of rotation in the plane of plane-polarized light

increases, which was the start of what is now known as optical rotatory dispersion (ORD). Polarimetry and ORD are similar in that in both methods the magnitude of the rotation of the plane-polarized light is measured. ORD is the continuous spectral analysis of this phenomenon while polarimetry is accomplished using measurements made at one or a few wavelengths.

Haidinger in 1847 observed that there was unequal absorption of circularly polarized light when it was transmitted through a crystal of amethyst quartz. But it wasn't until 1896 that Aime Cotton first observed the same effect when light from a sodium flame was passed through a solution of potassium chromium tartrate. He concluded, like Haidinger did, that the phenomenon of CD was in fact due to the unequal absorption of the left and right circularly polarized components of plane-polarized light.

In 1848 Louis Pasteur discovered a way to physically separate the optically active components of a racemic solution. Crystals of sodium ammonium tartrate that were slowly separated from a slightly supersaturated solution resulted in crystal patterns of two distinct shapes that were themselves non-superimposable mirror images and easily distinguishable under a microscope. He redissolved each type of crystal, in equal molar amounts, to form two separate solutions. Polarimetry measurements showed equal but opposite angles of rotation for the two solutions.

From this work came the concept of the tetrahedral carbon atom. That asymmetry was due to the bonding configuration was furthered by

Van't Hoff and Le Bel. The tetrahedral arrangement of saturated carbon atoms is currently the most common basis for three dimensional molecular shapes. The tetrahedral carbon atom became the focal point of the study of chirality. It was now possible to begin the process of determining the interatomic linkages that made up molecules. Though others have suggested that an asymmetric carbon atom would be a requirement of optical activity, Pasteur's original concept that any molecule whose mirror image is not superimposable is still considered to be the only requirement.

## II. THEORETICAL BACKGROUND

The ability of a molecule to rotate plane polarized light in many cases is a result of the presence of a carbon atom which has four unequal substituents attached to it. This led to the realization that two-dimensional representations of chiral molecules are insufficient to account for these optical properties. This conclusion led to the study of the positions in space of each atom in a molecule, i.e. molecular stereochemistry. Though rotatory power initiated the study of the absolute configuration of molecules, that method of analysis cannot be used for configuration determination.

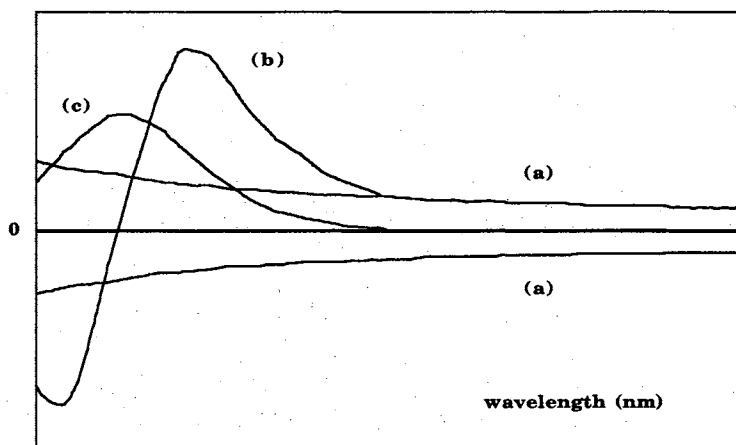


Figure 2.2: Examples of rotatory dispersion: a) normal ORD curves, b) anomalous ORD curve or Cotton effect associated with the absorption band of the chromophore, c) absorption spectra for the material in b).

Rotatory power, as pointed out by Biot, varies with wavelength. When the incident light is far from the absorbing region, the usefulness of this technique can become quite limited. Rotatory dispersion, ORD, was a more interesting property of a substance than a simple measurement at a given wavelength. Cotton ultimately introduced the study of ORD to wavelengths that lie within the absorbing region of a chromophore.

The resulting anomalous curve has been called a Cotton effect. Ordinary dispersion produces a monotonic change in the angle of rotation that increases as the wavelength decreases. Anomalous dispersion occurs over the wavelength range of an absorption band whenever a chromophore is present in an asymmetric (chiral) molecule (Figure 2.2b).

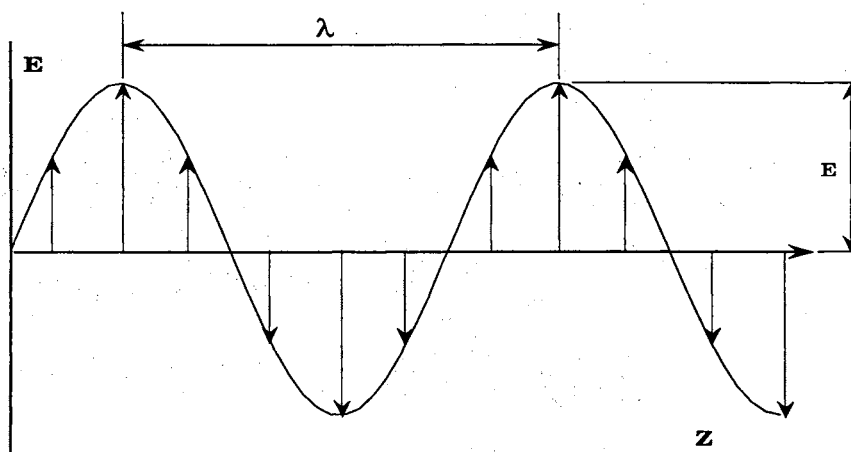


Figure 2.3: Electronic component of plane polarized light: each arrow represents the direction of the transverse wave. The z axis is the direction of propagation,  $\lambda$  is the wavelength, and  $E$  is the amplitude.

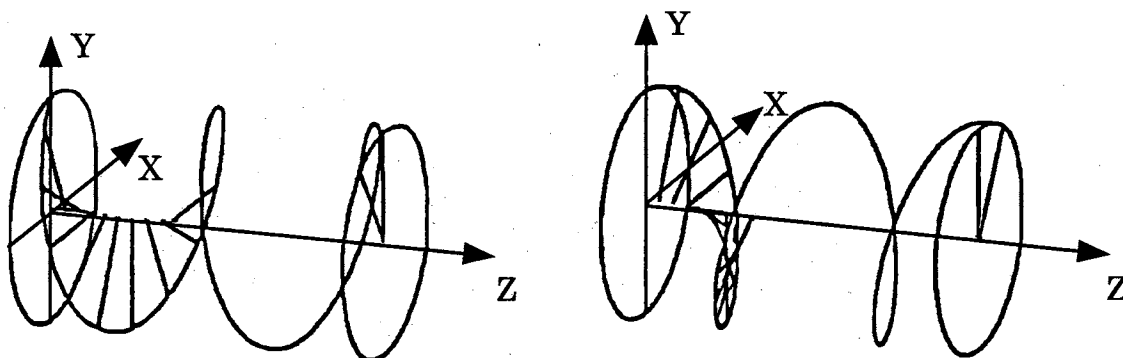


Figure 2.4: Vector representation of right- and left- circularly polarized light. Z in the axis of propagation.

This anomalous behavior could only be explained after Fresnel applied the new transverse wave theory of light to the phenomenon of optical activity. He described the physical connection between linearly polarized light and circularly polarized light. In the former the electric vector lies along a single axis as it propagates through space (Figure 2.3) and in the latter the electric vector rotates around the axis of propagation (Figure 2.4). The connection as it was proved by Fresnel is that linearly polarized light is actually the resultant of two coherent circular components of opposite sense of rotation. With linear or plane polarized light thus defined optical activity is determined by the relative indices of refraction for left- and right- circularly polarized light through the birefringent material.

Since the frequency of a light ray is independent of the index of refraction the resulting difference observed through the medium corresponds to a decrease in the speed of one component relative to the

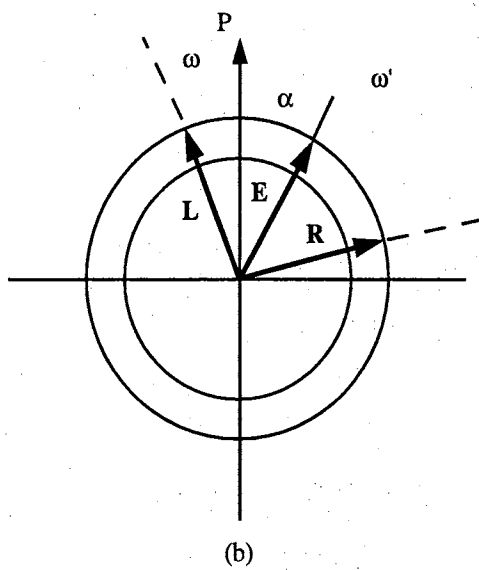
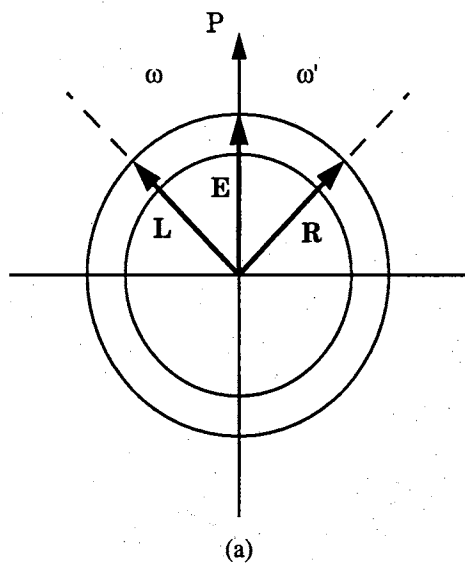


Figure 2.5: Effect of Plane Polarized Light on a) optically inactive material and b) optically active material where  $n_L > n_R$ .

other. If  $n_L$  is the index of refraction relative to the left- circularly polarized component and  $n_R$  represents the index of refraction for the right- circularly polarized component, then it can be concluded that  $n_L - n_R$  cannot be equal to zero for materials that exhibit optical rotation (Figure 2.5). If  $k$  is defined to be the absorption coefficient of a substance then  $k_L$  and  $k_R$  are the absorption coefficients of a substance relative to the left- and right- polarized components. When the incident light is under an absorption band and  $n_L - n_R \neq 0$  then it can be derived that  $k_L - k_R \neq 0$ . The quantity  $k_L - k_R$  is defined as the circular dichroism of the material,

or the differential absorption between the left- and right- circularly polarized light.

The differential absorption caused by the

circular birefringent material changes the lengths of the vectors associated

with the left- and right- components of the beam. If  $E_L$  and  $E_R$  are the



lengths of the left and right circularly polarized components respectively,

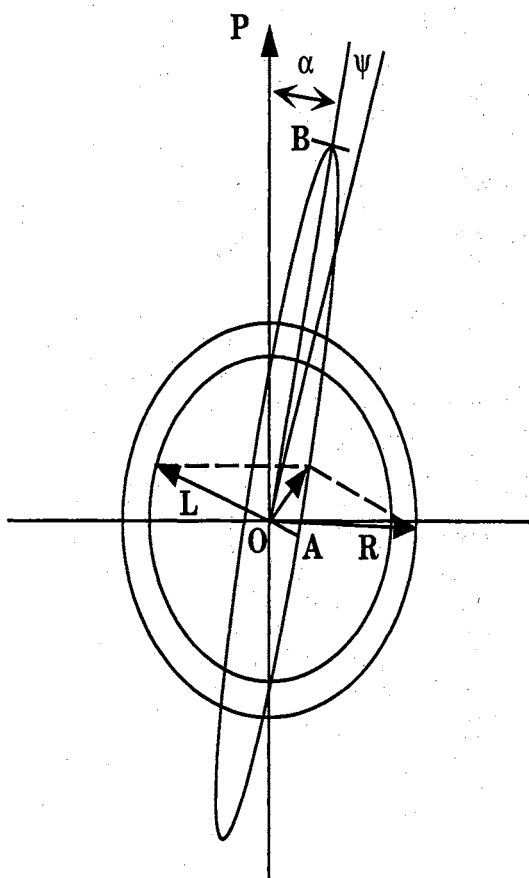


Figure 2.6: Ellipse created when the components of plane polarized light are unequally absorbed. **L** and **R** ( $E_L$  and  $E_R$ ) are the intensity of the left and right components.

then  $E_L + E_R$  is now out of the path of the

circle and the resulting vector traces out an ellipse. The resulting light is said to be elliptically polarized and in quantum mechanical terms the transition probabilities are different for left- and right- circularly polarized light (8).

From Figure 2.6 it can be seen that the ratio of the minor axis **A** and the major axis **B** of the elliptically polarized light is the measure of the tangent of the angle of ellipticity  $\Psi$ . The angle of rotation  $\alpha$  of the plane polarized light is the difference between the original (incident light)

plane of polarization and the major axis **B** of the resulting ellipse. If **E** is defined as the amplitude of the plane polarized

light then  $E_L$  and  $E_R$  would be the amplitude of the left- and right- circularly polarized components. The minor axis **A** then would be the difference between the two amplitudes and the major axis **B** would be the sum of the amplitudes from the two components. Then the tangent of the angle of ellipticity could be written this way.

$$\tan \psi = \frac{E_L - E_R}{E_L + E_R} \quad (1)$$

The amplitude (E) is related to the absorption coefficient  $\kappa$  by

$$E = E_0 e^{-2\pi\kappa/\lambda} \quad (2)$$

Since the angle of ellipticity ( $\Psi$ ) and the differences between  $\kappa_L$  and  $\kappa_R$  are small, the following relationship is found to be a good approximation

$$\Psi = \left[ \frac{\pi}{\lambda} \right] * [\kappa_L - \kappa_R] \quad (3)$$

Remembering that Fresnel had expressed the angle of the rotation as

$$\alpha = \left[ \frac{\pi}{\lambda} \right] [\eta_L - \eta_R] \quad (4)$$

then the angle of ellipticity can be expressed in terms of molar extinction coefficients by converting  $\kappa$  using

$$\kappa = 2.303\epsilon C \quad (5)$$

where  $\epsilon$  is the molar extinction coefficient and  $C$  is the concentration in moles per liter. The formalism adopted when describing the angle of rotation of a substance is to report the specific rotation  $[\alpha]$ , and this is defined to be

$$[\alpha] = \left[ \frac{\alpha}{c} \right] \left[ \frac{1800}{\pi} \right] \quad (6)$$

The units defined to  $\alpha$  are degrees per decimeter,  $c$  is the concentration expressed in the units of grams per milliliter of solution and the  $1800/\pi$  is the conversion factor giving the specific rotation in degrees per decimeter. The specific ellipticity would then be defined by

$$[\Psi] = \left[ \frac{\Psi}{l * c} \right] \quad (7)$$

where  $\Psi$  is measured in degrees,  $l$  is the path length in decimeters, and  $c$  is the concentration in grams per milliliter of solution. The molecular ellipticity  $[\Theta]$  is the most suitable experimental quantity for comparing different substances and it can be expressed

$$[\Theta] = 3300 * [\epsilon_L - \epsilon_R] = 3300 * \Delta\epsilon \quad (8)$$

where the ellipticity is dependent only upon the difference of the extinction coefficients. In this sense CD is no more than a modified form of absorbance spectrophotometry. All the properties of the latter apply to CD.

## CHAPTER THREE

### INSTRUMENTATION

#### Description of the instrument

The JASCO (Japan Spectroscopic Co. Ltd.) model J-500A was the instrument used to conduct all CD measurements (14).

This automatic recording spectrophotometer has a 450 watt xenon arc lamp as a light source. Dehydrated, spectra grade nitrogen protects the optical system from damage due to ozone formation. The spectrophotometer is

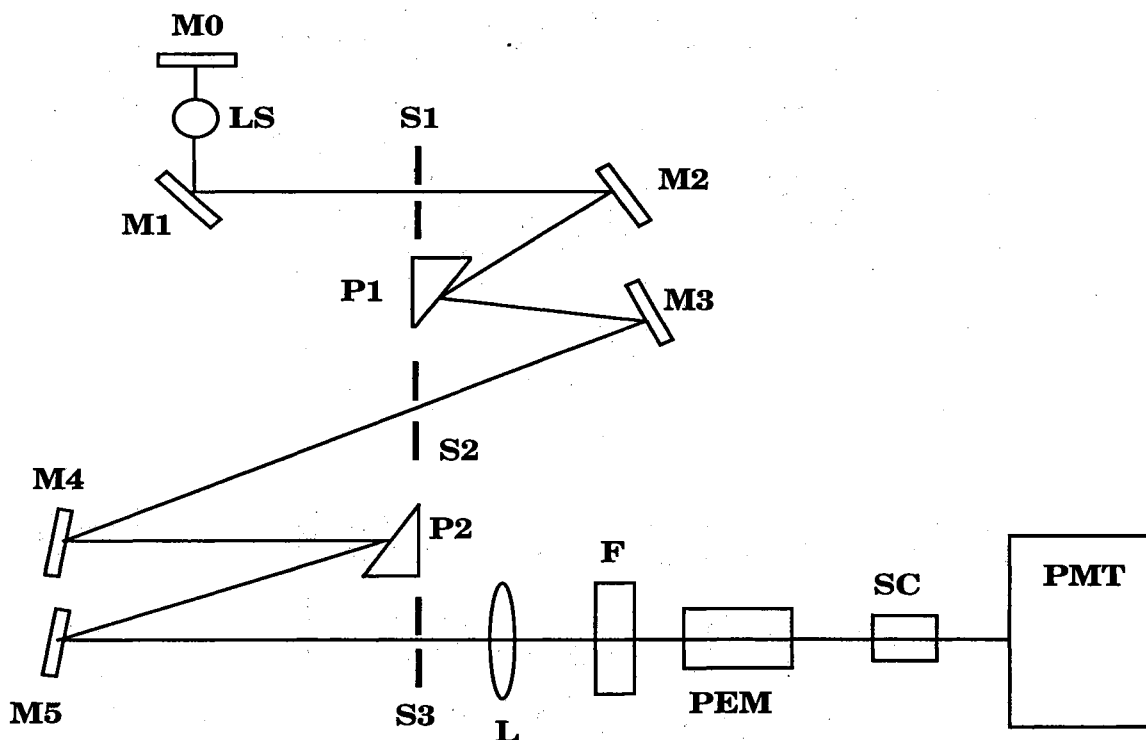


Figure 3.1. Optical system of the J-500A recording spectrophotometer

$M_0, M_1, M_2, M_3, M_4, M_5$ : spherical mirrors

LS: xenon arc lamp

$S_1, S_2, S_3$ : slits;  $P_1, P_2$ : prisms

L: lens; F: filter

PEM: piezo elastic modulator

SC: sample compartment

PMT: photomultiplier tube

controlled through a JASCO IF-500 II serial interface unit. This unit allows instrument parameters to be set and stored via a personal computer which can also save the spectral data as it is collected.

From the xenon arc lamp, see Figure 3.1, the light is focused onto the slit  $S_1$  by the spherical mirror  $M_1$ . The slit  $S_1$  is the entrance to the double monochromator. This optical set up reduces stray light, which is an interference in CD measurement. Each part of the monochromator is identical, from  $S_1$  to  $S_2$  and from  $S_2$  to  $S_3$ . The spectropolarimeter contains two quartz prisms  $P_1$  and  $P_2$ , which are placed orthogonal to each other. The monochromator is used not only for wavelength selection, but also

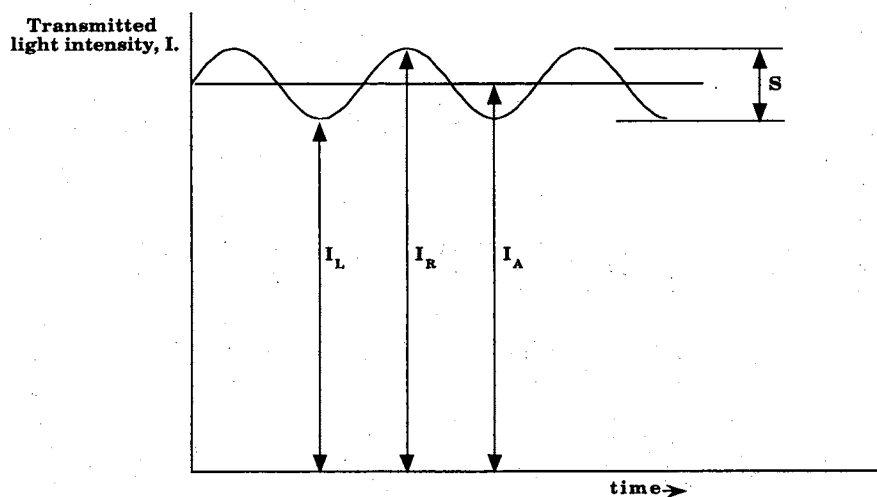


Figure 3.2: Signal intensity received at the PMT:  $I_L$  and  $I_R$  are the left- and right circularly polarized intensities,  $I_A$  is the average of the two signals, and  $S$  is the difference between  $I_L$  and  $I_R$ .

allows the beam to become linearly polarized in the horizontal direction. After leaving the double monochromator the light beam is refocused onto a filter to remove any stray unpolarized light. The resulting beam of linearly polarized light passes into a CD modulator. This modulator uses electrical pulses to produce left- and right- circularly polarized light by the piezoelectric effect. A sample is subjected to the beam intensity profile shown in Figure 3.2. The transmitted light falls onto a photomultiplier tube (PMT) that responds to the transmission by manipulating the voltage to maintain a constant current.

The piezoelastic modulator (PEM) cycles at a frequency of 50kHz. During each half cycle the PEM either left- or right- circularly polarizes the input beam. This modulation is carried on a DC signal that represents the average of the two circularly polarized beams. The birefringent material will experience difference absorption. This causes the intensity of the transmitted light to vary. The intensity fluctuations are transformed into voltage changes required to maintain a constant current from the PMT.

The maxima and minima in Figure 3.2 represent the intensity of the left- $I_L$  and right- $I_R$  circularly polarized light beam. From the theoretical treatment shown below these intensities correspond to the molecular ellipticities  $\epsilon_L$  and  $\epsilon_R$ . The CD signal is calculated at each wavelength directly by measuring the average signal intensity  $I_A$  and the peak difference between the left- $I_L$  and right- $I_R$  intensities.

From the theoretical treatment in chapter two it was mentioned that the molecular ellipticity  $[\Theta]$  can be expressed as

$$[\Theta] = \left[ \frac{4500}{\pi} \right] * [\epsilon_L - \epsilon_R] * \ln 10 \quad (9)$$

Where  $\epsilon_L$  and  $\epsilon_R$  are the extinction coefficients relative to the left- and right-circularly polarized light. The expression  $(\epsilon_L - \epsilon_R)$  or  $\Delta\epsilon$  can be written as

$$\Delta\epsilon = \frac{1}{L * C} \log_{10} \left[ \frac{I_R}{I_L} \right] \quad (10)$$

where L and C are the path length and the concentration.  $I_R$  and  $I_L$  are the intensities of the right- and left- polarized beams. These two equations can be combined to show that the molecular ellipticity is

$$[\Theta] = \left[ \frac{4500}{\pi * L * C} \right] * \ln 10 * \log_{10} \left[ \frac{I_R}{I_L} \right] \quad (11)$$

In practice, since the ratio of the intensities at every wavelength is nearly one the above equation is not a very accurate way of measuring the molecular ellipticity  $[\Theta]$ . It does show that a relationship exists between the beam intensities and the molecular ellipticity. From Figure 3.2 expressions for each beam intensity relative to the quantities  $S$  and  $I_A$  can be made

where

$$I_A = \frac{1}{2} [I_L + I_R] \quad (12)$$

and also

$$S = I_R - I_L \quad (13)$$

The molecular ellipticity can now be expressed as

$$[\Theta] = \left[ \frac{4500}{\pi * L * C} \right] * \ln 10 * \log_{10} \left[ \frac{1 + \frac{S}{2 * I_A}}{1 - \frac{S}{2 * I_A}} \right] \quad (14)$$

In this treatment it can be assumed that for most cases  $S/2 * I_A \ll 1$ . If  $E_A$  and  $E_S$  are defined to be the voltages associated with the average signal intensity and the difference of the signal intensity respectively then the molecular ellipticity can be expressed like this

$$[\Theta] = \left[ \frac{4500}{\pi * L * C} \right] * \ln 10 * \frac{E_S}{E_A} * \log_{10} e \quad (15)$$

These electrical quantities,  $E_A$  and  $E_S$ , are easily measured and are quite different from one another in magnitude. Jasco's J-500A calculates the molecular ellipticity by measuring these two quantities.

The Jasco J-500A is linked to a personal computer that is used to both control the instrument and save the data after collection. This set up has the flexibility to allow the operator to control all functions related to the collection of the data as well as many options related to the manipulation of the data once collected.



## CHAPTER FOUR

### I. IMPORTANCE OF CHIRALITY IN INDUSTRY

#### **Introduction**

The discovery of chiral molecules dates back to before Pasteur's time. The difficulty of separating enantiomers is well known and has been studied extensively. These studies have led to various conclusions including; the production of molecules in an achiral environment tends to produce racemic mixtures; the physiological response of the enantiomer of a molecule may be completely different and unpredictable compared with the response of the desired isomer; the presence of one enantiomer may actually inhibit the desired response of the other (27). The importance of producing optically pure compounds has not been lost. It only takes one example of negative feedback due to an enantiomeric impurity to impress upon chemists how critical the measure and analysis of enantiomers really is.

Besides the possibility of creating an enantiomer with side effects that negatively outweigh the benefits to society there are many other reasons why optically pure materials are desirable to produce. Along with the fact that biological activity goes along with one enantiomer, there may be a mixture of good and bad effects produced. Only the production of the optically pure material would allow the separation of these physiological effects. Racemic production produces as much waste as it does product. As found with the Japanese beetle pheromone the activity of the optically pure compound may be dramatically improved over the racemate.

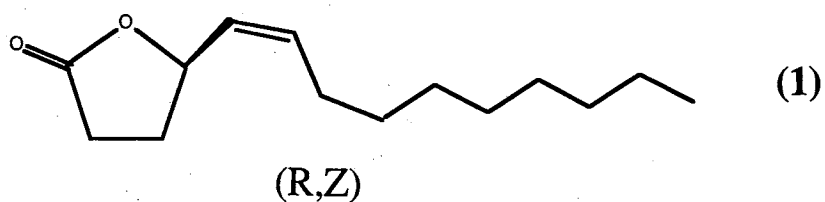


Figure 4.1: R,Z isomer of the Japanese beetle pheromone, which is the natural isomer that has been found to have reduced activity when even small amounts of the (S,Z) enantiomer is present.

It was found that as little as 1% of the (S,Z)-isomer inhibits the (R,Z)-isomer (60). The Food and Drug Administration as well as other regulating agencies in- and outside the US are contemplating whether the production of material as the required enantiomer should become a question of law. From an economic standpoint, the production of the optically pure material would save as much as half the resources of racemate production.

A routine method of producing optically pure materials in a generic manner may never be found. Very specific protocols must now be developed for each new molecule of interest. Rigorous testing must be maintained throughout the R &D of any new product to ensure that the pharmacological effects have been justifiably attributed to the correct substance. Enantiomers possess identical physical characteristics with the exception of how they rotate plane polarized light. The analysis of optically active material has been possible with spectroscopic techniques since the time of Cotton (1896). The separation of optically active compounds has been available only since the middle of the 1970's with the advent of the chiral stationary phase column used in HPLC.

## II. CHROMATOGRAPHIC METHODS USED TO ANALYZE PROTEINS

A common non-spectroscopic way of analyzing optically active molecules for the determination of enantiomeric purity is to associate the sample with a pure chiral substance (28, 71). While the physical properties of enantiomers are identical those of diastereoisomers are different. For example if the sample of interest is comprised of both the L- and D- forms of the molecule and this sample is allowed to associate with another molecule that is the pure L- isomer than the diastereoisomeric pairs of L-, L- and L-,D- which have dissimilar properties can now be distinguished. This method is used when separating enantiomers using a chiral chromatographic column. The diastereomers that are formed sometimes will have different retention times when eluting from a column. This indirect way of evoking a separation may destroy the sample. A direct method that is primarily used with HPLC is to attach to the stationary phase of the column a pure, optically active material that will associate with one enantiomer to a greater extent than the other. This method allows the stereoisomers in the sample to be separated without compromising the original sample itself (62, 64).

Besides HPLC methods there have been, in the last few years, advances in capillary electrophoresis (CE) methods that have effectively separated enantiomers by similar approaches already discussed (63, 65-67). One advantage of CE is that much smaller volumes can be used because the mobile phase is actually driven by a large voltage source. Like a large

magnet that pulls the solvent and the sample at different speeds depending on their mobility, separation is accomplished by the addition of an optically pure substance that associates with each stereoisomer of the sample. To facilitate the detection of the very small amounts in CE the optically pure substance added is also a strongly absorbing or fluorescing conjugated pi system that enhances the signal output. Often CE experiments are preceded by steps that enhance detection by covalently bonding fluorescing groups to the analyte. Cyanine-labeling as well as dansylation are both used to enhance the limit of detection (65, 66, 71).

There are dozens of different variations that have been applied to enhance the signal, increase the resolution, or decrease the analysis time of these HPLC and CE methods (71). One drawback to these methods is that the largest signal to noise occurs when a fifty fifty enantiomeric mixture is being analyzed. The least accurate samples to measure are those that have just a few percent of one enantiomer (60). To achieve detectability on very low concentrations the indirect method is needed which adds another chiral species to associate with the sample. With the analysis of a sample that contains only a few percent of an enantiomeric impurity the purity of the marker or tagging molecule becomes critical. If this agent has an optical purity which is less than 100% then this will lead to erroneous results in the determination of the enantiomeric purity of the sample (28).

For therapeutic agents that are being mass produced for public sale and consumption there is not so much a need for ultra low limits of detection as much as there is a need for methods that are accurate, rugged,

simple, quick, and can be automated. When we broaden the focus to therapeutic agents that contain more than one chiral center like proteins for instance the difficulty compounds itself. With the addition of another stereogenic center the number of stereoisomers possible multiplies by a factor of two. Enzymes and proteins contain many stereogenic centers. The introduction of the ability to synthetically create proteins to be administered to those that lack the ability to create their own and to make new proteins that perform specific therapeutic functions has led to an increase in the manufacture of protein drug forms.

Since before World War II the need for synthetically derived proteins has been acutely felt. Diseases such as diabetes were correlated with the deficiency of the hormone insulin. Other biologically active peptide drugs have been prepared and used to perform some new specific function or model the behavior of some missing enzyme (25, 27). A challenge with producing proteins and their use as pharmacological agents is that each amino acid (except for glycine) is chiral and optically active. The number of possible stereoisomers from a protein like insulin that contains 50 amino acid residues is staggering. If one chiral center is assumed to be contained on each residue then the number of stereoisomers is  $2^{50}$  or  $1.126 \times 10^{15}$ . The production of a large and complex protein is usually not undertaken in a stepwise fashion that adds each amino acid on as one would beads to a chain but the point is still important. A simple peptide of six residues would have  $2^6$  or 64 possible isomers. The difficulty of obtaining

chromatographic information on even a simple dipeptide with only two chiral centers has been shown (67).

CD can provide some information easier than many separation methods by focusing on the overall chirality of the analyte that is in a realitively pure form. The collection of QC information is the overall goal of this work and seeking a novel way of analyzing and reporting purity information of peptides and proteins.

## CHAPTER FIVE

### ANALYZING STRUCTURALLY SIMILAR PEPTIDES

#### Introduction

CD has been used to study protein structure (38, 39). These experiments have usually been carried out in the ultraviolet (UV) region. Because of the interferences that can be found in that region, CD spectrophotometric detection can be shifted into the visible region with the addition of a color derivatizing agent. The bands found in the visible range are broader than those found in the UV. CD-active derivatizing agents of choice are typically aqueous solutions of first row transition metal ion complexes in which the metal ion contributes the color and the chirality is located on the protein ligands. Asymmetric splitting of the ground and excited states of the ligand/metal electronic transitions is the origin of the CD activity (8).

Preparation of protein samples prior to inspection included creating a modified biuret reagent. The biuret reagent was first reported back in the 1940's as a method to measure total protein concentration (114). The reagent consists of copper ions mixed with tartrate in an alkaline medium. This solution is a light blue color due to the complex formed between the copper ion and the deprotonated tartrate. A measure of the intensity of the absorbance was taken using UV/Vis spectrophotometry. With the addition of protein, usually from human serum, the tartrate would be replaced by the deprotonated amide nitrogens along the peptide backbone. The complexation begins with the amino nitrogen binding first. Forming stable

five membered rings, each amide nitrogen binds onto the copper ion in a square planar arrangement until all four sites are occupied (99). The deep purple color that is a result of the protein complexed with the copper can be linearly correlated to the concentration of the protein complexed.

### **Experimental**

The system that was used for the experiments presented here used a modified biuret reagent. The host ligand has been changed to one of the ephedrine stereoisomers complexed onto the copper ion in an alkaline medium. The concentration of the copper ion was 2.0 mM and the concentration of the host ligand was kept at a constant 8.0 mM. The pH of the solution was maintained at a constant 13.0 by maintaining a NaOH concentration at 0.10 M. In most cases a stock solution of the modified biuret reagent was made up in bulk at a slightly higher concentration of copper (3.0 mM) and host ligand (12.0 mM) which could then be diluted by adding volumes of 0.10 M NaOH that was spiked with the analyte of interest.

*Modified biuret used with each specific peptide group:* To prepare the synthetic neuropeptides a modified biuret reagent was prepared with (+)-ephedrine as the host ligand. The biuret reagent prepared for the Dynorphin A set and the hexapeptide series contained (-)-ephedrine as the host ligand.

*Reagents:* The proteins studied in the first two sections of this chapter are all neurological peptides (or derivatives of that group) that belong to the



opioid peptides or endogenous morphine-like peptide group (113). These peptides are released by the body to relieve pain, produce euphoria, ease anxiety, and facilitate sleep and are generally called endorphins. A common theme in finding new therapeutic agents is to start with the sequence of a naturally produced protein and modify the primary sequence by a single addition, transposition, substitution, or deletion to produce a new function similar to the original function. For example, as is discussed in chapter six, the insulin Lispro is a synthetically derived analog of the naturally occurring human insulin that was created because its structure produces a much faster response time, thus entering the blood stream quicker (95). In the same way the set of peptides used in the first section consist of endorphin analogs that were created to treat various mental illnesses. This set of experimental neuropeptides include DTLET, DSLET, DAGO, and DADLE (See Table 5.1 for the primary sequence). The residues that are in bold print are common among the first set of peptides.

**Table 5.1: Primary Sequence of the Synthetically Engineered Analogs.**

DSLET	HO-TLFGs <b>Y</b> -H	HO-The-Leu- <b>Phe-Gly</b> -(D)-Ser- <b>Tyr</b> -H
DTLET	HO-TLFGt <b>Y</b> -H	HO-The-Leu- <b>Phe-Gly</b> -(D)-Thr- <b>Tyr</b> -H
DAGO	HO-GFG(N-Me)a <b>Y</b> -H	HO-Gly- <b>Phe-Gly</b> (N-Me)-(D)-Ala- <b>Tyr</b> -H
DADLE	HO-IFGa <b>Y</b> -H	HO-(D)-Leu- <b>Phe-Gly</b> -D-Ala- <b>Tyr</b> -H

The second section consists of a group of natural peptides. This group of endorphins including Dynorphin A(1-9), Dynorphin A(1-11), Dynorphin A(1-13), and Dynorphin A(1-13)amide were all porcine derived products. Table 5.2 shows the highlighted similarities between the

experimental neuropeptides or synthetic analogs and the Dynorphin set.

**Table 5.2: Some Naturally Occurring Dynorphins.**

Dynorphin A(1-9)	H-YGGFLRRIR-OH	<b>H-Tyr-Gly-Gly-Phe</b> -Leu-Arg-Arg-Ile-Arg-OH
Dynorphin A(1-11)	H-YGGFLRRIR-PK-OH	<b>H-Tyr-Gly-Gly-Phe</b> -Leu-Arg-Arg-Ile-Arg-Pro-Lys-OH
Dynorphin A(1-13)	H-YGGFLRRIR-PKLK-OH	<b>H-Tyr-Gly-Gly-Phe</b> -Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-OH
Dynorphin A(1-13) amide	H-YGGFLRRIR-PKLK-NH <sub>2</sub>	<b>H-Tyr-Gly-Gly-Phe</b> -Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-NH <sub>2</sub>

The third section includes a group of peptides whose therapeutic value is unknown but whose structural similarities are easily identified.

Table 5.3 shows the primary sequences for this third set where each peptide is a different stereoisomer and the differences are highlighted.

**Table 5.3: Primary Sequence of the Six Residue Stereoisomers Included in the Third Section.**

All L-parent	HO-RFMWMR-H	HO-Arg-Phe-Met-Trp-Met-Arg-H
D-1	HO-RFMWMr-H	HO-Arg-Phe-Met-Trp-Met-( <b>D</b> )Arg-H
D-2	HO-RFMWmR-H	HO-Arg-Phe-Met-Trp-( <b>D</b> )Met-Arg-H
D-3	HO-RFMwMR-H	HO-Arg-Phe-Met-( <b>D</b> )Trp-Met-Arg-H
D-4	HO-RFmWMR-H	HO-Arg-Phe-( <b>D</b> )Met-Trp-Met-Arg-H
D-5	HO-RfMWMR-H	HO-Arg-( <b>D</b> )Phe-Met-Trp-Met-Arg-H

The sequence for the natural occurring peptides are written starting with the amine end on the left and the carboxylic end on the right. The synthetic peptides are written with the carboxylic acid end on the left and the amine end on the right. When the single letter abbreviation for each amino acid is used the capital letters signify an L-amino acid and the lower

case letters signify the D-amino acid residues.

## Results and Discussion

*Experimental Neuropeptides subset:* Figure 5.1 shows the ellipticity measured versus the wavelength for the experimental neuropeptide series.

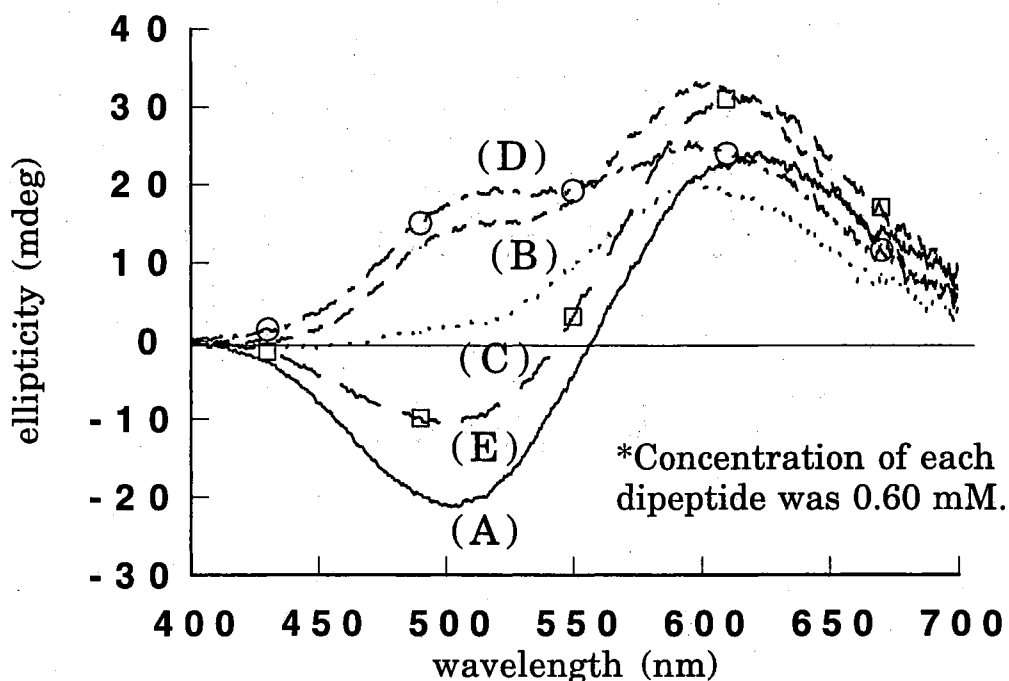
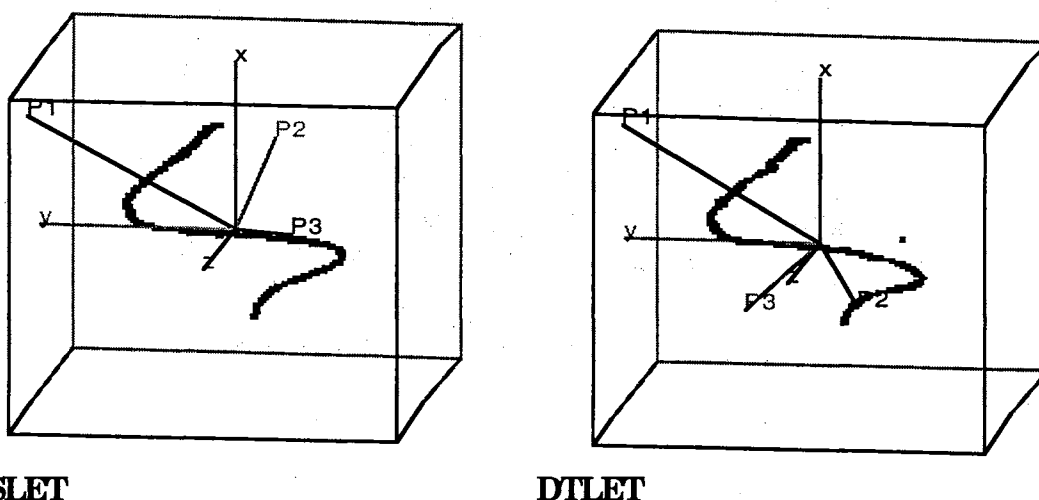


Figure 5.1: CD data for the (A) (+)-ephedrine, (B) DTLET, (C) DSLET, (D) DADLE, and (E) DAGO complexes.

The differences these data show could be reported by comparing the magnitudes of the band around 500 nm with the band around 625 nm. This method would only need the ellipticities at two wavelengths ignoring most of the data. A preferred method is to use Principal Component Analysis (PCA) see Appendix A. PCA provides solutions to a 3-D Spin Plot<sup>®</sup> which can be used to characterize the spectral differences. Figure 5.2 shows the

Spin Plots<sup>®</sup> for DTLET and DSLET with the wavelength as the first variable, the spectral data for the (+)-ephedrine as the second, and the analyte complex the third. Table 5.4 shows the PCA data for these two Spin Plots<sup>®</sup>.



DSLET

DTLET

Figure 5.2: 3-D Spin Plots<sup>®</sup> for the experimental neural peptides DSLET, and DTLET. The Principal Components P1, P2, and P3 have been included for each plot. The P2 and P3 vectors in the DSLET plot are in different directions from the DTLET plot.

Factor PC33 was chosen over all other factors because it showed the best discrimination for all four different analytes in the set. The bold-faced values in Table 5.4 show the PC33 factor that was chosen to discriminate between these CD spectra. The PC33 values for the whole set are 0.6491 for DSLET, 0.0531 for DTLET, 0.2481 for DADLE, and -0.7241 for DAGO.

This first series of analytes shows that the CD spectra are different for peptides that have some structural similarities. By choosing an appropriate concentration, the differences in the CD spectra from peptide to peptide can be increased. Also the host ligand used in the modified biuret

**Table 5.4: Data Reduction by PCA of a Spinning Plot Presentation of CD Data for the Host and Mixed Copper Complexes of DSLET and DTLET.**

<b>Principal Components</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>
<b>Eigenvalues</b>	2.4450	0.3505	0.2045
<b>Eigenvectors /(nm)</b>	0.5577	0.8215	0.1184
(+)-ephedrine	0.5912	-0.2930	-0.7515
(+)-eph/DSLET	0.5826	-0.4891	<b>0.6491</b>
<b>Eigenvalues</b>	2.2620	0.4421	0.2959
<b>Eigenvectors /(nm)</b>	0.5909	-0.3474	-0.7281
(+)-ephedrine	0.5866	-0.4346	0.6834
(+)-eph/DTLET	0.5539	0.8309	<b>0.0531</b>

reagent increases the sensitivity as well as the selectivity. The sensitivity is increased when the starting spectra is away from the baseline. The signal to noise ratio is the poorest when the ellipticity is zero so having a chiral host ligand enables very small concentration changes to be measured with less interference. The selectivity is enhanced when the starting signal is something other than a flat line. A nonlinear starting point allows bands to contribute uniquely from one analyte to another.

Because of the possibilities of both positive and negative bands the starting spectra of the host ligand can increase the ellipticity change due to the addition of the analyte. The formation constant will favor the complexation of the analyte with the copper. The “ending” spectra for the analyte with the copper will reach some end point when all the binding sites on the copper have become occupied by the analyte. The concentration of the analyte when this occurs will be the same if an achiral or chiral host ligand was used. The achiral host ligand has a CD spectra of a flat baseline. A

chiral ligand can be chosen with a starting CD spectra that starts hundreds of mdeg on the other side of the origin from the analyte.

Figure 5.3 shows an example of a host ligand that does not enhance the selectivity as well as (+)-ephedrine. The spectra for DSLET and for DADLE denoted (C) and (D) are very similar. The spectra for the DTLET is similar in shape to the DSLET and the DADLE as well. Figure 5.3 shows that achieving maximum spectral differences requires choosing the right host ligand as well as the right analyte concentration.

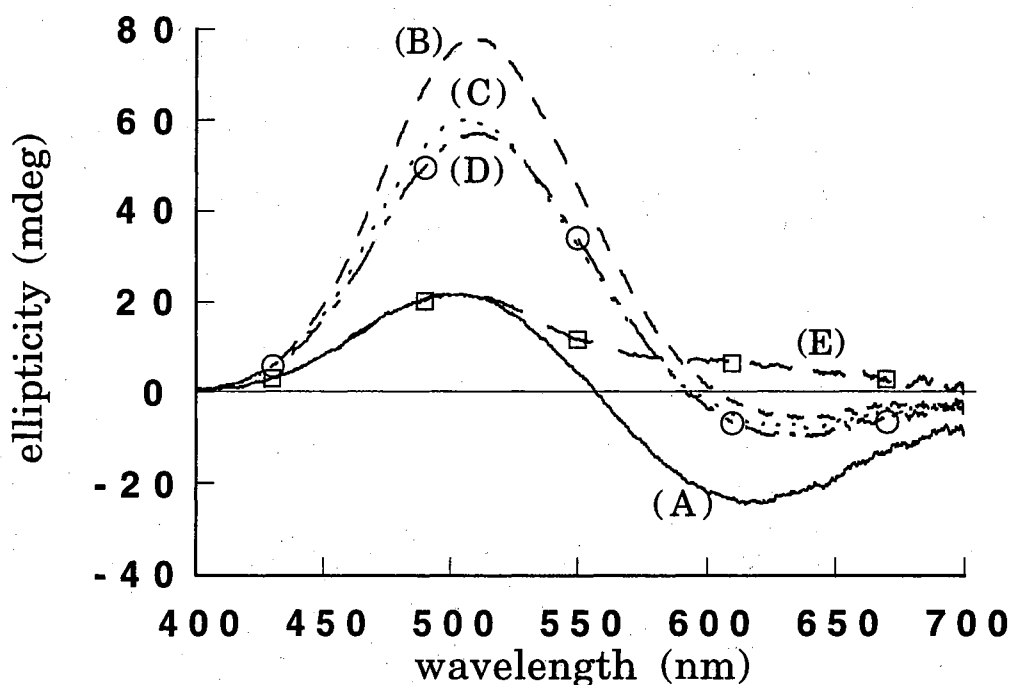


Figure 5.3: Spectral CD data for the (A) (-)-ephedrine, (B) DTLET, (C) DSLET, (D) DADLE, and (E) DAGO.

Figure 5.4 shows another example of the spectra for the same set of analytes prepared with the same ligand as Figure 5.1 only the concentration of the analyte added in each case was 0.94 mM rather than 0.60 mM.

As Figure 5.4 shows, the spectra for the DSLET and the DADLE, denoted (C) and (D), are more similar at the 0.94 mM concentration than they were at the 0.60 mM concentration. The DTLET spectra, denoted as (B), has very few spectral features to distinguish it from the (C) and (D) other than the size of the overall spectra.

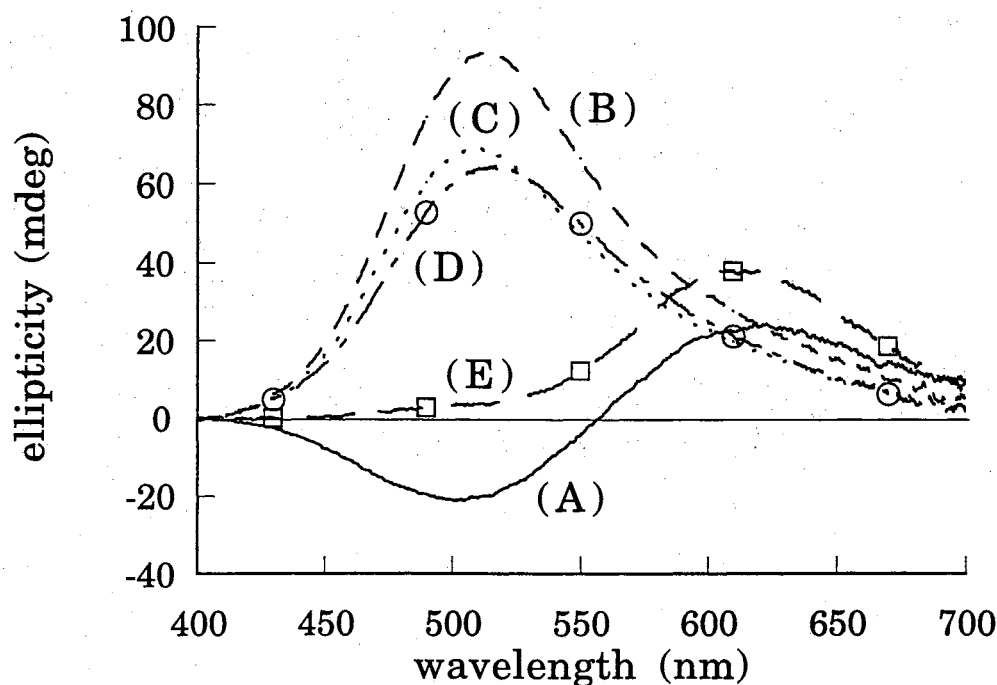


Figure 5.4: CD spectra for (A)(+)-ephedrine, (B) DTLET, (C) DSLET, (D) DADLE, (E) DAGO at 0.94 mM concentration of each.

Each analyte may have a slightly different formation constant with copper. This may account for some of the differences observed in Figure 5.4. The CD spectra obtained as the concentration of the peptide is becomes greater shows more similarities between these analytes than if the concentration of each peptide is lower. There appears to be some range, a working range, where the spectral differences between these structurally similar peptides

are more pronounced.

*Dynorphin A subset:* Since the binding of the copper ion to the peptide occurs at the amine terminus the addition of amino acids to the carboxylic end of the chain should not have as great an effect on the CD spectra as a substitution at one of the first three amino acids from the amine terminus.

The spectral differences observed for the Dynorphin A set (Figure 5.5) are actually similar in magnitude to those observed in Figure 5.1. This

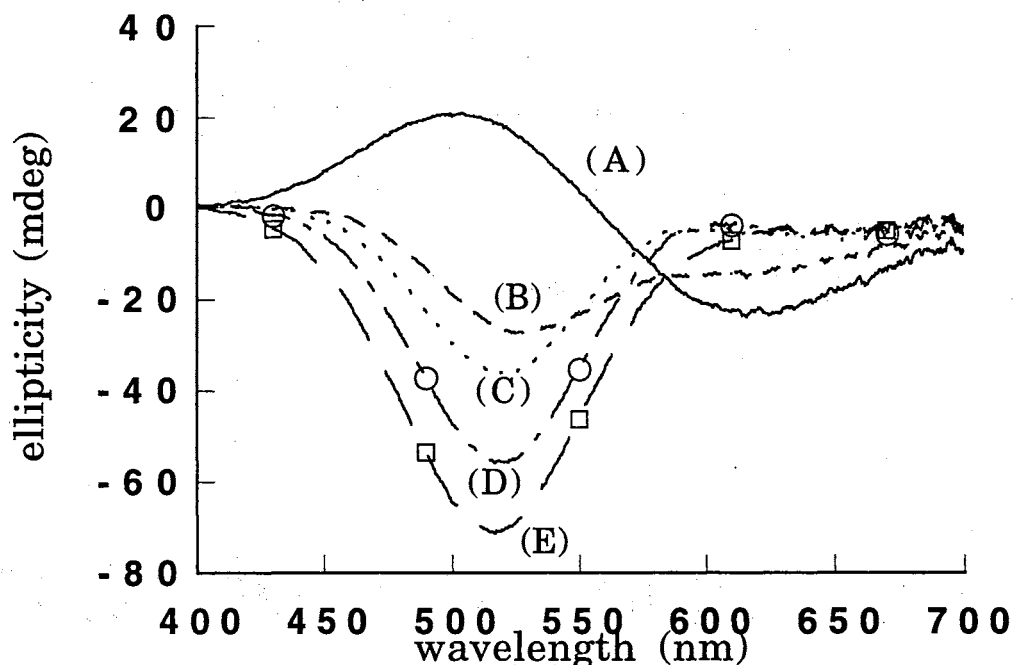


Figure 5.5: CD spectral data for the (A) (-)-ephedrine stock; (B) Dyn (1-9); (C) Dyn (1-11); (D) Dyn (1-13); and (E) Dyn (1-13)NH<sub>2</sub>. Each Dynorphin was at a constant 0.60 mM concentration.

implies that there must be some contribution to the CD spectrum from the equilibrium position that the non-ligated part of the protein assumes. This result suggests that the CD spectrum for a copper ion and a protein may in part be due to the overall structure that the protein assumes. By



scanning in the visible range, only the interaction between the chiral ligand and the colored copper center effects the resulting spectrum. The implication is that the degree of chirality maybe effected due to interactions of other groups on the protein with those groups bonded onto the color center.

The differences between the Dynorphin A spectra can be measured by comparing the principal component eigenfactors. Table 5.5 shows the PCA performed on each Dynorphin A at the 0.60 mM concentration.

The 3-D Spin Plots<sup>®</sup> were created for each analyte by plotting the

**Table 5.5: PCA factors Selected to Characterize the Dynorphins.**

Principal Components	PC1	PC2	PC3
nm	-0.7202	-0.1079	0.6853
(-)-ephedrine	0.6701	-0.3639	0.6469
(-)-eph/Dyn A(1-9)	0.1796	<b>0.9252</b>	0.3344
nm	0.5372	-0.6626	0.5219
(-)-ephedrine	-0.6889	0.0125	0.7248
(-)-eph/Dyn A(1-11)	0.4868	<b>0.7489</b>	0.4497
nm	0.5225	0.7122	0.4688
(-)-ephedrine	-0.6700	0.0029	0.7423
(-)-eph/Dyn A(1-13)	0.5273	<b>-0.7020</b>	0.4787
nm	0.5207	0.7357	0.4332
(-)-ephedrine	-0.6580	0.0225	0.7527
(-)-eph/Dyn A(1-13)NH <sub>2</sub>	0.5440	<b>-0.6769</b>	0.4958

wavelength, the spectral data for the (-)-ephedrine, and the spectral data for each Dynorphin A at 0.60 mM. The bold-faced value, PC23 is the factor that shows the greatest difference between the spectra.

*Hexapeptide subset:* The spectral data for the synthetically derived hexapeptide series are shown in Figure 5.6. In each stereoisomer there has been one stereogenic center inverted from the all L-parent analog. The CD spectra for each stereoisomer was quite different from the all L-parent as well as each other. Because of the limited supply of these analytes that were available, only one concentration of each stereoisomer was added to the host ligand. Each of these hexapeptides was prepared by adding 0.60 mM to the modified biuret in which (-)-ephedrine was used as the host ligand. The individual spectra for each stereoisomer shows maxima and minima shifts. Figure 5.6(D) shows the spectra for the stereoisomer in which the amino acid Tryptophan has been switched to the D- orientation. This spectra has a maximum at the shorter wavelengths while all the other isomers show a minimum in this region. Also notice that the amino acids

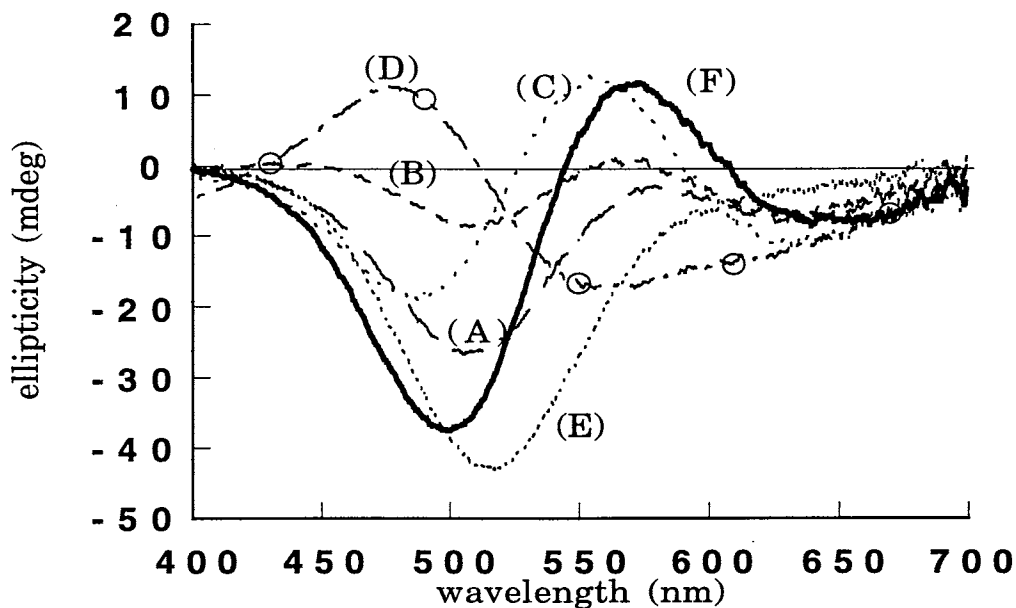


Figure 5.6: CD spectra plot of the stereoisomers of HO-RFMWMR-NH<sub>2</sub> where (A) is the all L-parent; with all L-residues except for a D-residue at position (B) 1-D; (C) 2-D; (D) 3-D; (E) 4-D; and (F) 5-D from the amine end.

at position two and four are both methionine and that the spectra for each, labeled (C) and (E) in Figure 5.6, are quite different from one another. This observation leads to the conclusion that the CD spectra can selectively show the stereogenic orientation of the first methionine compared to the second.

Spin Plots<sup>®</sup> were created for each analyte by plotting the wavelength, the spectral data for the (-)-ephedrine, and the spectral data for each hexapeptide. Table 5.6 shows the results of the PCA performed on these plots. The bold-faced values were used to characterize each stereoisomer.

**Table 5.6: PCA factors to Compare the Hexapeptides.**

Principal Components	PC1	PC2	PC3
nm	0.5240	0.7108	0.4692
<b>(-)-ephedrine</b>	-0.6687	0.0022	0.7435
<b>(-)-eph/RFMWMr(all L-parent)</b>	0.5275	<b>-0.7034</b>	0.4765
nm	-0.7168	-0.0636	0.6944
<b>(-)-ephedrine</b>	0.6599	-0.3831	0.6463
<b>(-)-eph/RFMWMr(D-1)</b>	0.2249	<b>0.9215</b>	0.3166
nm	0.6848	-0.2119	0.6973
<b>(-)-ephedrine</b>	-0.6975	0.0867	0.7113
<b>(-)-eph/RFMWmR(D-2)</b>	0.2111	<b>0.9735</b>	0.0884
nm	-0.5728	0.6266	0.5284
<b>(-)-ephedrine</b>	0.6010	-0.1172	0.7906
<b>(-)-eph/RFMwMR(D-3)</b>	0.5573	<b>0.7705</b>	-0.3095
nm	0.5213	0.7340	0.4354
<b>(-)-ephedrine</b>	0.6582	0.0210	0.7526
<b>(-)-eph/RFmWMR(D-4)</b>	0.5432	<b>-0.6788</b>	0.4941
nm	0.5383	-0.7017	0.4667
<b>(-)-ephedrine</b>	-0.6516	0.0046	0.7586
<b>(-)-eph/RfMWMR(D-5)</b>	0.5345	<b>0.7124</b>	0.4547

As with the other examples in this chapter, if the CD spectral data is visibly different then the PCA will produce factors that reflect these differences.

### **Summary**

The purpose of this preliminary chapter is to set the ground work for the analyses performed in the latter chapters. The experiments discussed in this chapter were performed to answer initial questions concerning the viability of this method. The modified biuret reagent was found to provide greater selectivity when the host ligand was chosen so that the spectral differences of the peptide complexes were enhanced. At this point there are a few rules that can aid in the choosing of the host ligand; (a) the host ligand must be exchangeable to allow the peptide to ligate to the copper; (b) in most cases the sensitivity is increased when the spectra of the host ligand complex is on the other side of the baseline from the peptide complex spectra for a large part of the visible region; and (c) the biphasic nature of the (-)-ephedrine complex produces spectral features that can enhance the selectivity. The method of choosing CD as a detector and using a modified biuret reagent to prepare the proteins for analysis becomes more important with the addition of the data reduction method.

PCA is used in the analysis of the spectra of different proteins to provide a measure that is unique for that spectra and that protein complex. Traditional spectroscopy reporting techniques uses the absorbance (for instance) at the maximum to identify a specific band. This method ignores most of the information gathered which is generally Gaussian in shape.

Specifically, absorbance spectroscopy does not contain as many features as CD spectra because the absorbance can only be positive while ellipticity can be both positive and negative. New techniques are needed to report CD spectral information both because there is more information in each spectrum and also because the personal computer makes data manipulation easier to access information that was formerly hidden by tedious data reduction. As the later chapters will show, PCA is sensitive to concentration changes associated with the addition of a chiral impurity. A factor, which is a cartesian coordinate of a particular principal component, can be monitored to find a distinct value to represent the particular analyte as was done in the three examples given here. This is done to show that quality control information can be derived using the factors obtained from PCA and routine analysis using the method outlined in this and other chapters. The focus of this work is the ability to distinguish between structurally related proteins and peptides, a more rigorous approach is presented in the next chapter.

## CHAPTER SIX

### CHIRAL PROPERTIES OF INSULINS

#### Introduction

Since Frederick Sanger determined the complete amino acid sequence of the bovine polypeptide hormone insulin in 1953 (Figure 6.1) scientists have known that proteins have unique structures which guide their molecular mechanisms of action. Later, with the advent of

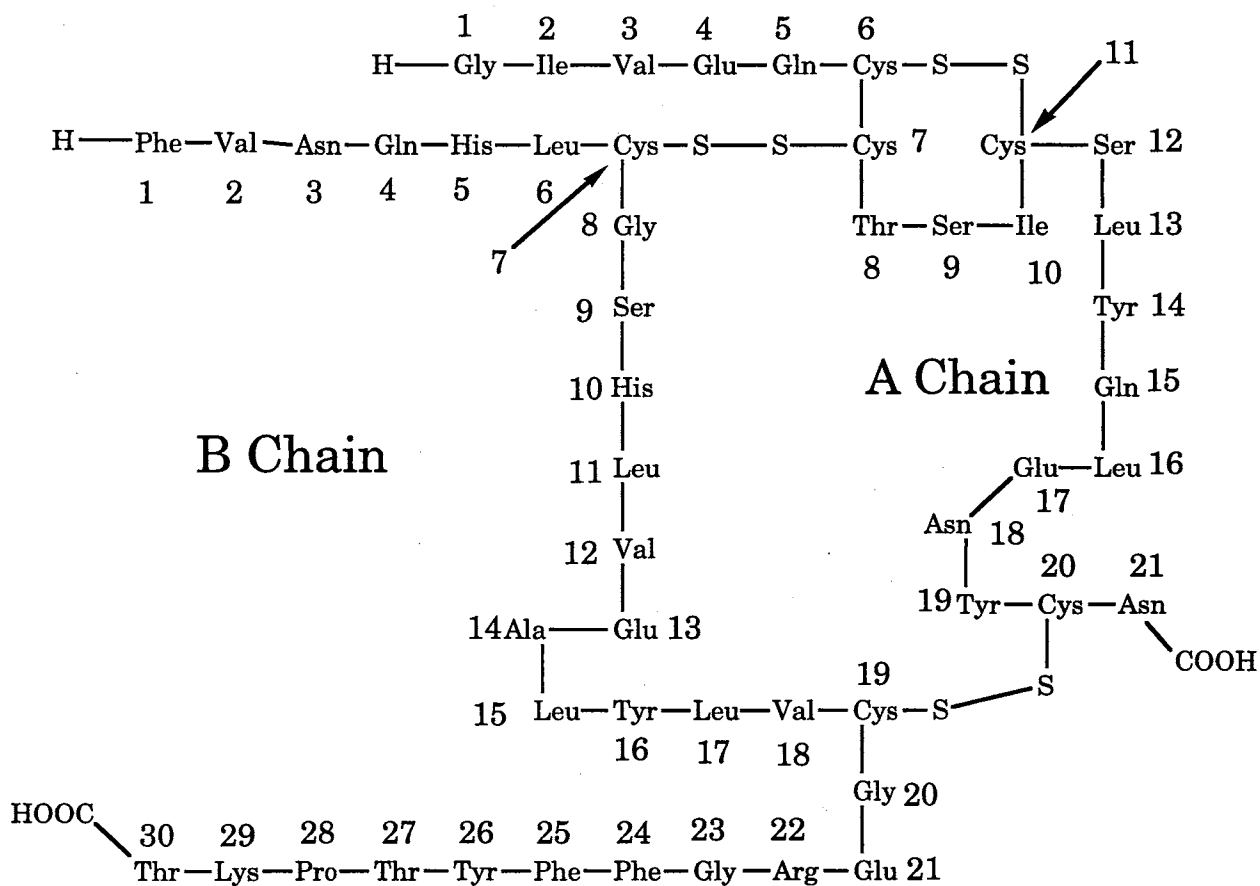


Figure 6.1: Human Insulin. The B Chain and the A Chain are connected by sulfur-sulfur bonds through the amino acid groups Cysteine. The only difference in sequence between Porcine and Human insulin is Alanine for Threonine at position 30 on the B Chain. Bovine insulin also has an additional sequence substitution at position 8 and 10 of the A Chain with the amino acid groups Alanine and Valine respectively.

X-ray and NMR analyses which are used to assist in the computer generated images of supposed structures of proteins and enzymes, the manufacture of these molecules were possible for clinical applications.

The amino acid sequence of human insulin was published in a 1960 edition of *Nature*. Soon after this the Novo Research Institute began attempting to substitute the B30 alanine residue of porcine insulin with a threonine residue, thereby semisynthetically converting porcine insulin to human insulin. Complete synthesis of insulin types were carried out in a few weeks with specific activity comparable to that of the natural hormone (96-98).

Since 1982 Eli Lilly has been using bacteria to mass produce human insulin. Despite the fact that an exact chemical copy of the endogenous insulin secreted by the pancreas is available, the time course of action for these hormones are slow in onset and prolonged in duration (90). Higher incidence of blindness and liver failure result in diabetics due to the inconsistencies of their blood sugar level. Eli Lilly & Co. developed an analog, lispro, that has a much faster onset time and reduced duration by switching the proline and the lysine on the B chain at the position 28 and 29. There is no difference in strength between regular human insulin and insulin lispro other than it is absorbed faster, peaks quicker, and doesn't last as long.

Following the modified biuret test used on the small peptides in chapter five, the question was raised whether this spectroscopic method could discern and record spectral differences between regular human

insulin, bovine insulin, porcine insulin, and insulin lispro for QC applications.

## **Experimental**

*Reagents:* The analytes were the insulin themselves received from the manufacturers in various forms. The sources of these insulins were Eli Lilly & Co., Novo Nordisk, and Sigma. Altogether 38 lots were investigated, but not all were used quantitatively. Some lots contained the Zn-crystalline form (Lilly, Novo) while others were Zn-free (Novo). The regular human insulin came from E. Coli (Sigma), yeast (Sigma), and recombinant DNA expressions (Lilly, Novo). The bovine insulins came from either Sigma or Lilly and were described as Beef Purified or from bovine pancreas. Lilly and Sigma were the sources of the Pork Purified Zn-porcine insulin. Several different lots of the insulin lispro were obtained from Lilly. The separated chains of bovine insulin were obtained from Sigma.

*Derivatizing agent:* Stock solutions of Cu(II)-(D-Histidine) host complex were made such that  $[\text{Cu}^{2+}] = 3.0 \text{ mM}$ , and  $[\text{D-Histidine}] = 12.0 \text{ mM}$ . The pH was adjusted to 13.0 using NaOH. The stabilizer (KI) was also added to a final concentration of 0.03 M. The stock solution was made to be used within the week but would keep for several weeks if needed. The working solutions were made by taking 10.0 mL of the stock solution and adding 5.0 mL of 0.10 M KOH for a final volume of 15.0 mL and final concentrations of  $\text{Cu}^{2+} = 2.0 \text{ mM}$  and D-Histidine = 8.0 mM at a pH of 13.0.



Appropriately weighed out samples of insulin were added to the 15.0 mL working solution for an insulin concentration of 150  $\mu\text{M}$ . This concentration is far below the copper ion concentration and showed in preliminary results to produce the greatest spectral differences between the host and the different mixed complexes.

*Measurement:* Spectra were measured using a Jasco 500-A as described in the preliminary chapters. The experimental parameters were: wavelength range 400-700 nm; sensitivity 20 mdeg/cm; time constant 0.25 sec; scan rate 200 nm/min; single scan; pathlength 5.0 cm; temperature ambient. The CD spectra of the working solution of Cu(II)-(D-Histidine) was used to check the day to day reproducibility of the system. Standard deviations for the reproducibility of the maximum ellipticities measured at wavelengths 487 nm and 682 nm were  $7.42 \pm 0.07$  mdeg and  $-214 \pm 0.60$  mdeg respectively.

The additions of each insulin sample were made immediately prior to analysis. The spectral measurements were made 3-5 minutes after mixing. This was to insure that any spectral changes observed were not a function of unfolding or other rearrangement that may occur over time.

## **Results and Discussion**

The D-Histidine concentration is 500 times that of the insulin, and is, according to the law of mass action, heavily favored to bind with the copper ion. The complex formed between the insulin and the Cu(II) is much less than stoichiometric so the final product of the exchange is a mixture of

original host and the Cu(II)-(D-Histidine-insulin) complex. The insulin samples that were identified as containing  $Zn^{2+}$  had no observable spectral changes compared to those that did not contain the  $Zn^{2+}$ , because the concentration difference between the Cu(II) and the Zn(II) was a factor of 500. The equilibrium between the two CD-active complexes showed a nonlinear dependence upon the concentration of the insulin. Therefore the same total concentration of insulin was maintained for each sample. The change in the molecular weight of different insulin types was negligible compared to the total molecular weight so the same mass was used for each sample. At a final pH of 13.0 the formation constants for each of the insulin types were considered to be the same. Differences observed in the CD spectra were the result of the differences in the amino acid sequence itself.

CD spectra for each insulin type were averaged and the resulting mean spectra are shown in Figure 6.2. Spectra for the human and porcine insulins appear to be equivalent. This is not unexpected due the similarity between the two, only the amino acid 30 in chain B is different. The physiological activity of human and porcine insulin is very similar as well (30). Spectral differences do exist between this pair and the bovine and insulin lispro. Thus, initially, three out of the four insulin types show spectral differences.

Mean spectra were the results of many repetitive measurements from different lots obtained from each manufacturer. Ten human insulin

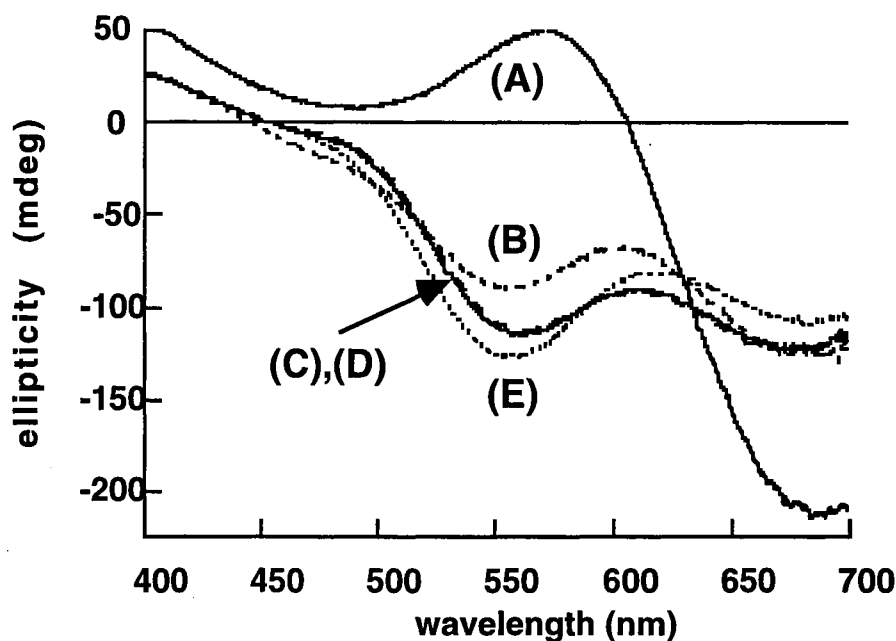


Figure 6.2: Visible range CD spectra for: (A) the Cu(II) complex with D-Histidine; and the mixed Cu(II) complexes of D-Histidine with (B) bovine insulin; (C) human insulin; (D) porcine insulin; and (E) human Lispro insulin.

master lots obtained from Lilly (n=31), Novo (n=22), and Sigma (n=2); three porcine lots from Lilly (n=24) and Sigma (n=7); four bovine lots from Sigma (n=7) and Lilly (n=8); and three lispro lots from Lilly (n=22).

To evaluate the robustness, comparisons were made between lots obtained from the same manufacturer and lots obtained from different manufacturers. Also comparisons were made between preparations made from the same lots at different times, often up to a month apart. The spectral reproducibilities for the mixed complexes were compared by monitoring the ellipticities at local maxima in the visible region. Values for the same insulin types agreed to better than  $\pm 2.0\%$  showing the

ruggedness of this method of analysis.

Overlaying the spectra for the same insulins from different manufacturers creates only a subjective test that does not verify that all samples are equivalent and that all conform to the specifications of a single standard reference material. Other ways of quantifying the data are essential for the creation of a quality control protocol.

*2-D Data Reduction Algorithm for Enhancing Selectivity* : To determine if two spectra are equivalent, two spectra from the same lot were plotted against one another. The resulting plot, if the spectra for each sample are the same, will show a perfect linear correlation with a slope of 1.0, e.g. Figure 6.3, for the human insulins. For materials that are the same but of unequal purity, the correlation coefficient will still be 1.0 if the impurity is either non-chiral or enantiomeric but the slope will be greater or less than 1.0. Correlation lines are neither straight nor have a slope of 1.0 for materials that are not the same, Figure 6.4, where human insulin data are plotted against data for insulin lispro, bovine, and porcine forms.

Results from Figure 6.4 provide further evidence that there are real chiral differences between the human insulin and the lispro and bovine insulins when complexed to Cu(II) ion. The plot also shows a more mathematical way of reporting the spectral differences of Figure 6.2. Considering that the only difference between the human insulin and the human lispro is the exchange of proline and lysine on chain B at position 28 and 29 there is a surprising amount of spectral difference observed in Figure 6.2 that corresponds to the nonlinearity shown in Figure 6.4. The

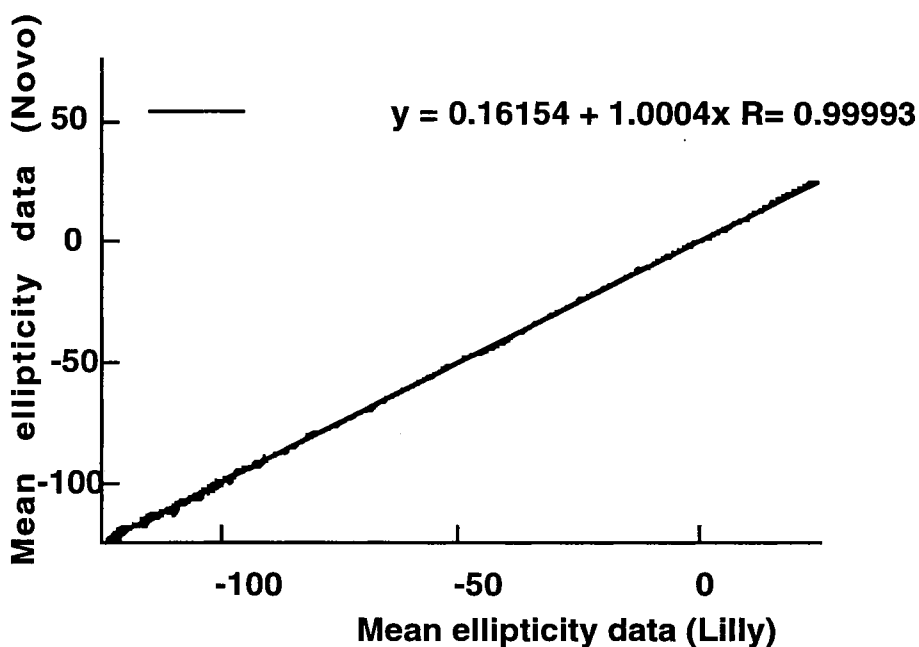


Figure 6.3: Correlation of the full spectral data for the mean of all Lilly human insulin mixed complex spectra (x-axis) against the mean of all Novo human mixed complex spectra (y-axis). The slope and the correlation coefficient are indicative of their relative equivalence.

only spectral evidence that indicates the human and porcine insulins are not collinear is the - 100 mdeg region where slight spreading can be observed. This nonlinearity may account for the 2.1% increase in the slope to 1.021 but may not provide enough evidence to prove that differentiation has been achieved. The only structural difference between the human and porcine insulin is the substitution of an alanine residue for a threonine on chain B at position 30, which is far removed from the amine end that is involved in bonding to the Cu(II) ion (99) and therefore is expected to contribute little to the spectral differences observed.

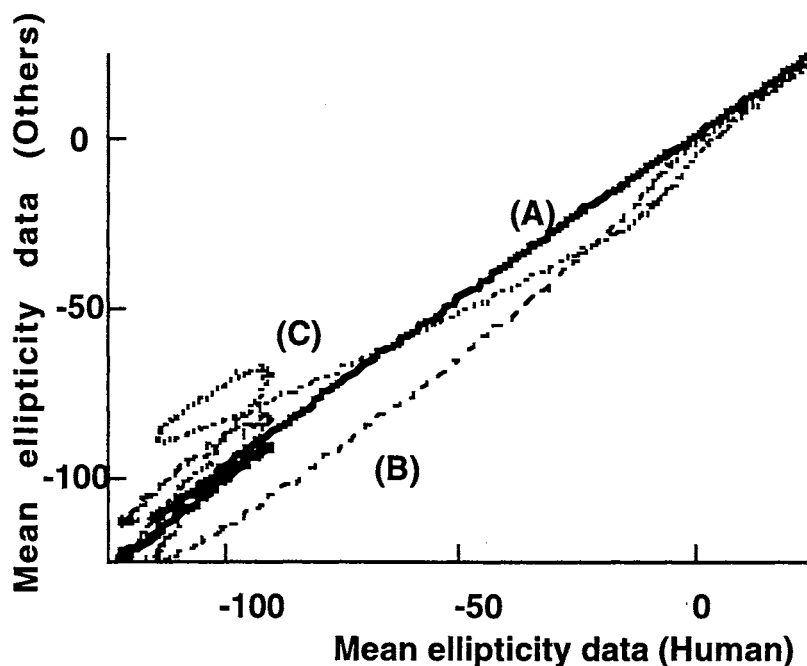


Figure 6.4: Correlations of the full spectral data for the mean of all Lilly human insulin mixed complex spectra (x-axis) against the means for the corresponding data for (A) porcine; (B) human lispro; and (C) bovine insulins. Only the human vs. the porcine correlation approaches linearity for which the regression equation is  $(y = 1.2021 + 1.021x, R = 0.998)$ .

The spectral differences observed with the exchange of proline with lysine in the human insulin/human lispro comparison show great sensitivity to the primary sequence of a protein and may be the largest spectroscopic difference observed for these two insulin forms. The comparison of the spectral data for one insulin type with another gives a numerical way of deciding whether substances are equivalent which is an improvement over the simple superposition method shown in Figure 6.2.

*A 3-D Data Reduction Algorithm for Enhancing Selectivity* : When

making 2-D plots shown in Figure 6.3 and 6.4, the wavelength is an implied variable. If the wavelength is used as the x-axis, ellipticity data from the host complex the y-axis, and ellipticity data from the mixed complex the z-axis, a 3-D plot is generated. Spinning Plots<sup>®</sup>, which are 3-D plots that can be viewed from all positions, were derived for each mixed complex spectrum taken (29, 115). Ellipticities can be positive, zero, and negative values. Therefore, when two CD spectra are plotted against each other, four sign combinations are possible at every wavelength. For example when looking at Figure 6.2 the ellipticity value at 550 nm is going to be positive for the host complex and negative for each of the mixed complexes, while at 650 nm the sign is negative for all of the ellipticities measured. In a 3-D plot the ellipticity values for the host complex and the mixed complex can coincide at more than one place but because the wavelength was chosen as the z-axis and it does not have two values that coincide the result is a 3-D plot where the function “wraps around” the z-axis. In retrospect what are observed as 2-D plots in Figure 6.3 and 6.4 are actually 3-D plots projected onto the x-y plane. Figure 6.5 shows the 3-D plot of human lispro and bovine insulin. Subtle differences can be seen in both the length and the positions of the principal components. P1 and P2 for the bovine insulin are pointing in noticeably different directions from the lispro. The added value of the third dimension is that with Principal Component Analysis a 3x3 matrix of eigenvectors can be produced that characterizes the spectral features and numerically allowing for analytical analysis.

*Data Reduction by Principal Component Analysis* : To be able to

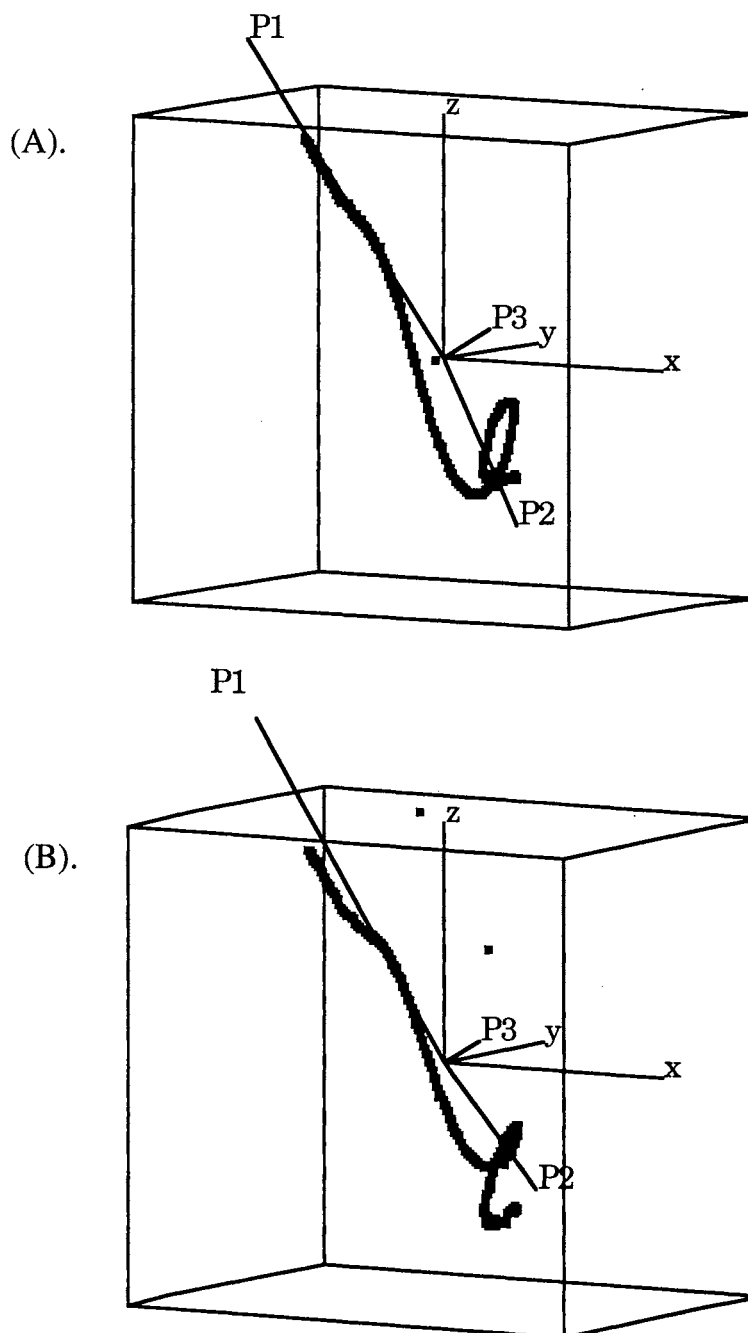


Figure 6.5: Spinning plots for the spectral data for (A) human lispro and (B) bovine insulins. The variables are the wavelength (x-axis), the full spectral data for the host complex (y-axis), and the full spectral data for the mixed complex (z-axis). P1, P2, and P3 are the principal component axes from the factor analysis.

quantitatively compare the spectra for the mixed complexes, data reduction



was carried out using Principal Component Analysis (PCA). This factor analysis method of data reduction was performed on the full range of the spectral data (1500 data points). Appendix A contains more detailed information about this process. PCA results are expressed in matrix form as eigenvalues and eigenvectors for each of the principal components for each of the analytes. Table 6.1 is an example readout for human insulin. The software package used for the analysis was JMP 3.2.2 from SAS Institute Inc. This program would quickly (in a matter of seconds after the data was collected) allow the data to be analyzed and could easily be incorporated into the data collection procedure.

<b>Table 6.1</b>			
<b>Data Reduction by Principal Component Analysis of a Spinning Plot Presentation of CD Data for the Host and Mixed Copper Complexes of Human Insulin.</b>			
<b>Principal Components</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>
<b>Eigenvalues</b>	2.4993	0.4813	0.1094
<b>Eigenvectors</b>			
nm	-0.6270	0.1122	0.7709
D-Histidine	0.5290	0.7878	0.3155
D-Hist/Human Insulin	0.5719	<b>-0.6056</b>	0.5533

The pictorial 3-D plot is not necessary because any differences observed between mixed complexes are subjective and not quantitative unless factor analysis has been performed. Visible differences between Figures 6.5 (A) and (B) are clarified when principal component axes are included. Lengths and angular directions differ. For example PC1 is

longer in (B) than in (A) and the PC2 is oriented differently in space between the two.

For the spectral data obtained from each mixed complex PCA data reduction was performed and the eigenvectors compared. The factor that was the most sensitive to the identity of the insulin ligand is the PC2 eigenvector associated with the D-Histidine/insulin mixed complex, i.e. the highlighted value in Table 6.1, bottom row. Correlating the variation of these numbers with the identity of the mixed complex has the potential to show 100% differentiation between all four insulin types. The mean values of the PC2 data for the mixed complexes are listed in Table 6.2, as well as the standard deviations (SD) for each type, manufacturer, and lot of insulin.

Since the SD's are all below 0.010, the mathematical and statistical evidence confirms that all of the human insulin lots from Lilly and Novo that were sampled are equivalent. This result correlates well with the plot in Figure 6.3. Sigma human samples do not correlate as well but the sample size ( $n=2$ ) is not statistically significant to validate this apparent difference. Lispro and bovine insulin PC2's are statistically different from each other and from the human and porcine PC2 values. The ability to judge the type of insulin sample by monitoring one single number is a big improvement over the nonlinear 2-D plots shown in Figure 6.4. A single decimal fraction, which is one normalized coordinate to the PC2 vector, could become a unique value for each insulin type. Bovine and porcine PC2 values are different enough to be able to determine the proportion of each in

**Table 6.2:**  
**Comparison of Mean Values for PC2 Eigenvectors for D-Histidine/Insulin Complexes Among Insulin Types, Manufacturers, and Lot Numbers.**

Type	Source/repeats	Lot	PC2(means)	SD
Human	Lilly(n=1) r.DNA, Zn cryst.	LH-1	-0.59866	
	Lilly(n=5) r.DNA, Zn cryst.	LH-2	-0.60679	0.0101
	Lilly(n=13) r.DNA, Zn cryst.	LH-3	-0.59786	0.0015
	Lilly(n=6) r.DNA, Zn cryst.	LH-4	-0.60734	0.0062
	Lilly(n=6) r.DNA, Zn cryst.	LH-5	-0.60513	0.0072
	<b>Lilly(n=31)</b>		<b>-0.60256</b>	<b>0.0049</b>
Human	Sigma (n=2) E.coli	SH-1	-0.58383	0.0014
Human	Novo Nordisk(n=6) r.DNA	NH-1	-0.59979	0.010
	Novo Nordisk(n=4) r.DNA	NH-2	-0.60573	0.010
	Novo Nordisk(n=5) r.DNA	NH-3	-0.61534	0.0036
	Novo Nordisk(n=7) r.DNA	NH-4	-0.60533	0.0056
	<b>Novo Nordisk(n=22)</b>		<b>-0.60617</b>	<b>0.0071</b>
<b>Human</b>	<b>Lilly, Sigma, Novo (n=55)</b>		<b>-0.60332</b>	<b>0.0057</b>
Porcine	Sigma (n=7) Zn cryst.	SP-1	-0.59974	0.0067
	Lilly (n=18) Zn cryst.	LP-1	-0.59332	0.0097
	Lilly (n=6) Zn cryst. (repeats)	LP-1	-0.59354	0.0060
<b>Porcine</b>	<b>Sigma and Lilly (n=31)</b>		<b>-0.59481</b>	<b>0.0083</b>
Lispro	Lilly (n=4) Zn cryst.	LL-1	-0.65921	0.0028
	Lilly (n=3) Zn cryst.	LL-2	-0.66617	0.0021
	Lilly (n=15) Zn cryst.	LL-3	-0.67700	0.0056
<b>Lispro</b>	<b>Lilly (n=22)</b>		<b>-0.67229</b>	<b>0.0046</b>
Bovine	Sigma (n=7) pancreas	SB-1	-0.51931	0.0100
	Lilly (n=3) Zn cryst.	LB-1	-0.52148	0.0025
	Lilly (n=2) Zn cryst.	LB-2	-0.54216	0.0036
	Lilly (n=3) Zn cryst.	LB-3	-0.52649	0.0080
	<b>Sigma and Lilly (n=15)</b>		<b>-0.52423</b>	<b>0.0073</b>

a binary mixture. An even greater improvement in sensitivity could be obtained if larger concentrations were used in the working solutions.

The PC2 values for the human (n=55) and the porcine (n=31) insulins are very similar in magnitude and the SD's overlap enough to make the mean values not statistically different. Recall however that a difference was observed as a slight loss of linearity in the spectral correlation found in Figure 6.4. The chiral recognition of a singular change in a peptide sequence at a position so far removed from the active binding site shows the great potential of this method to validate the quality of peptide drug forms.

*Bovine chain A and chain B* : Chain A and chain B of bovine insulin were examined separately using the same protocol. From Figure 6.1 it is shown that the initial sequence of chain A is NH<sub>2</sub>-Gly-Ile-Val-Glu- and chain B is NH<sub>2</sub>-Phe-Val-Asp-Gln- and is common to all four insulins. The spectral results of chain A and chain B (Figure 6.6) show that each chain can exchange with the D-Histidine. The resultant CD spectra are completely different from each other and from the spectrum of the intact bovine insulin. As expected this determines that the PC2 values shown in Table 6.3 are different between chain A, chain B, and bovine insulin. Results indicate that a sensitivity to the initial sequence exists but that the presence of the sulfur-sulfur bonds, which change the folding and absolute conformation of the molecule, may affect the spectral changes as well. CD spectra that are measured for a prepared equimolar mixture of chain A and chain B are superimposable on the spectrum that is the calculated

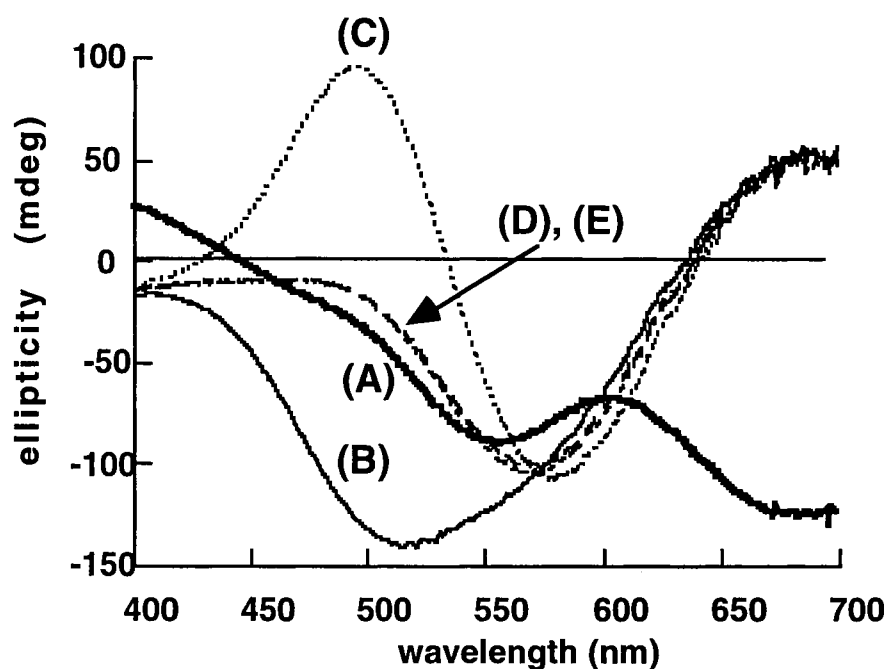


Figure 6.6: CD Spectral data for mixed complexes of D-Histidine with: (A) intact bovine insulin; (B) bovine insulin chain A; (C) bovine insulin chain B; (D) an equimolar mixture of chain A and chain B; and (E) the average value of the sum of the spectra for chain A and chain B.

mean of the spectra for the separate chains, see Figure 6.6. This result indicates that the two chains exchange independently, at least when they are separated, but it does not show that both chains are involved when the insulin exchanges. It may be that multiple substitution is an important factor to consider when accounting for analyte selectivity. This study, among other things, gave good evidence that quantitative work can be done that focuses on the chirality of a protein at more than one location in a manner that is time effective and may be applied to other systems.

**Table 6.3:**  
**Bovine Chain A and Chain B PC2 Values for Separate Mixtures,**  
**Equimolar Mixtures, and the Spectral Average of the Two.**

Type	Source	Lot	PC2 value
Bovine Chain A	Sigma	SBA-1	-0.69856
Bovine Chain B	Sigma	SBB-1	0.91557
Chain A, Chain B Equimolar Mixture	Sigma	SBA-1 + SBB-1	0.71310
Chain A, Chain B Spectral Average	Sigma		0.71226

### Summary

A method to validate the primary sequence and absolute conformation has been described which has also been shown to measure the chemical purity of proteins and peptides for the purposes of quality control. The ruggedness as well as the reliability of the method were shown to exist over several months of testing, various manufacturers, and several different master lots. Mathematical models were introduced that give characteristic numerical information that can aid in the determination of the identity of an insulin.

CHAPTER SEVEN  
QUANTITATIVE VALIDATION OF THE CHIRAL  
PROPERTIES OF PEPTIDES

**Introduction**

Pharmaceutical and other biotechnology companies are committed to the production of chiral drug substances (27). Manufacturers have the option to prepare chiral drugs either as pure single substances (enantiomers) or as racemic mixtures. Many times the racemic mixture is produced with a nonspecific synthesis and a pure substance is produced with a natural process. Racemic mixtures are easier and cheaper to produce though there are many good reasons for choosing to manufacture enantiomerically pure forms of a drug molecule. These reasons include but are not limited to considerations of the relative therapeutic effects of each enantiomer, toxicity of each enantiomer, and wasted resources used in the production of a form that does not contribute to the activity. More than just good reasons are needed to be able to manufacture enantiomerically pure compounds. A routine analytical method is needed that is inexpensive, quick and adaptable to all chiral drug forms. This test can be used by a manufacturer to control costs by closely monitoring the chirality of the product, and can be used by regulating agencies to insure the safety of the consumers.

The current analytical options available for these types of analysis include tests that involve the simultaneous derivatizations of both enantiomers in a racemic or partial racemic mixture to their

corresponding diastereoisomers by selective reactions with a third chiral species. Unlike enantiomers, diastereoisomers can be differentiated by physical properties other than just the direction of rotation of linearly polarized light. Chiral chromatography is currently the benchmark method in analysis of chiral compounds, either by direct separation, where there is a chiral stationary phase, or indirect where a chiral derivatizing agent is added without a chiral stationary phase. Since the diastereoisomers elute at different retention times, simple absorption, fluorescence, electrochemical, or mass spectrometers detectors can be used. The limit of detection is different for each.

When determination of the chemical and enantiomeric purity of a material is desired without first performing some method of separation, dual detectors methods are usually employed which measure different properties of the same sample, i.e. optical activity and absorption. Using a combination of CD and absorbance detection, spectral differences or spectral ratios can be used to find chemical and enantiomeric purity information. Most of these methods are based on single wavelength detection data.

There is precedence in a method that employed multiple wavelength detection with modern chemometric methods for data analysis. The procedure exploits the best characteristics of both of the other methods; use of a single chiral detector; bulk *in situ* derivatizations; no separation was needed. Results obtained in that study showed that the enantiomeric purities determined using visible CD detection for the four ephedrine



stereoisomers complexed to Cu(II) ion, were an improvement over what was capable at that time by either the chiral chromatographic or two-detector methods.

It has been reported that chiral chromatography is limited by the number of chiral centers that can be resolved for any given species (67). The number of potential stereoisomers increases by a factor of  $2^n$  where  $n$  is the number of chiral centers. With elution times being very similar for these isomers the complexity of resolution becomes insurmountable, for example with 64 isomers possible for only a six residue peptide. Baseline fluctuations in the detector make determinations of the enantiomeric purity best with a 50:50 racemic mixture and worst when impurities of just a few percent are being measured. This makes the measurement of very impure samples more accurate but errors of 200% have been reported at the 1% impurity levels (108).

The pharmaceutical industry is focused on the chirality of the whole molecule and how that affects its therapeutic value. The total CD signal for a metal-peptide complex is not determined by just the number and the order of chiral centers in the primary peptide sequence. It also includes contributions from longer range chiral interactions between sidechain substituents that modify the ternary structure when peptides are coordinated to metal ions. Experimental conditions must carefully control the parameters that would affect these pH-sensitive structural modifications. On the other hand because they are sensitive and accumulative, these chiral properties could produce a degree of analytical

selectivity that is unmatched by other detectors.

The object was to prepare a series of simple dipeptides which have one or two chiral centers and complex these with Cu(II). Since the molecules are small they will not have a predominant portion of each molecule outside of the coordination sphere of the copper ion. The question to be answered is whether these very similar molecules will show spectral differences that can be used to show both enantiomeric and chemical purity information.

Dipeptides chosen for the study have sequences that are permutations of only three L- and one D- amino acid monomers. The molecules have no ternary structure, so the sequence variation is the only factor that will affect the selectivity.

## Experimental

*Chemicals* : Nine of the dipeptides used in the study are analogs of glycine (G), L-Alanine (A), and L-Tyrosine (Y), Figure 7.1. Each amino acid occupies a position at either end of the peptide chain, GG, GA, and GY;

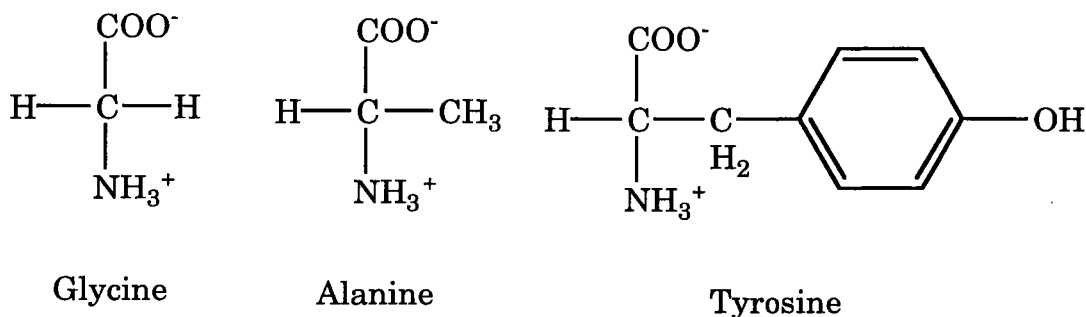


Figure 7.1: Covalent structures of the amino acid monomers used. Alanine (A) and Tyrosine (Y) are optically active while Glycine (G) is not.

AG, AA, and AY; YG, YA, and YY (the amine end is always listed on the left).

The last peptide is the D-Alanine enantiomer of glycylalanine, abbreviated G(D)A. All peptides were supplied by Sigma Chemical Co. which reported an enantiomeric purity in excess of 99.8%. Reagent grade D-Histidine, (from Sigma) used in the manner described in Chapter VI, was used to calibrate the CD scale. Fisher Scientific was the supplier of the reagent grade  $\text{Cu}(\text{SO}_4) \cdot 5\text{H}_2\text{O}$ .

*Solution Preparations* : Aqueous stock solutions were made for each of the Cu(II)-dipeptide complexes in pH 13.0 solutions with the  $[\text{Cu}^{2+}] = 3.0 \text{ mM}$  and  $[\text{dipeptide}] = 12.0 \text{ mM}$ . To stabilize the solutions 0.030 M of KI was added to each stock solution. To create working solutions the stock solutions were diluted with 0.10 M NaOH, so that the final concentrations were  $[\text{Cu}^{2+}] = 2.0 \text{ mM}$  and  $[\text{dipeptide}] = 8.0 \text{ mM}$ .

Mixed dipeptide solutions were made which contained a predetermined amount of two dipeptides with ratios of 100:0, 99.0:1.0, 97.0:3.0, 95.0:5.0, and 90.0:10.0 of the host to the impurity. Typical working solution concentrations are shown in Table 7.1. When G(D)A was added as an impurity to the GA host, ratios of 65.0:35.0, 52.0:48.0, and 50.0:50.0 were also prepared.

The chemistry for the derivatization reaction was described in Chapter VI and uses the same modified biuret color reaction for detection of

**Table 7.1:**  
**Ratios and Concentrations of the Dipeptides used in the Chemical Purity Study where GA was the Host and the Other Nine dipeptides the Impurities.**

Ratio	[Host]	[Impurity]
100:0	8.0mM	0mM
99:1	7.92mM	0.08mM
97:3	7.76mM	0.24mM
95:5	7.60mM	0.40mM
90:10	7.20mM	0.80mM

the chirality of the dipeptides in the visible region. The host ligand was the dipeptide GA and the chemical impurities were the eight dipeptides GG, GY, AG, AY, AA, YA, YY, and YG. The enantiomeric impurity was the G(D)A dipeptide.

*Measurements* : CD were measured using the Jasco 500-A as described in Chapter VI with the same parameters chosen including the wavelength range. There is not a danger of protein tertiary structure unfolding with peptides that are so short but the mixed complexes were still analyzed in 5-10 minutes after the solutions were made to insure comparability.

## Results and Discussion

*Cu(II)-amide complexes* : The local microsymmetry of the Cu(II) ion in aqueous solution is essentially square-planar due to Jahn-Teller distortion, or axial elongation, of the local octahedral symmetry adopted by most first row transition metal ions. Complexation serves to keep the Cu(II) ion in solution at high pH conditions (99).

In a highly alkaline environment the amide-nitrogen protons are fully ionized on each dipeptide which allows them to compete for binding on the copper ion. The D-Histidine, used as the host ligand, is thought to ligate to the copper at high pH through a three-coordinate arrangement binding at the carboxylate oxygen, the amine nitrogen atom, and the pyrimidine nitrogen atom. The stoichiometry for the D-Histidine copper ion complex is 1:1 (99).

Complexing a peptide to Cu(II) at high pH involves first attachment through the nitrogen atom of the terminal amine group followed by ring closure through bonding with the nitrogen atoms of the successive amide bonds until maximum thermodynamic stability is achieved. The side chains of the amino acid residues lie outside the coordination plane and may be factors in inter- and intramolecular interactions, unless a Histidine residue, or a Lewis base is present. The amides of the protein bind to the copper at the square planar sites while the axial positions are thought to be occupied by hydroxide ions. This feature complicates the stoichiometry of the simple metal/protein complexation equilibrium, Reaction 16,



unless the pH of the system is maintained at a high level.

The stoichiometry of the Cu(II)-dipeptide complexes would only be an issue if the complexation constants were somehow needed in this study to determine analytical selectivity. The fact that they are all relatively similar increases the difficulty in finding spectroscopic differences between dipeptides. If the stoichiometry were to be different from one peptide to

another then more selectivity would be expected. Other analytical methods may require that the stoichiometry be rigorously explored. This method has no such prerequisite.

The CD activity of the chiral Cu(II) complexes is the result of disymmetric perturbations of the ground and excited state ligand field

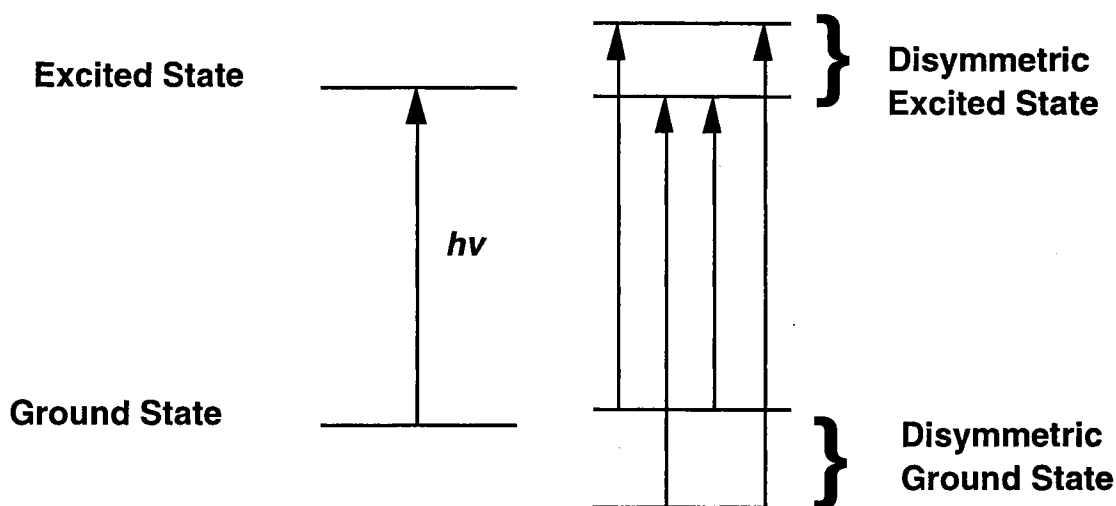


Figure 7.2: The chiral ligands can effect the ground state, the excited state or both. The splitting that occurs, results in unequal absorption of the left- and right- circularly polarized components.

orbitals by the chiral ligands, Figure 7.2. Bands in the UV range, which are a result of the chirality of the ligands whether they are bound or not, lack sensitivity to the coordinating metal ion. Though these bands are very intense, their lack of sensitivity coupled with a 10-fold increase in analysis time at those wavelengths are among the major reasons why these studies ignored the UV region for analysis.

*Visible CD spectra for Cu-complexes* : The CD data for the copper complexes with nine of the dipeptides as well as D-Histidine are shown in

Figure 7.3. The GG spectra coincides with the baseline and is not shown.

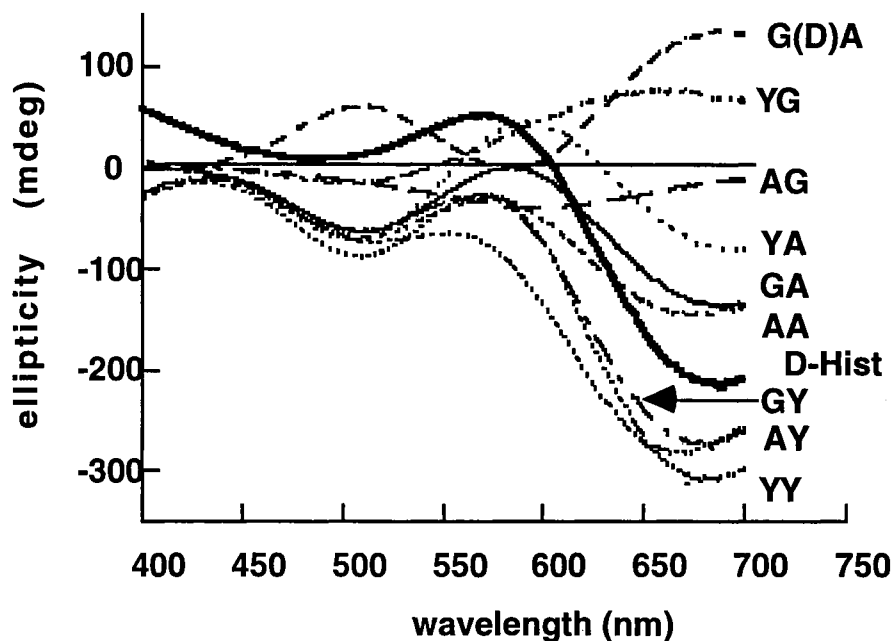


Figure 7.3: Visible CD spectra for the Cu(II) complexes of the nine chiral dipeptides. The two spectra most similar are for the GY and YY ligands, especially at wavelengths shorter than 625nm.

The spectral differences, as described in the caption, are sufficient for all of the peptides except the GY and the YY which have very similar spectra when the wavelength is less than 625 nm.

An initial survey of the spectral data reveals that there are variations of note between dipeptide pairs that have an inverse order, GA and AG, GY and YG, AY and YA. The residue whose nitrogen atom is directly involved in the ring closure step has the greatest affect on the magnitude of the spectral changes, Figure 7.4. G (glycine) is achiral but its relative position

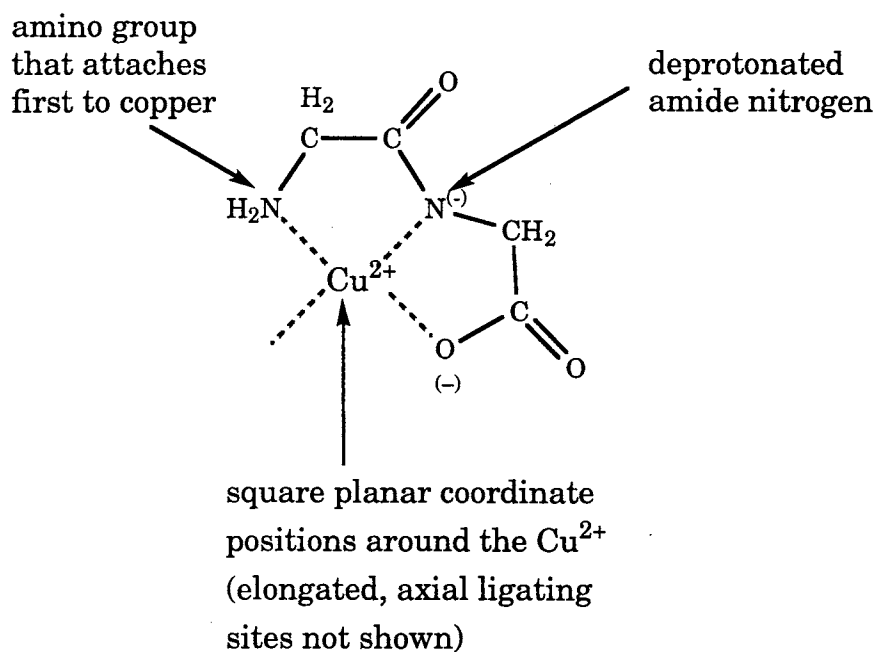


Figure 7.4: Bonding scheme of the dipeptide glycylglycine (GG) to the copper ion. The initial chelation through the amino nitrogen is followed by ring closure involving the deprotonated amide nitrogen which is completed by the binding of the carboxylic oxygen at the third position

in the sequence is important and affects the CD spectra quite dramatically.

The side chain of the second amino acid group can not be involved in the binding but is definitely a reason why spectral variations are observed.

*Factors that affect the selectivity* : The factors in solution that may contribute to the observed spectral differences include the stoichiometry of the complexes, relative stabilities of each complex, and the total analyte concentration. The analyte concentration is held constant and the intensities of the CD spectral data are not orders of magnitude different from one another suggesting that the equilibrium constants are similar.

The association constant was also verified using absorbance data, not shown, that showed very small absorption differences (<.05 AU) at the



maximum from one dipeptide to another. CD measurements were also made of each dipeptide with different ratios of the copper ion to find the spectral saturation point for each complex. Figure 7.5 shows an example of

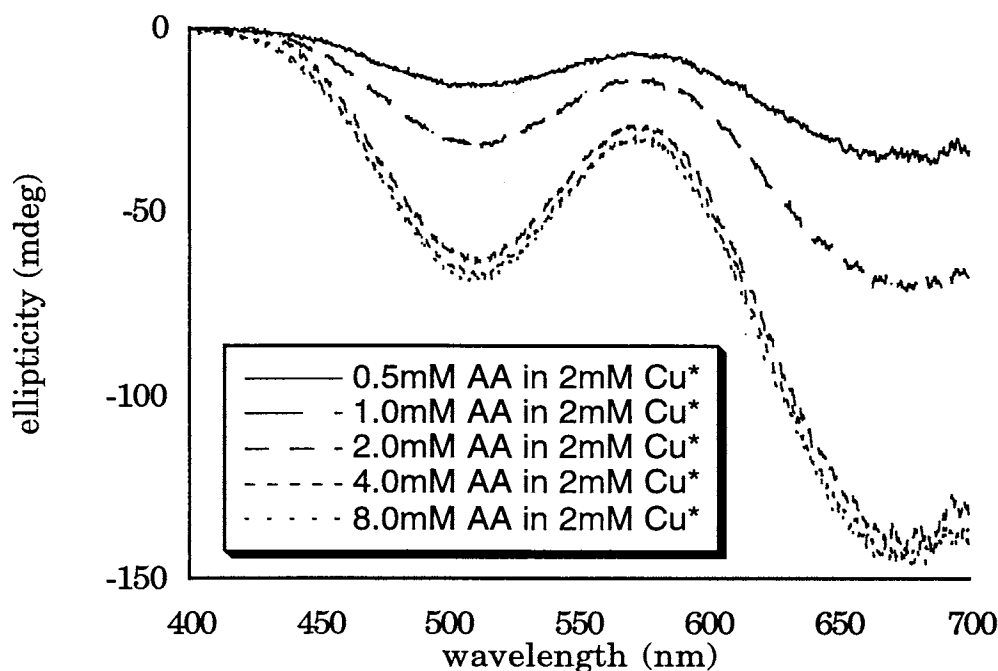


Figure 7.5: Dipeptide AA added to  $\text{Cu}^{2+}$  at different concentrations. The spectra doesn't change very much after the stoichiometric the ratio of 1:1 has been reached.

the type of data collected. The information from these experiments confirmed that at a dipeptide concentration of 8.0 mM most if not all the binding sites on the copper are filled. Visual inspection of the dipeptide complexes with copper at various concentrations indicated that 8.0 mM showed the greatest spectral differences between the spectra.

*Data Reduction Algorithms for Enhancing Sensitivity* : Data reduction for the dipeptide complexes were done by the method described in Chapter VI.

The use of the full 1500 data points prescribed by these methods allows

very subtle spectral changes to be distinguished and assigned a numerical value. This procedure relies heavily on the personal computer and data reduction software that enables in a few seconds what would take hours by hand. The graphical images produced by instruments of analysis have been important in understanding the relationship between the independent and dependent variables. The break down of that graphical data into information that can be correlated linearly with the ligand concentration has always depended upon a single independent value. The resulting analytical determinations from a single wavelength lack the accuracy which is gained when all the data points are used. Figure 7.6 shows the CD spectral data for the dipeptide GA when AG is added as an impurity at

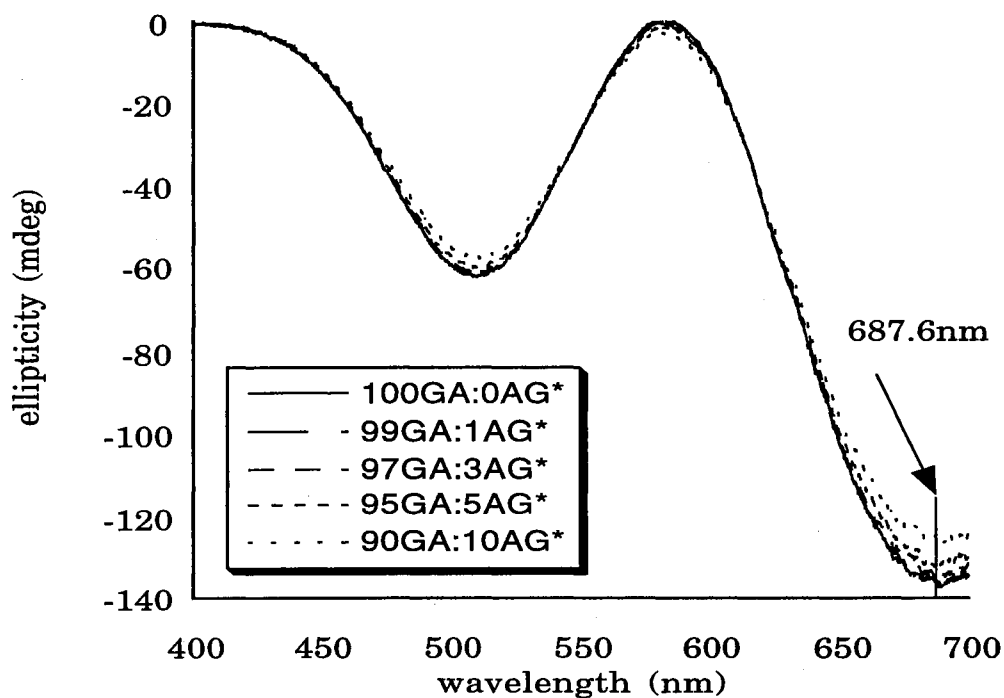


Figure 7.6: CD spectra of the GA complexed with AG added as an impurity at the levels of 1,3,5, and 10%. The ellipticities at the 687.6nm wavelength was chosen to compare with the full spectrum analysis algorithm.

levels of 1, 3, 5, and 10%. If a vertical line is placed on the plot where the ellipticity is the most sensitive to impurity concentration, the graph in Figure 7.7 can be made which shows ellipticity vs. concentration of the

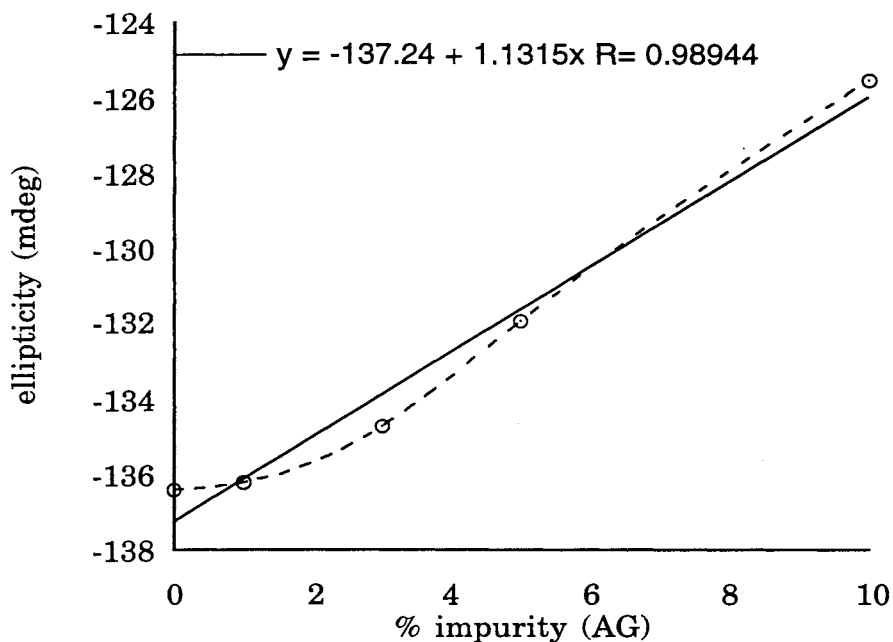


Figure 7.7: Plot of ellipticity vs %AG added to GA as an impurity. The solid line represents the best linear fit that can be made to the data while the dashed line shows the basic trend that the data is following.

impurity added. This graph shows deviations from the linear that may be due to a smaller signal to noise ratio at the higher wavelengths or is due to the very small signal change for the concentrations measured. The limit of detection between two CD signals at this wavelength range is about 2.0 mdeg. Without greater signal changes the percent impurity that is detectable with single wavelength analysis is limited to four or five percent at best.

*2-D Data Reduction Algorithm* : The first step taken to simplify the data is to plot the CD spectral data for the dipeptide GA against itself. This can be

done to check the reproducibility of the GA complex from day to day by monitoring the slope and correlation coefficient for deviations from unity.

CD data for the dipeptides G(D)A, GG, AG, GA, and AA was plotted against the CD data for the GA complex in Figure 7.8.

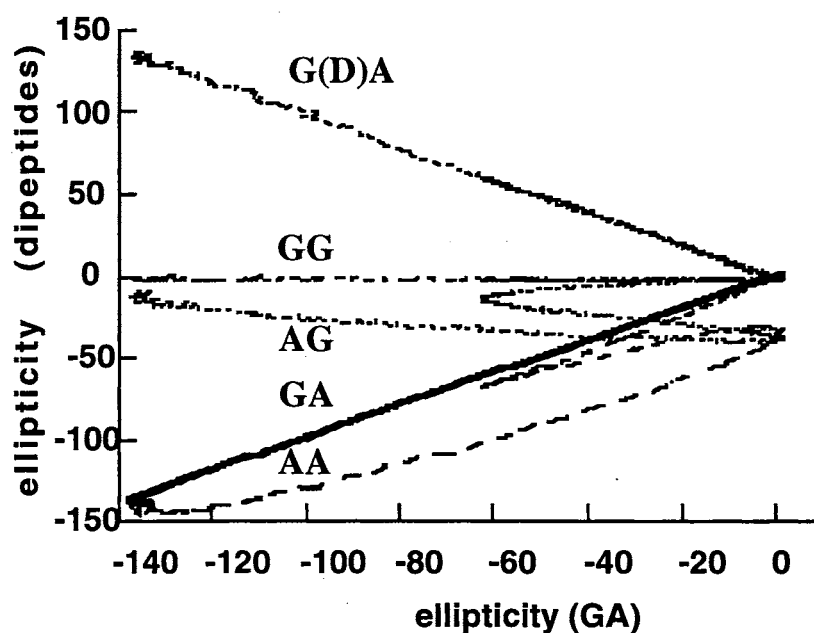


Figure 7.8: Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for the analogous complexes with G(D)A, GG, AG, GA, and AA.

Dipeptides that contain only Glycine and Alanine show large differences when plotted against the ellipticity data for GA. Figure 7.9 is a plot of the ellipticities for the dipeptides YG, YA, GA, GY, YY, and AY against the GA complex.

In Figure 7.9 the plot of GA versus itself divides the dipeptides into two subgroups, those that contain the amino acid tyrosine (Y) at the amine end and those that have tyrosine (Y) at the carboxylic terminus.

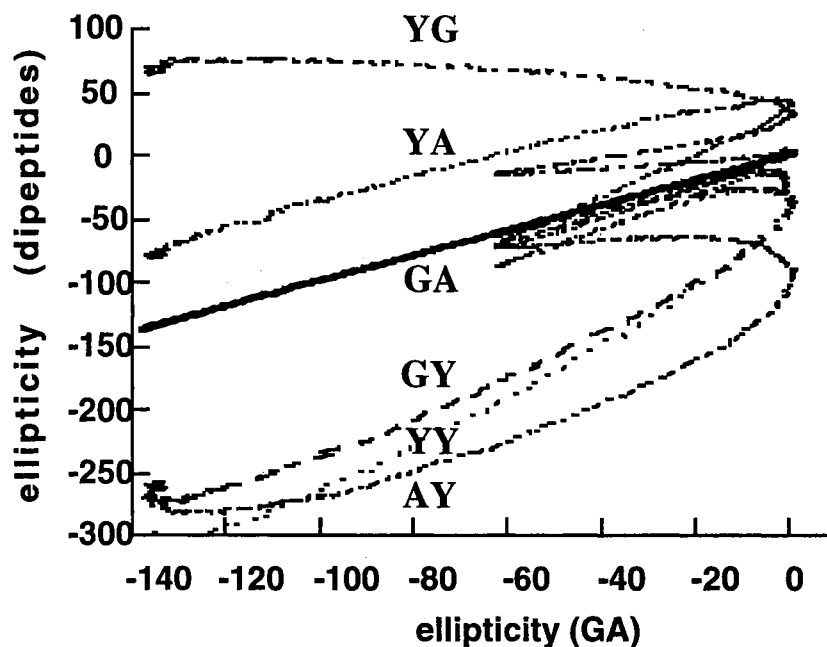


Figure 7.9: Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for the analogous complexes of YG, YA, GA, GY, YY, and AY.

(a) *Enantiomeric Mixtures of GA + G(D)A*. The wavelength chosen for the single wavelength analysis is shown in Figure 7.10. Using the 2-D method of data reduction the plots of G(D)A versus GA are shown in Figure 7.11. The data was the same for both, the only difference being that the method of data reduction.

If both enantiomers are of equivalent purity then the slopes of line (a) and (h) will be  $\pm 1.0$  respectively. For the correlation plot of GA vs GA to have a slope of +1.0 means that at every wavelength the ellipticity of the GA stock is the same from one measurement to another. This will always be true unless a human error is made, the complexation changes over time, or an impurity exists from one GA source to another. For the correlation plot of

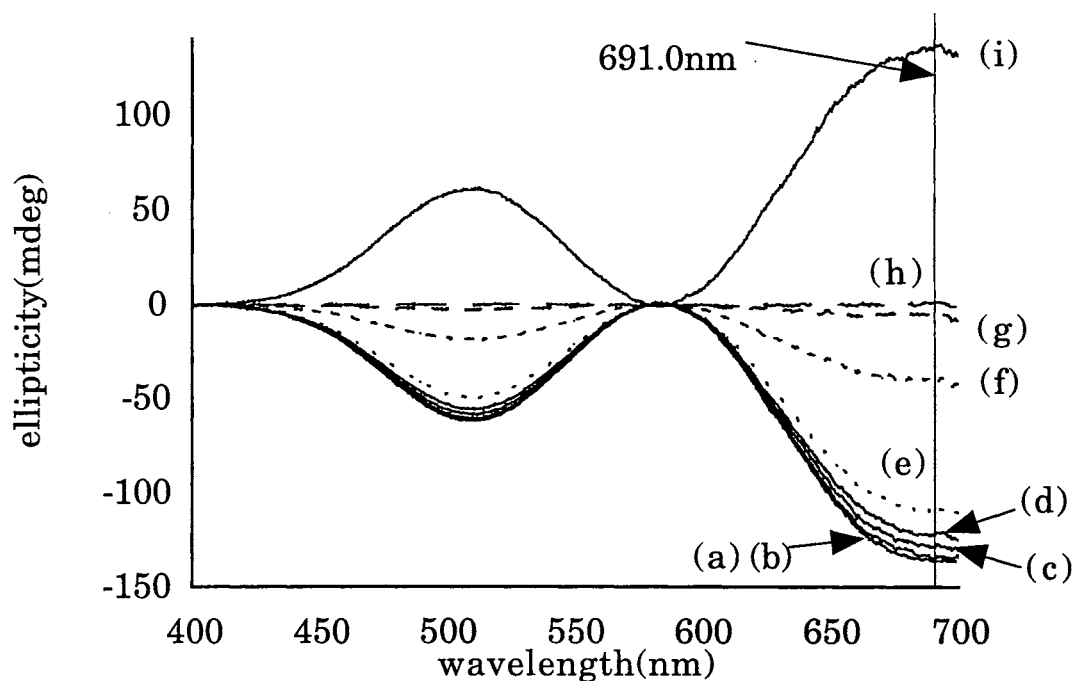


Figure 7.10: CD spectra of the GA with G(D)A added as an impurity;(a) 100% GA; (b) 99%; (c) 97%; (d) 95%; (e) 90%; (f) 65%; (g) 52%; (h) 50%;and (i) 0%GA 100% G(D)A.

GA versus G(D)A to have a slope of -1.0 it is required that the ellipticity of the GA stock is the opposite in sign and equal in magnitude to the ellipticity of the G(D)A stock at every wavelength. For (h) the slope of the line is -0.9989, which, when placed in the enantiomeric excess equation 17 given below, equates to an impurity of approximately 0.05%.

$$\text{Slope} = \frac{[\text{Cu}^{2+}(\text{GA})] - [\text{Cu}^{2+}(\text{GA})_{1-x}(\text{G(D)A})_x]}{[\text{Cu}^{2+}(\text{GA})] + [\text{Cu}^{2+}(\text{GA})_{1-x}(\text{G(D)A})_x]} \quad (17)$$

This equation alone is not enough to solve for the %GA in each mixture, equation 18 is needed as well, which simply states that the total concentration of each sample must remain constant, 8.0mM for this case.

$$8.0\text{mM} = [\text{Cu}^{2+}(\text{GA})] + [\text{Cu}^{2+}(\text{G(D)A})] \quad (18)$$

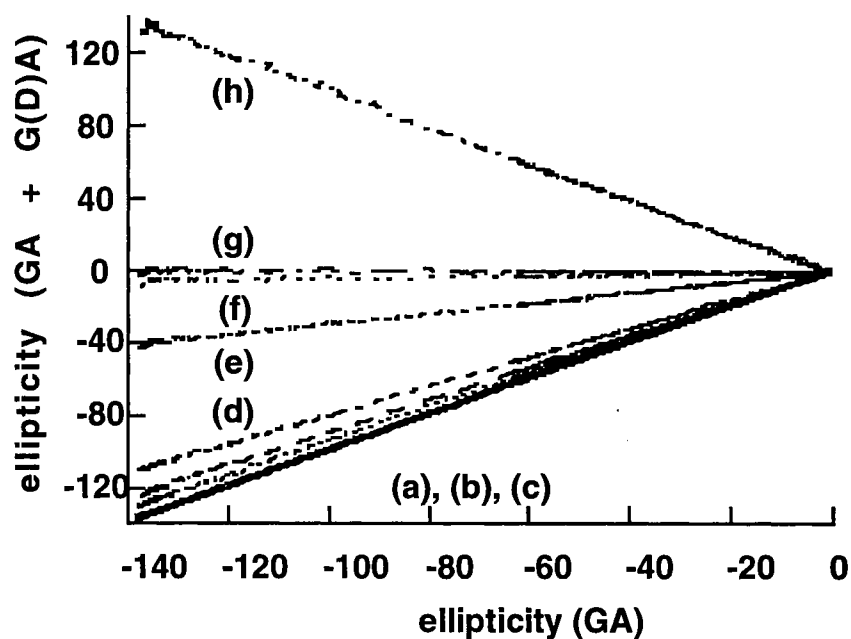


Figure 7.11: Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for the enantiomeric mixtures of percent compositions: (a) 100% GA; (b) 97; (c) 95; (d) 90; (e) 65; (f) 52; (g) 50; and (h) 0% GA or 100% G(D)A.

Using the correlation slopes from Figure 7.11 and equation 17, percent purities for GA were calculated for the prepared mixtures, Table 7.2. Standard deviations are less than 0.25%. When these numbers are compared to the results where data reduction was done on the CD data measured at one wavelength the SD has been improved by almost a factor of ten. Comparisons with the single wavelength method finds that the SD is closer to 1.9% for impurity values 10% or less. The 2-D algorithm has effectively reduced the 1500 spectral data points to one number (the correlation slope), from which enantiomeric purity can be determined with

better accuracy.

Chemical impurities can be recognized using the 2-D algorithm but the quantitation of a chemical impurity is more easily done with the 3-D algorithm.

**Table 7.2: Determination of Enantiomeric Purities for Prepared Binary Mixtures of GA and G(D)A.**

%GA in solution	Regression Slope	Regression Coefficient	%GA Calculated
100	1.0	1.0	100
99.0	0.9837	0.99994	99.05±0.15
97.0	0.9438	0.99995	97.01±0.19
95.0	0.8998	0.99994	95.03±0.03
90.0	0.8016	0.99990	90.05±0.04
65.0	0.2975	0.9995	64.96±0.08
52.0	0.0400	0.96388	52.13±0.13
50.0	0.0040	0.3391	50.34±0.17
0	-0.9989	1.0	0.05±0.05

(b) GA + "Chemical Impurity" levels for all other dipeptides. Among the several ways that impurities might occur while preparing a peptide for pharmaceutical purposes are the following. First an enantiomeric impurity might be produced or racemization might occur after production which will change the chirality of the stereogenic center. Second is the production of a completely different chemical substance. Synthesis of the peptide by sequentially linking each amino acid residue together in a chain there can be accidental reversal of the sequence or perhaps improper sequences due to the monomer reacting with itself rather than with the



target.

The question with respect to “chemical impurities” is not how great the absolute difference is between the CD spectra when both compounds are host molecules but rather is the difference enough to quantitate the “chemical impurity” when the impurity amounts to only a few percent?

Rationalizations based on the potential needs a pharmaceutical manufacturer led to the following line of thinking.

In the hypothetical large scale manufacture of the “drug” GA, “chemical impurities” might be G- and A- monomers, and GG, AG, and AA dimers. Monomers are not thermodynamically competitive with the dimers when binding to copper so they would not contribute to the visible spectrum but the mass difference that they would cause would result in a correlation slope less than 1.0, indicating that there is some impurity. The correlation slope would still be straight and so monomeric impurity might be mistaken for the achiral GG or enantiomeric G(D)A impurity.

Low amounts of dipeptide impurities will cause the linear 2-D GA vs GA plot to be split. Observing these deviations can easily be done using the 2-D algorithm but being able to mathematically quantitate the amount and type of chemical impurity is better done with the 3-D algorithm and Spinning Plots<sup>®</sup>.

*3-D Data Reduction Algorithm:* In the 2-D plots shown in Figures 7-8 and Figures 7-9 wavelength is an implied variable. For the generation of a 3-D plot the wavelength is used as the third dimension. CD is the measurement of the difference in absorption between the left- and right- components so

there are positive ellipticities as well as negative. When CD spectra are plotted against one another there exists four possible sign combinations at every wavelength,  $+/+$ ,  $+/-$ ,  $-/+$ , and  $-/-$ . Points that share the same ellipticity values may have wavelengths that are not adjacent to one another in the spectra. This causes the 3-D plots to have three dimensional structure that the 2-D plots lack. The 2-D plot of the CD spectra in reality is the projection of the 3-D plot on the x-y coordinate plane. The real benefit of adding the wavelength was to have an increase in analytical selectivity as the spectra for different dimers are compared. Figure 7.12 (A) and (B) show the visual differences between plotting the AG and GA in the 3-D format where the wavelength and the data for the GA complex are the x and y coordinates

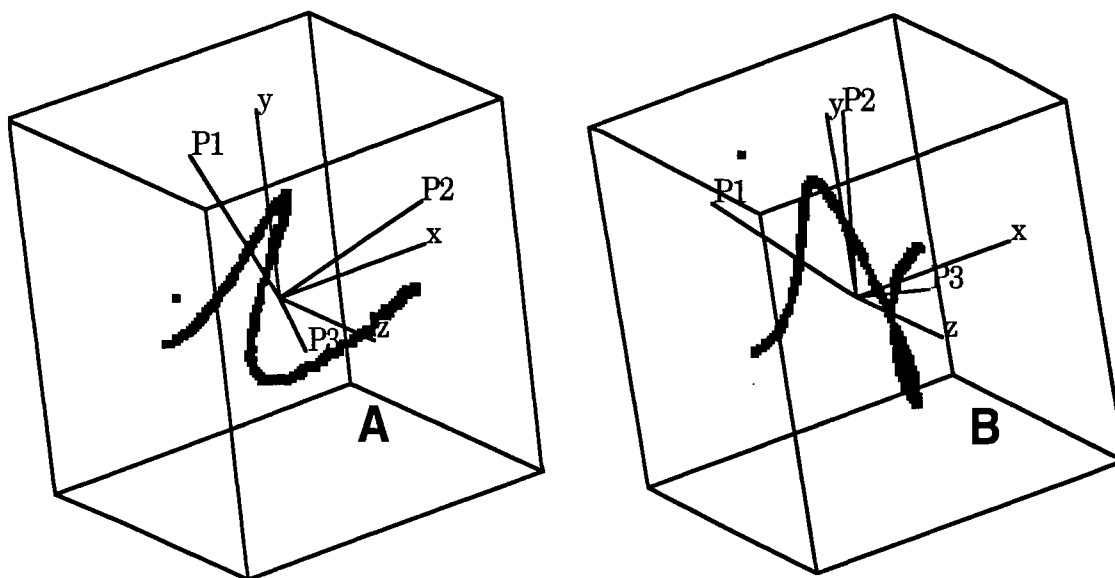


Figure 7.12: Spin Plots for the presentation of wavelength (x-coordinate); spectral data for the GA complex (y-coordinate); and spectral data for (A) the AG, and (B) GA complexes. The lines P1, P2, and P3 are the principal component axes.

respectively. Figure 7.12 (B) shows the 3-D baseline of the GA complex plotted against itself.

*Factor Analysis of Spinning Plot<sup>®</sup> Data* : Data reduction was done on these 3-D plots using the Principal Component Analysis (PCA) algorithm. Table 7.3 shows the eigenvalues and eigenvectors for the three principal components, PC1, PC2, and PC3 calculated for the GA/AG combination plot shown in Figure 7.12 (A). The CD spectra which showed the greatest similarities were the GY and the YY. Using the PC23 values generated from each dipeptide complex in a 3-D spin plot, even these two can be distinguished.

**Table 7.3:**  
**Principal Component Values Calculated for the GA versus AG System:**

		PC1	PC2	PC3
<b>Eigenvalues</b>		+1.9681	+1.0008	+0.0311
<b>Eigenvectors</b>	<b>nm</b>	-0.70725	-0.00007	+0.70697
	<b>GA</b>	+0.52044	-0.67686	+0.52058
	<b>AG</b>	+0.47848	<b>+0.73611</b>	+0.47874

Table 7.4 shows the collection of the PC23 values for all the dimer complexes.

**Table 7.4: Corresponding PC23 Values for GA Plotted Against Spectral Data for the Dipeptides in 8.0mM Solution.**

AA	AG	GG	YA	GY	YG	AY	YY	G(D)A
0.1064	0.7361	0.9992	0.6906	0.0983	0.5334	-0.18498	0.1720	-0.36786

There is also a very clear distinction between the correlations for the achiral

GG and enantiomeric G(D)A peptides. Standard deviations calculated for spectral data from 3-5 independent repeat measurements are  $\pm 0.002$ , substantiating the earlier claim that differentiations can be arrived at by a simple mathematical inspection of the 3-D plots. The small difference between the magnitude of the PC23 values for GA and G(D)A is consistent with the small difference in the slopes of the correlation lines in the 2-D treatment, implying that the enantiomeric purity for these are not exactly the same. The 3-D algorithm has reduced 1500 spectral data points to one number, PC23, with which potential chemical impurities can be qualitatively identified.

Quantitative determination of the amount of a "chemical impurity", that is on the order of 1-10% of the amount of GA, followed quite simply from the PC2 determination because the PC23 value is linearly correlated with the percent of impurity. Figure 7.13 is the plot of PC23 values for the different dimers used as impurities in the GA vs percent composition. With the exception of GG and G(D)A, correlation slopes give analytical sensitivities that are at least ten times more accurate than analogous plots of the ellipticity values measured at a maximum wavelength plotted against concentration. Referring back to Figure 7.6 and Figure 7.7 where CD spectral data are plotted for GA with AG as an added impurity it can be seen that, since the best possible resolution for the CD instrumentation used in this study is  $\pm 2.0$  mdeg, there are clearly cases among these analytes where the S/N is of insufficient quality to obtain a determination at all.

The slope of the PC23 versus percent impurity line for YA in

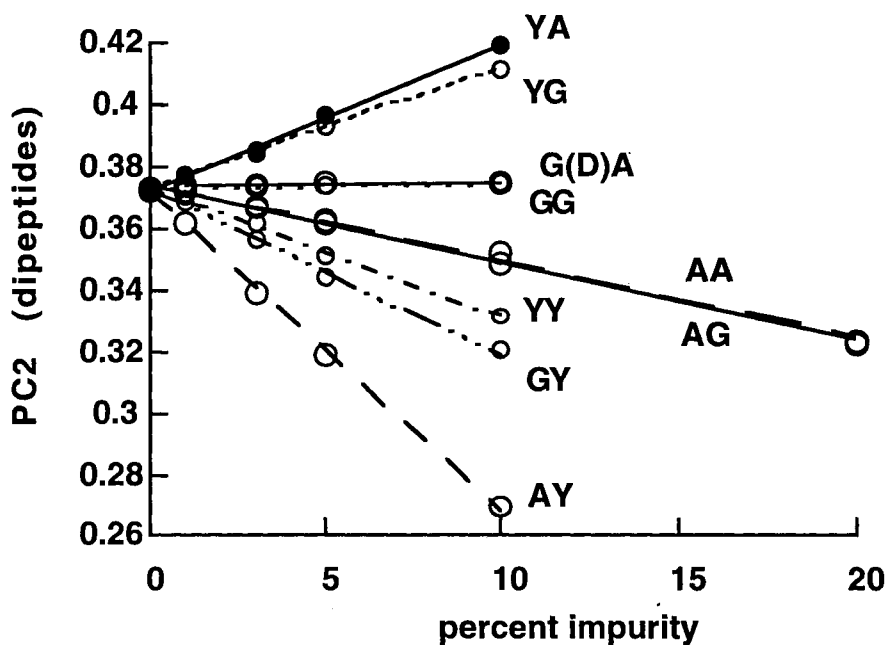


Figure 7.13: Linear correlation plots of PC2 values versus the percent "chemical impurity" for all nine dipeptides added to GA.

Figure 7.13, which is approximately the mean value for all of the slopes, is more than two times the  $\pm 0.002$  SD in the means for PC23 values, which implies that impurity levels as little as 1-3% can be measured with confidence.

While the 2-D method succeeded in providing accurate values for enantiomeric purity measurements, the 3-D method provides for a quantitative measure of the non-enantiomeric chiral impurities.

### Summary

Using a modified biuret method that employs chiral sensitivity with

two novel data reduction algorithms for the handling of the CD data, a potentially effective quality control procedure for peptides, oligopeptides, and proteins has been developed. This method does not suffer, as other detection protocols do, from increased error when the impurity level drops below 10%.

A typical procedure begins with the measurement of the visible CD spectrum for the Cu(II) complex of the purest available form of the substance being regulated. Data are archived in a PC file on-board the spectrometer and are updated each time the reference material is measured. Spectra for aliquots taken from each newly manufactured product lot are plotted against the standard and successive on-screen visual comparisons are made.

Deviation from a slope of 1.0 in the 2-D test is an instant indication that the purity is less than that of the reference standard. If the 2-D linear regression coefficient indicates an enantiomeric "impurity", the enantiomeric purity is calculated from the regression slope. Separation in the correlation line from the standard reference material gives instant recognition that a "chemical impurity" is present whose identity is confirmed by the PC23 value calculated from the Spinning Plot<sup>®</sup> algorithm. The percent impurity is calculated from the correlation slope of the PC23 versus impurity line.

This method could be an alternative to chromatographic or mass spectrometry methods for the quality control of small peptides.

CHAPTER EIGHT  
 TRIPEPTIDE DISCRIMINATION USING  
 CIRCULAR DICHROISM DETECTION

**Introduction**

While very similar to the situation discussed in the previous chapter, tripeptides offer some new challenges. The complex formed between the copper ion and a dipeptide may or may not have a stoichiometric ratio of 1:1. An example was shown (Figure 7.5) where the ratio of Cu(II) and the dipeptide GA was changed and for concentrations above a 1:1 ratio with the copper ion there was no appreciable CD signal change. That could not

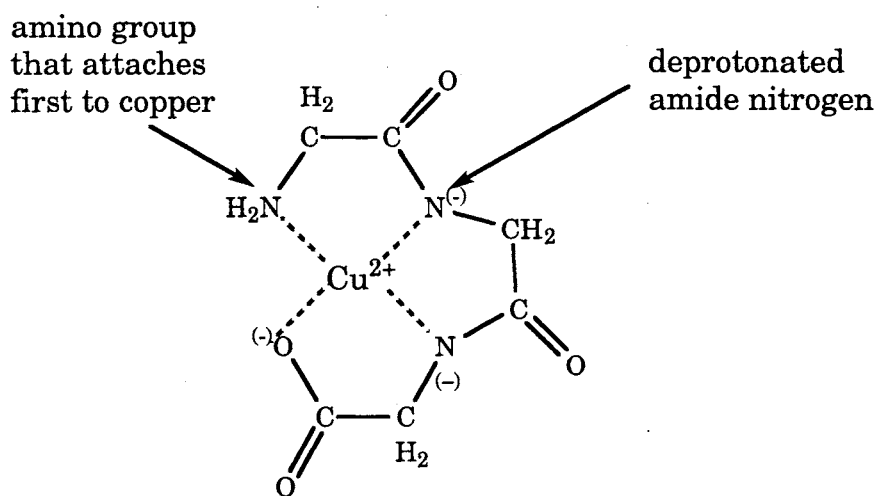


Figure 8.1: Bonding scheme of the tripeptide glycylglycylglycine (GGG) to the copper ion. All four of the square planar binding sites are occupied by a three residue peptide. It is for this reason that it is thought that the first three residues contribute the most to the shape of the CD signal.

be assumed to be the case for all dipeptides and it was stressed that any differences in complexation between the dipeptides would only lead to greater selectivity. Tripeptides are long enough and have enough binding sites in the appropriate places on the peptide backbone to saturate the tetragonal binding sites on the copper ion (Figure 8.1). With all the potential binding sites filled but without any remaining residues left to allow possible chiral enhancement due to intra- and inter- molecular associations the tripeptides may offer the biggest challenge in characterization.

Eight tripeptides were chosen for the study. Common to all eight are two glycine residues which occupy positions 1, 2-, 1, 3-, and 2, 3- in the sequence. The remaining residues are L-enantiomers of aliphatic and aromatic aminoacids. Again, the tripeptides have no stable ternary structure to speak of, so the variability of the sequence is really the only parameter affecting the chiral response of the CD detector. The order of residues in short peptides is crucial to the extension of the study to peptides and proteins as a whole because coordination of the latter to Cu(II) involves the first three aminoacids from the amine-terminus. If there is no CD selectivity associated with changes in the initial sequence, the method has no value in the study of oligopeptides and proteins where too frequently the same initial sequence is common to several potential analytes.

The experimental procedure is a combination of the methods that were used to discriminate among the dipeptides in Chapter VII and the insulins in Chapter VI and the method used to measure enantiomeric



purities for glycyl-L-alanine in chapter seven. Data reduction and spectral differentiations are done using variations on standardized mathematical algorithms and PCA.

## **Experimental**

*Chemicals:* Tripeptides used in the study were glycyglycyl-L-alanine (GGA), glycyglycyl-L-isoleucine (GGI), glycyglycyl-L-leucine (GGL), glycyglycyl-L-phenylalanine (GGP), glycyl-L-histidylglycine (GHG) and its D-enantiomer (GhG), L-leucyl-glycyglycine (LGG) and its D-enantiomer (lGG), and L-tyrosylglycyglycine (YGG) and its D-enantiomer (yGG). All eight L-enantiomers were supplied by Sigma Chemical Co. which reported an enantiomeric purity in excess of 99.8%.

The D-enantiomers GhG, lGG, and yGG of GHG, LGG, and YGG were prepared by Multiple Peptide Systems (MPS), San Diego. Certificates of Analysis described them as unpurified off white powders. Percent purities as determined by RP-HPLC analyses were reported as 86.57, 99.10, and 97.86 respectively. The low value for GhG is explained as being due to two elution peaks that correspond to the same compound. The percentage is based on the relative area of the first peak which corresponds with most of the material eluting with the void volume peak. The second peak is related to the hydrophobicity of the molecule that causes it to stick to the column to be eluted later. D-Histidine was also a Sigma Chemical Co. product with an enantiomeric purity reported at better than 99.8%. Reagent grade  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  was obtained from Fisher Scientific.

*Solutions Preparation:* Solutions for this set of experiments were made in the same manner as those discussed in Chapter VII. A modified biuret reagent was made that contained [2.0 mM]  $\text{Cu}^{2+}$  and [8.0 mM] of the tripeptide or tripeptide mixture. Also included in the reagent is  $[\text{OH}^-]$  at 0.10 M and  $[\text{I}^-]$  at 0.030 M added as a stabilizer.

Quantitative tests were done on two kinds of mixtures. For the first kind, GGA was arbitrarily selected as an enantiomerically pure “reference” material. Aliquots from the stock were spiked with “chemical impurities” or small volume aliquots of the other L-tripeptide stocks to cover the impurity range from 1-10%, prior to dilution with the hydroxide. For the second, “enantiomeric purity” tests, aliquots of YGG, LGG, and GHG stocks were spiked with smaller volume aliquots of the corresponding D-enantiomer stocks, yGG, lGG, and GhG over the same 1-10% impurity range.

*Measurements:* CD spectra were measured using the same instrumentation discussed in the earlier chapters and the same instrumental setup parameters (See Chapter VI *measurements* section).

## **Results and Discussion**

*Visible CD spectra for Cu(II)-tripeptide complexes:* Spectra for all eight copper-L-tripeptide complexes, in which  $[\text{Cu(II)}] = 2.0 \text{ mM}$  and the ligand concentrations is 8.0 mM, are shown in Figure 8.2. Only GGH, GHG, LGG, and YGG are uniquely differentiable by their zero order CD spectra.

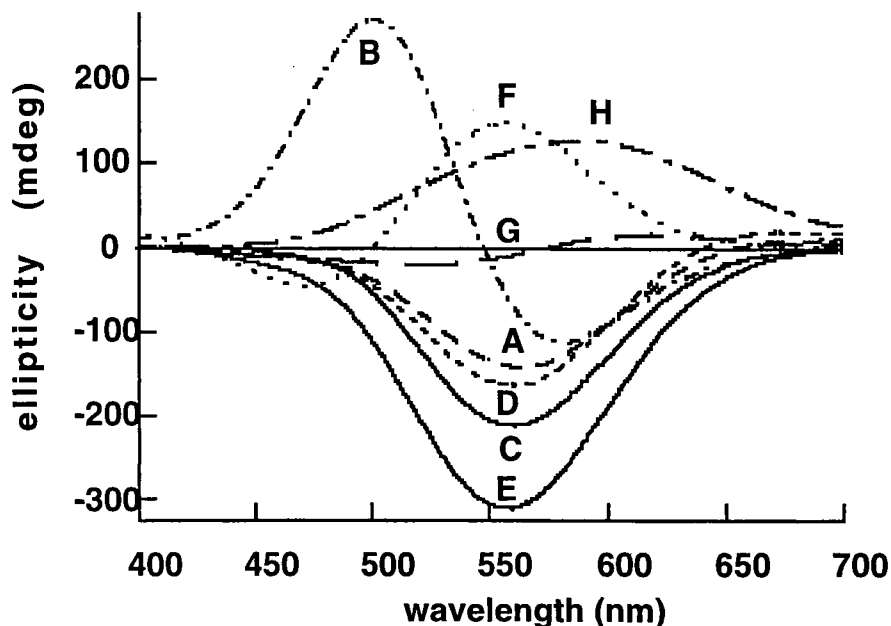


Figure 8.2: Visible CD spectra for the Cu(II) complexes of: (A) GGA; (B) GGH; (C) GGI; (D) GGL; (E) GGP; (F) GHG; (G) LGG; and (H) YGG. Similarities are greatest for the GGA, GGI, GGL, and GGP complexes over the entire wavelength range.

Spectra for the histidyl-containing ligand complexes, GGH and GHG, are blue shifted compared with the Cu(II)-D-Histidine complex itself which has an intense negative band with a maximum at 689 nm and a weaker positive maximum at 570 nm. The magnitude of the shift is greatly dependent upon the position occupied by the histidyl residue. The sensitivity of the CD spectral response to the histidine position is a significant first result in the context of possibly sequencing short peptides by this spectroscopic method.

Of the five GGX peptides, only the spectrum for the GGH is unique, which might be attributable to the special involvement of the imidazole N-atom in binding to Cu(II). The remaining four have but one broad negative band that maximizes around 550 nm. Aromaticity in the side chain may

(YGG) or may not (GGP) induce a spectral change, which with further developments, might be exploited for short range sequencing. There is ambiguity in differentiating among GGA, GGI, GGL, and GGP unless the solution concentrations are carefully controlled.

The L-leucine structural isomers can be qualitatively differentiated in a quality control context. The lack of band intensity for the LGG is a potential problem in quantitation. Although the glycy residue is achiral, the relative positions that it occupies affect the CD spectra quite dramatically.

*2-D Data Reduction Algorithm For Enhancing Selectivity:* GGA is arbitrarily assigned the status of an enantiomerically pure standard reference material. In a pharmaceutical context, GGA might represent a commercial drug product. The others fill the roles of potential “chemical” and “enantiomeric” impurities.

The concept used in previous chapters is used again where a plot is made of the 1500 data points for the 8.0 mM GGA spectrum (on the x-axis) against analogous data for 8.0 mM solutions for each of the others (on the y-axis). To get a baseline reference check for the absolute enantiomeric purity of GGA, its CD spectrum is also plotted on both axes. The correlation is a straight line of unit slope and zero intercept. Spectra for the remaining seven tripeptides complexes are plotted against GGA in Figure 8.3 for the GGX subseries and in Figure 8.4 for the other sequences.

Plots are non-linear and individually distinct from one another. The ellipsoidal shapes for GGI and GGL might appear similar but with

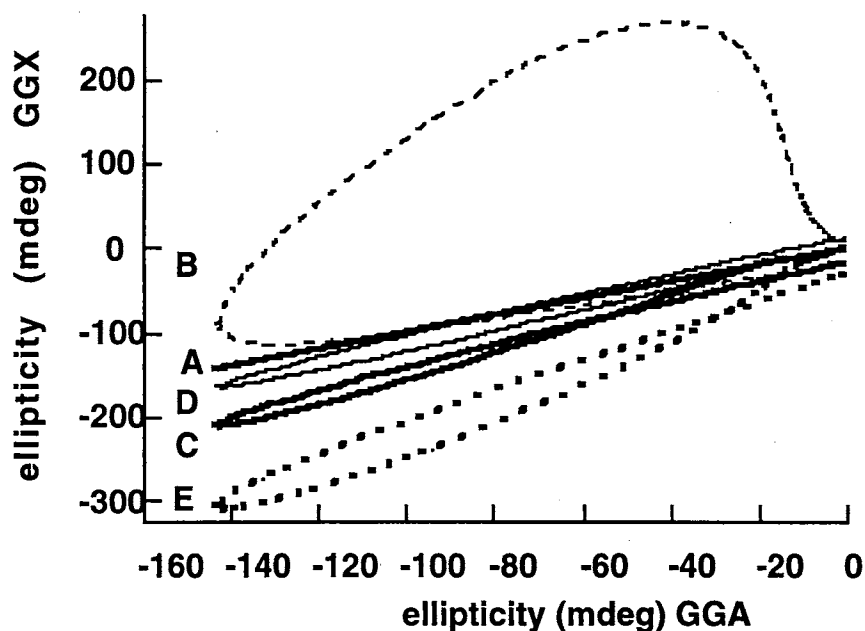


Figure 8.3: Correlation plots of ellipticity for the Cu(II)-GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of (A) GGA; (B) GGH; (C) GGI; (D) GGL; and (E) GGP.

different virtual slopes. On further examination the ellipse for the GGI is seen to “fold over” on itself in a partial figure 8 implying a latent 3-dimensional property in these plots. The same phenomenon can be seen more clearly for the plot of the GHG analog vs. GGA in Figure 8.4. Differentiation among enantiomerically pure forms of the tripeptides has apparently been achieved at least when the number of possibilities is limited to a small closed set, as seen here.

The only other possible correlation line where the slope is unity (but opposite in sign) and the intercept is zero is for the plot of GGA versus the D-enantiomer, GGa, but only if their purities are equivalent. This is a

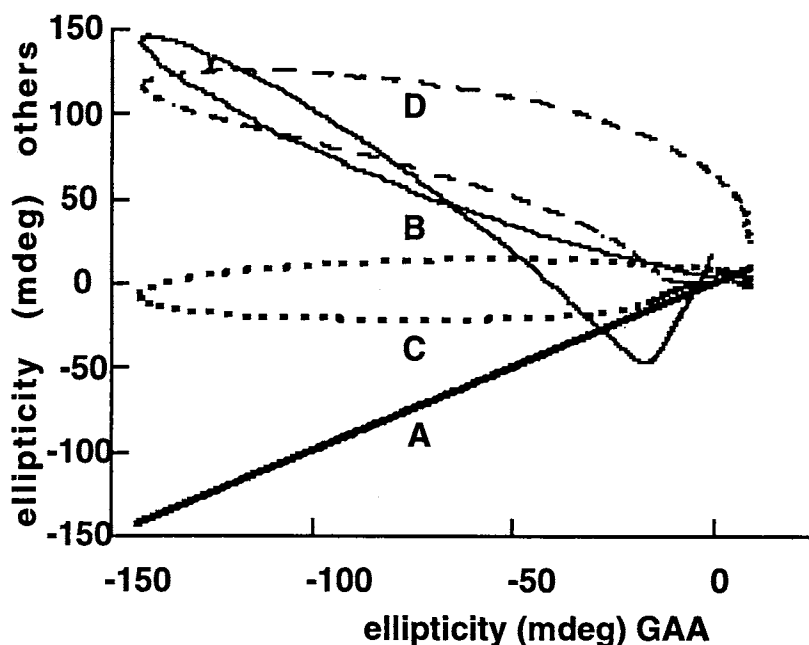


Figure 8.4: Correlation plots of ellipticity for the Cu(II)-GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of (A) GGA; (B) GGH; (C) LGG; and (D) YGG.

consequence of their being chemically identical. The feature common to all cases where spectra for chemically dissimilar compounds are correlated is splitting of the correlation line compared with the ideal reference line.

Some splittings are extreme, as can be seen in Figure 8.3 and Figure 8.4.

Conversely, if the correlation plot of the CD spectrum for a newly manufactured lot of GGA versus the reference is linear with a slope less than one, and shows no evidence of splitting, that is evidence for the presence of the enantiomer or an achiral impurity. Splitting is instant evidence for the presence of a chemical impurity. Linear regression would not give accurate information if applied to these correlation plots simply

because they are not linear. Quantitation of these “chemical impurities” require PCA.

*Quantitation of Enantiomeric Mixtures:* Enantiomeric purity tests were made on three analyte pairs, GHG/GhG, LGG/lGG, YGG/yGG. As the D-enantiomers were added in increasing amounts, over the range 1, 3, 5, and 10% of the L-enantiomer concentration, the slopes of the correlation lines decreased.

Judging by the regression coefficient of 0.9998 for the GHG versus GhG plot, there is no significant loss of linearity compared to the reference baseline, meaning that the enantiomeric purity of GhG is equivalent to that of GHG. The explanation given in the Experimental Section for the low percent purity for GhG, as described in the MPS Certificate of Analysis, is apparently vindicated by the results of this spectroscopic method.

Splitting of the YGG/yGG correlation line is consistent with the MPS reported purity level of 97.86% or total impurity of 2.14%.

Noise on the LGG/lGG correlation line conceals whether there is splitting of the line or not. A poor S/N ratio is expected since the CD spectral intensity for LGG is the weakest, as observed in Figure 8.2, being approximately a tenth of the band intensities for the other tripeptides.

Equations 17 and 18 in Chapter VII were used again to determine the percent composition of one enantiomer present in an enantiomeric mixture. Finding the slope of the correlation mixture and knowing the total amount of analyte added will allow one to solve these equations to find the percent impurities. Calculated values for spiked GHG solutions are in

excellent agreement with the measured values for prepared mixtures, Table 8.1.

Imprecisions based on data from 3-5 repeat measurements, are an improvement by almost a factor of ten over results obtained from the analysis of binary ephedrine mixtures in which a chemometric analysis method, Partial Least Squares, was applied to data at five wavelengths (36).

**Table 8.1: Determination of Enantiomeric Purities for Prepared Binary Mixtures of GHG/GhG; LGG/IGG; and YGG/yGG.**

<b>% L-form in Prepared Solution</b>	<b>Regression Slope (Enantiomeric Excess)</b>	<b>Regression Coefficient</b>
<b>GHG/GhG</b>		
99.0	0.9897	0.9998
97.0	0.9692	0.9998
95.0	0.9501	0.9951
90.0	0.8983	0.9999
<b>LGG/IGG</b>		
99.0	0.9974	0.9951
97.0	0.9723	0.9957
95.0	0.9455	0.9946
90.0	0.8819	0.9932
<b>YGG/yGG</b>		
99.0	0.9893	0.9996
97.0	0.9671	0.9996
95.0	0.9380	0.9996
90.0	0.8858	0.9995

In spite of the splitting of the YGG/yGG and the noise in the LGG/IGG best-fit correlations, agreements between calculated and measured



enantiomeric purities are still much better than previous methods.

*GGA + "Chemical Impurity" Levels for all Other Tripeptides:* Again the question to address with respect to "chemical impurities" being detectable in the 8.0 mM solution of GGA is not whether they are detectable at equimolar concentrations but whether they are detectable and measurable when these amount to only a few percent. The answer to the question lies in how sensitive the CD detector is in discovering splitting of the correlation line when spectra for "impure" samples are plotted against the spectrum for the primary reference standard.

Spectra were measured for mixtures in which GGA solutions were spiked with small volumes of the other L-tripeptides at levels of 1, 3, 5, and 10%. Splittings, where they are observed, may be accompanied by a change in the virtual slope compared to the reference line. Because of the absence of splitting, the plots fail to confirm the presence of GGI, GGL, or GGP at a level of 5%, see Figure 8.5.

The greater the splitting the more we can be assured that a chemical impurity exists but the less sure the quantity of the impurity can be measured using the slopes of the correlation plots.

Again it was shown that the 2-D algorithm effectively reduced the 1500 spectral data points to one number (the correlation slope), from which enantiomeric purities can be determined with excellent accuracy. Recognition of a potential "chemical impurity" is elementary for a limited number of cases but its analytical determination is not easily done using the 2-D method.

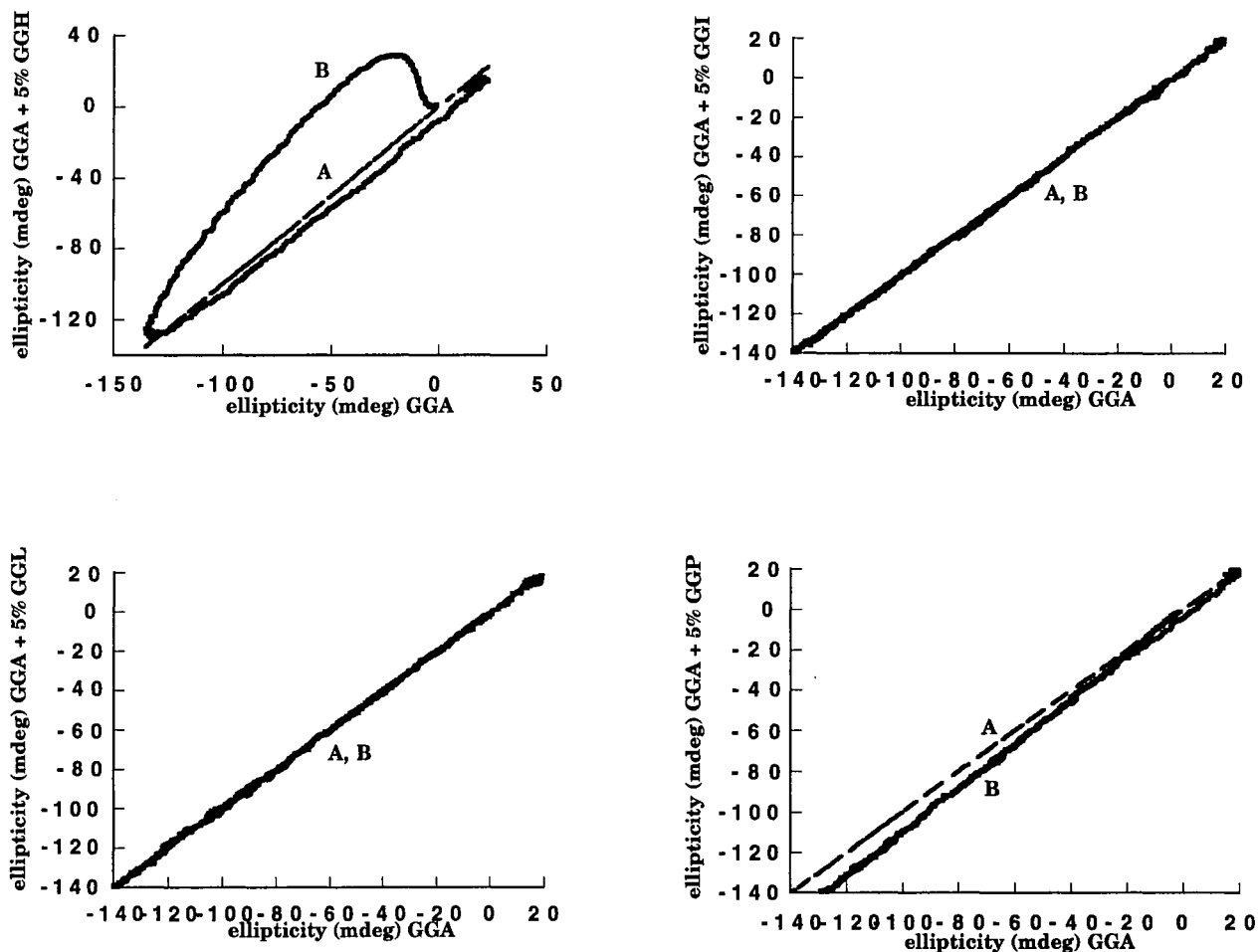


Figure 8.5: Correlation plots of ellipticity for the Cu(II)-GGA complex versus ellipticities for 5% chemical mixtures with GGH, GGI, GGL, and GGP. In each case line A is for GGA against itself and line B is for GGA against the mixture.

*The 3-D Data Reduction Algorithm for Enhancing Selectivity:* The objectives that relate to this second data reduction algorithm were to discover if the GGA, GGI, GGL, and GGP series can be completely differentiated both qualitatively and quantitatively. The objectives were achieved when the

same algorithm was applied to the set of dipeptides in Chapter VII.

Using the same Spin Plots<sup>®</sup> procedure where the wavelength is

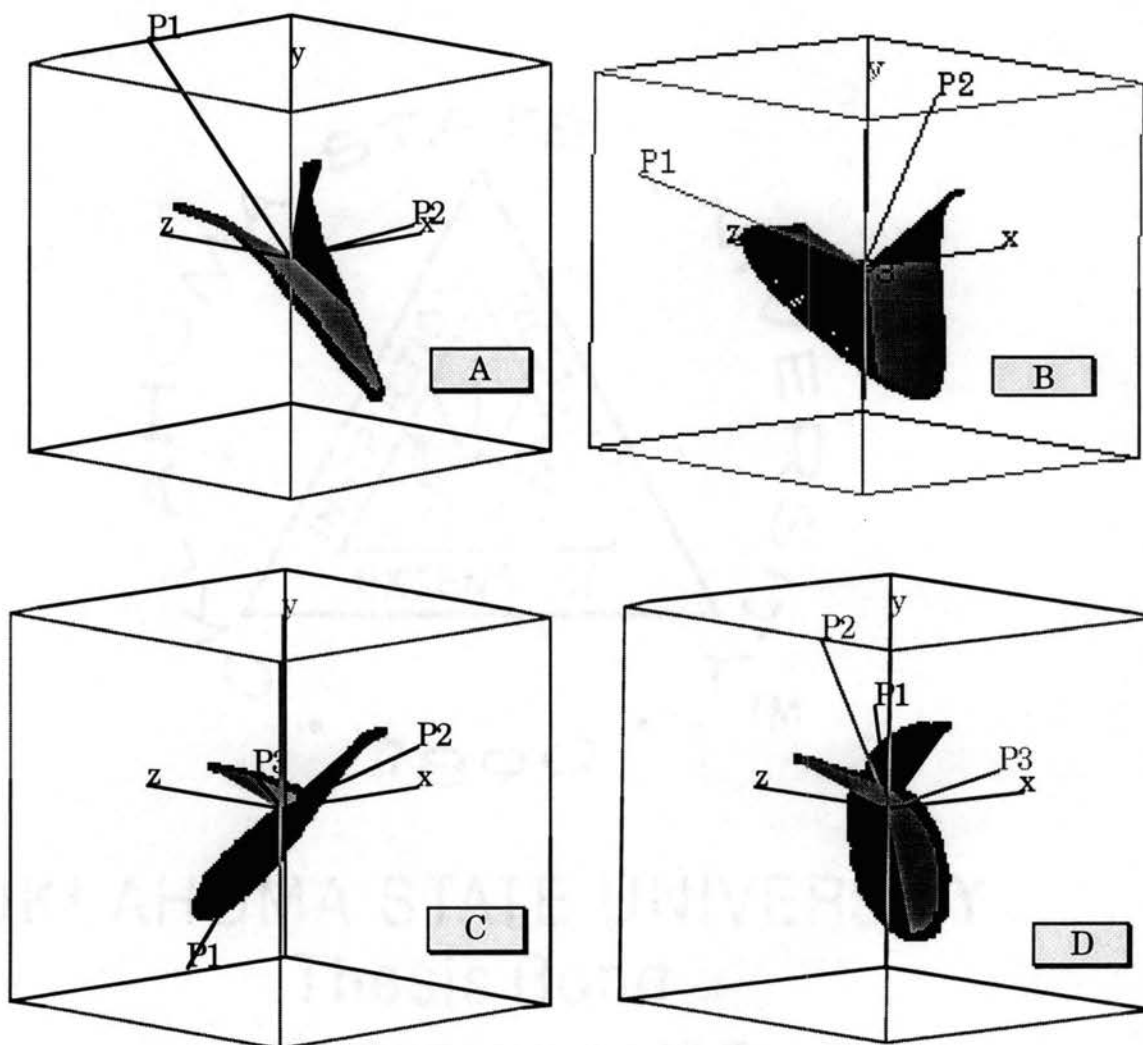


Figure 8.6: Spinning Plots<sup>®</sup> for the presentation of wavelength (x-coordinate; spectral data for the GGA complex (y-coordinate); and spectral data for (A) GGA; (B) GGH; (C) GHG; and (D) LGG complexes. The lines P1, P2, and P3, are the principal component axes for the PCA solutions. Rays have been added to the presentation to enhance the three dimensional effect. Dark and light areas distinguish the front four quadrants of the cube from the rear four quadrants.

assigned to the x-axis, the ellipticity data for the GGA complexed with the

copper is placed on the y-axis, and the ellipticity data for the GGA with the chemical impurities are plotted on the z-axis, four 3-D plots (Figure 8.6) were created for the purpose of applying Principal Component Analysis.

*Factor Analysis of Spinning Plot® Data:* To derive a quantitative mathematical algorithm, data reduction was done on each Spinning Plot® created. These values (eigenvalues and eigenvectors) from the three Principal Components calculated for the GGA/GGH combination plot are shown in Table 8.2.

**Table 8.2: Principal Components Calculated for the GGA vs. GGH system**

		PC1	PC2	PC3
<b>Eigenvalues</b>		1.9603	0.9798	0.0599
<b>Eigenvectors</b>	<b>nm</b>	0.19154	0.97273	-0.13080
	<b>GGA</b>	-0.68522	<b>0.22794</b>	0.69175
	<b>GGH</b>	0.70270	-0.04287	0.71019

Spatial projections of these same principal components are shown in Figure 8.6, plot B as P1, P2, and P3.

Of the twelve eigenfactors, the one that is most sensitive to variations in the identity of the analyte is PC22, highlighted in bold type in Table 8.2. Comparative PC22 values for all combinations with GGA are listed in Table 8.3.

**Table 8.3: Comparative PC22 Values for all the L-Tripeptides against GGA**

GGA	GGH	GGI	GGL
0.04519	0.88194	0.13171	-0.01932
GGP	GHG	LGG	YGG
0.10312	0.22794	0.89248	0.57491

Standard deviations in PC22 determined for data from 3-5 independent repeat measurements are  $\pm 0.002$ , meaning that based on the values in Table 8.3 total analytical selectivity was accomplished. The 3-D algorithm effectively reduced the 1500 original spectral data points to a single discretionary number, PC22.

The remaining question concerning the tripeptides and the ability of the 3-D method to distinguish subtle differences in the concentration is to determine if the PC22 values that show sensitivity to each type of tripeptide are selective when one of these structural similar tripeptides is added to GGA as an impurity of a small percent. If PC22 values were to correlate linearly with the amount of "chemical impurity", as in the dipeptide case, then the amount of impurity and possibly the identity of the impurity can be determined by differences.

Representative plots of PC22 versus percent impurity for solutions of GGA spiked with GGH, GGP, LGG, and YGG are shown in Figure 8.7. With the exception of GGI and GGL, correlation slopes are more than two times the  $\pm 0.002$  SD in the mean for PC22 values, which indicates that an impurity level as little as 1-3% can be measured with confidence. Analytical sensitivities are at least ten times more accurate than analogous plots in which maximum ellipticity values measured at a single wavelength are plotted against concentration.

Referring to Figure 7.6 and Figure 7.7 and the discussion that follows the changes in the maximum ellipticity values at a single wavelength over the 1-10% impurity range are less than the best resolution of  $\pm 2.0$  mdeg for

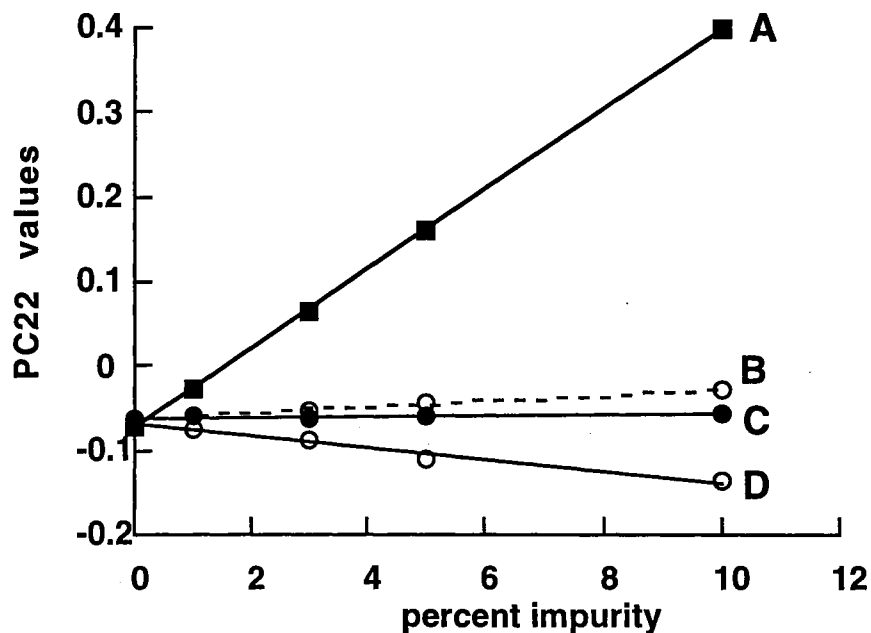


Figure 8.7: Plots of the percent chemical impurity vs. the eigenfactor PC22 for Cu complexes of (A) GGH; (B) GGP; (C) LGG; (D) YGG.

the CD instrumentation that was available for this study. The additional accuracy comes from the ability to conveniently include data at 1500 wavelengths.

#### Summary:

Again, where the 2-D method succeeded in providing a means to get accurate values for enantiomeric purities, the 3-D method provides a way to get a quantitative measure of non-enantiomeric chiral impurities. The slopes of the lines in Figure 8.7 are characteristic values that will identify a particular impurity.

The object of the study was to develop a protocol for distinguishing the

spectra of peptides and proteins and reporting the differences by using a color derivitization reaction, circular dichroism detection, and a novel use of principal component analysis. The method employs a modified biuret reagent where the host ligand has been exchanged for either a chiral ligand or the analyte itself. The detection method used was circular dichroism rather than UV/Vis spectrophotometry. Data analysis of the spectra was enhanced by running principal component analysis using the wavelength, the spectral data for the host ligand, and the spectra data for the mixed complex.

The color reaction which involves the binding of the protein to a copper ion was found to produce very specific spectra that could easily be observed in the visible region. Each small peptide(di- and tri-) that was analyzed showed spectral differences including those that contained the same residues but in a different order. These spectral differences could be characterized by performing principal component analysis and monitoring the eigenvector factors. These factors were found to be sensitive to chemical impurities that could be linearly correlated to the concentration of the impurity added and the slope of the factor versus concentration would identify the impurity. Quality Control information was also gathered using this method on human insulin and related analogs. This method used all of the spectral data gathered, increases the sensitivity by a factor of ten over the conventional single wavelength analysis, and is much faster than chromatographic techniques.

## LITERATURE CITED

- (1) *Encyclopedia of Pharmaceutical Technology, vol. 18*, Swarbrick, J.; Boylan, J. C., Eds.; Marcel Dekker, Inc.: New York, 1999.
- (2) *Analytical Applications of Circular Dichroism*; Purdie, N.; Brittain, H. G., Eds.; Elsevier:Amsterdam, 1994; Vol. 14.
- (3) Purdie, N.; Swallows, K. A.; Murphy, L.H. *Trends Anal. Chem.* **1990**, *9*, 136-142.
- (4) *ORD and CD in Chemistry and Biochemistry*; Crabbe, P.; Academic Press, Inc.:New York, 1972.
- (5) *Optical Rotatory Power*; Lowry, T. M.; Dover Publications, Inc.:New York, 1964.
- (6) *Optical Circular Dichroism: Principles, Measurements, and Applications*; Velluz, L.; Legrand, M.; Grosjean M.; Academic Press, Inc.:New York, 1965.
- (7) *Organic Structural Analysis*; Lambert, J. B.; Shurvell, H. F.; Verbilt, L.; Cooks, R. G.; Stout, G. H. Macmillan:New York, 1976.
- (8) *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*; Crabbe, P.; Holden-Day:San Francisco, 1965.
- (9) Armstrong, D. W. *Anal. Chem.* **1987**, *59*, 84A-91A.
- (10) *Optical Rotatory Dispersion: Applications to Organic Chemistry*; Djerassi, C.; MacGraw-Hill:New York, 1960.
- (11) Duffield, J. J.; Abu-Shumays, A. *Anal. Chem.* **1966**, *38*, 29A-58A.
- (12) Friedman, Michael "High Standards Maintained." USA Today **10 July 1998**: 14A.
- (13) Boswell, Clay "Setting Quality Standards for Dietary Supplements" *CMR Focus Report*. **13 July 1998**.
- (14) *Model J-500 Automatic Recording Spectropolarimeter Instruction Manual*, Japan Spectroscopic Co., Ltd.:Tokyo, 1979.
- (15) Anonymous *Industrial Engineering* **1995**, *27*, 1, 21.
- (16) Stinson, S. C. *Chem. Eng. News* **1993**, *71*, 38.



- (17) De Camp, W. H. *Chirality* **1989**, *1*, 2-6.
- (18) *Special Issues on: Chiral Discrimination*, Fell, A. F., Ed.; *Trends Anal. Chem.* **1993**, *12*, 125-189.
- (19) *Circular Dichroic Spectroscopy, Exciton Coupling in Organic Stereochemistry*, Harada, N.; University Science Books: Mill Valley, 1983.
- (20) *The Molecular Basis of Optical Activity: Optical Rotatory Dispersion and Circular Dichroism*, Charney, E.; John Wiley & Sons: New York, 1979.
- (21) *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*, Sneath, G., Ed.; Heyden & Sons Ltd.: London, 1967.
- (22) *Combinatorial Chemistry, Synthesis and Application*, Wilson, S. R., Czarnik, A. W., Eds.; John Wiley & Sons: New York, 1997.
- (23) *Encyclopedia of Spectroscopy*, Perkampus, H. H.; Verlagsgesellschaft mbH: Weinheim, 1995.
- (24) Purdie, N.; Swallows, K. A. *Anal. Chem.* **1989**, *61*, 77-89A.
- (25) *Chirality and the Biological Activity of Drugs*, Crossley, R.; CRC Press Inc.: Boca Raton, 1995.
- (26) *Spectroscopic Methods in Bioinorganic Chemistry*, Solomon, E. I., Hodgson, K. O., Eds.; American Chemical Society: Washington, DC, 1998.
- (27) *Chirality in Industry, The Commercial Manufacture and Applications of Optically Active Compounds*, Collins, A. N., Sheldrake, G. N., Crosby, J., Eds.; John Wiley & Sons: New York, 1994.
- (28) *Chromatographic Separations of Stereoisomers*, Souter, R. W.; CRC Press Inc.: Boca Raton, 1985.
- (29) *JMP® Statistics and Graphical Guide, Version 3.1*, Lehman, A., Sall, J.; SAS Institute Inc.: Cary, NC, 1995.
- (30) *Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry, Eighth Edition*, Doerge, R., Ed.; J. B. Lippincott Company: Philadelphia, 1982.

- (31) *The Merck Index of Chemicals and Drugs*; 7<sup>th</sup> ed.; Stetcher, P. G.; Finkel, M. J.; Siegmund, O. H.; Szafranski, B. M., Eds.; Merck & Co., Inc.:Rathway, N. J., 1960.
- (32) Anonymous *Industrial Engineering*, **1995**, *27*, 21.
- (33) *Circular Dichroism and Linear Dichroism*, Rodger, A.; Nordén B., Eds.; Oxford University Press:New York, 1997.
- (34) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman, G., Ed.; Plenum Press:New York, 1996.
- (35) *Circular Dichroism Principles and Applications*, Nakanishi, K.; et. al., Eds.; VCH Publishers, Inc.:New York, 1994.
- (36) Purdie, N.; et. al. *Journal of Pharmaceutical Sciences*, **1994**, *83*, 9, 1310.
- (37) Clennan, Edward L.; Heah, Poh Choo; *J. Org. Chem.* **1981**, *46*, 4107.
- (38) Kelly, S. M.; Price, N. C.; *Biochimica et Biophysica Acta* **1997**, *1338*, 161.
- (39) Woody, R. W.; *Methods in Enzymology* **1995**, *246*, 34.
- (40) Purdie, N.; Swallows, K. A.; *Biochimica Clinica* **1990**, *14(1)*, 29.
- (41) Purdie, N.; Swallows, K. A.; *Trends in Analytical Chemistry* **1990**, *9(3)*, 94.
- (42) Purdie, N.; Swallows, K. A.; *Trends in Analytical Chemistry* **1990**, *9(4)*, 136.
- (43) Purdie, N.; Swallows, K. A.; Murphy, L. H.; Purdie, R. B.; *Journal of Pharmaceutical & Biomedical Analysis* **1989**, *7(12)*, 1519.
- (44) Purdie, N.; Swallows, K. A.; *Analytical Chemistry* **1989**, *61*, 77A.
- (45) Camilleri, P.; *Analytical Proceedings* **1992**, *29*, 226.
- (46) Ariëns, E. J.; *Analytical Proceedings* **1992**, *29*, 232.
- (47) Goodall, D. M.; Wu, Z.; Lisseter, S. G.; *Analytical Proceedings* **1992**, *29*, 238.
- (48) Whelpton, R.; Buckley, D. G.; *Analytical Proceedings* **1992**, *29*, 249.

- (49) Hermann, T. W.; *Chem. Anal.* **1997**, *42*, 1.
- (50) Beary, J. F.; Siegfried, J. D.; Tavares, R.; *Drug Development Research* **1998**, *44*, 114.
- (51) Weiss, S. M.; *Molecular Diagnosis* **1998**, *3(1)*, 63.
- (52) Gladwell, Malcolm "Ingredient From New Supplier Failed in Generic Medication." *The Washington Post* **27 Sept. 1989**: 1A.
- (53) Anonymous "Hoechst Unit Stops Shipment of Antibiotics." *Chemical Market Reporter* **18 Nov. 1996**: 7.
- (54) McCoy, Michael "Biopharma Eyes Outsourcing." *Chemical and Engineering News* **27 July 1998**: 27.
- (55) Anonymous "Steris Receives Fine for Failing to Report Drug-Test Problems." *Wall Street Journal* **6 Feb. 1998**: B15.
- (56) Condon, E. U.; Altar, W.; Eyring, H.; *J. Chem. Phys.* **1937**, *5*, 753.
- (57) Moscovitz, A. *Advances in Chemical Physics*, Prigogine, I., Ed.; Vol. IV Interscience:New York, 1962.
- (58) Kondru Rama K.; et al.; *Science* **1998**, *282*, 2247.
- (59) Lodevico, Romulo G.; et al; *Talanta*, **1997** *44*, 1353.
- (60) Mannschreck, Albrecht; et al; *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 6, 469.
- (61) Sawada, Masami; et al.; *Chem. Commun.* **1998**, 1569.
- (62) Ferretti R.; et al.; *Journal of Chromatography B* **1998**, *710*, 157.
- (63) Xie, Guang-hua; et al.; *Biomedical Chromatography* **1997**, *11*, 193.
- (64) Han, S.; *Biomed. Chromatogr.* **1997**, *11*, 259.
- (65) Takashi, K., et. al.; *Journal of Chromatography A* **1998**, *805*, 295.
- (66) Wan, H.; Blomberg, L. G.; *J. Microcolumn Separations* **1996**, *8(5)*, 339.
- (67) Wan, H.; Blomberg, L. G.; *Journal of Chromatography A* **1997**, *758*, 303.

- (68) Wu, Z.; Goodall, D. M.; Lloyd, D. K.; *Journal of Pharmaceutical & Biomedical Analysis* **1990**, *8*(4), 357.
- (69) Hanna, G. M.; *Journal of Pharmaceutical & Biomedical Analysis* **1997**, *15*, 1805.
- (70) Horváth, P.; Gergely, A.; Noszál, B.; *Talanta* **1997**, *44*, 1479.
- (71) Bressolle, F.; Audran, M. Pham, T.; Vallon, J.; *Journal of Chromatography B* **1996**, *687*, 303.
- (72) Salam, A.; Meath, W. J.; *Chemical Physics* **1997**, *228*, 115.
- (73) Mannschreck, A.; Andert, D.; Eiglsperger, A.; Ghahl, E.; Buchner, H.; *Chromatographia*, **1988**, *25*(3), 182.
- (74) Horváth, P.; Gergely, A.; Noszál, B.; *J. Chem. Soc. Perkin Trans. 2* **1996**, 1419.
- (75) Däppen, R.; Voigt, P.; Maystre, F.; Bruno, A.; *Analytica Chimica Acta* **1993**, *282*, 47.
- (76) Lloyd, D. K.; Goodall D. M.; *Chirality* **1989**, *1*, 251.
- (77) Goodall, D. M.; *Trends in Analytical Chemistry* **1993**, *12*(4), 177.
- (78) Lu, W.; Cole, R. B.; *Journal of Chromatography B* **1998**, *714*, 69.
- (79) Spencer, K. M.; Edmonds, R. B.; Rauh, R. D.; *Applied Spectroscopy* **1996**, *50*(5), 681.
- (80) Beyrich, T.; Junghänel, J.; Mietz, G. Hauschke, J.; *Pharmazie* **1997**, *52*(7), 512.
- (81) Francotte, E. R.; *Chimia* **1997**, *51*, 717.
- (82) Smith, J. T.; *Electrophoresis* **1997**, *18*, 2377.
- (83) Andreas F. R.; Huhmer, G. I.; Aced, M. D.; Perkins, R.; et. al.; *Anal. Chem.* **1997**, *69*, 29R.
- (84) Crimmins, D. L.; *Analytica Chimica Acta* **1997**, *352*, 21.
- (85) Schwarz, M.; Kline, W.; Matuszewski, B.; *Analytica Chimica Acta* **1997**, *352*, 299.

- (86) *Biochemistry, 2nd Edition*, Voet, D.; Voet, J. G.; John Wiley & Sons, Inc.: New York, 1995.
- (87) Greenfield, N. J.; *Analytical Biochemistry* **1996**, *235*, 1.
- (88) Takayuki, I., et. al.; *Journal of Chromatography A* **1998**, *813*, 267.
- (89) Markussen J.; et al.; *Diabetes Care* **1983**, *6*, suppl. 1, 4.
- (90) Rodgers, Katie; *Drug Topics* **1994**, *138*, 58.
- (91) Ross, Philip E.; *Forbes* **1996**, *157*, 150.
- (92) Xiaoquig, C.; Jorgensen, A. M.; Bardrum, P.; Led, J. J.; *Biochemistry* **1997**, *36*, 9409.
- (93) Derewenda, U.; et. al.; *Nature* **1989**, *338*, 594.
- (94) Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A.; *Nature* **1994**, *372*, 746.
- (95) Sparrow, L. G.; et. al.; *The Journal of Biological Chemistry* **1997**, *272(47)*, 29460.
- (96) Chang, A. Y.; *Ann. Rep. Med. Chem.* **1974**, *9*, 182.
- (97) Catsoyannis, P.G.; *Science* **1966**, *154*, 1509.
- (98) Rittel, W., et al.; *Jelv. Chim. Acta* **1974**, *67*, 2617.
- (99) Sigel, H.; Martin, R. B.; *Chemical Reviews* **1982**, *82(4)*, 386.
- (100) Wilson, E. W.; Kasperian, M. H. Martin, R. B.; *Journal of the American Chemical Society* **1970**, *92(18)*, 5365.
- (101) Martin, R. B.; Tsangaris, J. M.; Chang, J. W.; *Journal of the American Chemical Society* **1968**, *90(3)*, 821.
- (102) Czarnecki, J. J.; Margerum, D. W.; *Inorganic Chemistry* **1977**, *16(8)*, 1997.
- (103) Tsangaris, J. M.; Martin, R. B.; *Journal of the American Chemical Society* **1970**, *92(14)*, 4255.
- (104) Wilson, E. W.; Martin, R. B.; *Inorganic Chemistry* **1971**, *10(6)*, 1197.
- (105) Biaser, H. U.; *Chemical Reviews* **1992**, *92(5)*, 948.

- (106) Formicka-Kozłowska, G.; Kozłowska, H.; Siemion, Z.; Sobezyk, K.; Nawrocka, E.; *Journal of Inorganic Biochemistry* **1981**, *15*, 201.
- (107) *Metal Ions in Biological Systems: Amino Acids and Derivatives as Ambivalent Ligands. Vol. 9.* Sigel, H., Ed.; Marcel Dekker, Inc. : New York, 1979.
- (108) Armstrong, D. W.; Zukowski, J.; Yubing, T.; Berthod, A.; *Anal. Chim. Acta* **1992**, *258*, 83.
- (109) Pearson, K.; *Phil. Mag.* **1901**, *6(2)*, 559.
- (110) Hotelling, H.; *The Journal of Educational Psychology*, **1933**, *24*, 417.
- (111) *Principal Component Analysis*, Jolliffe, I. T.; Springer-Verlag: New York, 1986.
- (112) *Multivariate Statistical Methods: Among-Groups Covariation*, Eds. William R. Atchley, Edwin H. Bryant; Dowden, Hutchinson & Ross, Inc.: Strousburg, 1975.
- (113) *Analysis of Neuropeptides by Liquid Chromatography and Mass Spectroscopy* Desiderio, D. M.; Elsevier: New York, 1984.
- (114) Kingley, G. R.; *J. Lab. Clin. Med.*, **1942**, *27*, 840.
- (115) Fisherkeller, M. A., Friedman, J. H., and Tukey, J. W.; *PRIM-9: An interactive multidimensional data display and analysis system*, SLAC-PUB-1408, Stanford, CA., 1974.
- (116) *Spectrochemical Analysis* Ingle, J.D.; Crouch, S. R.; Prentice Hall: Englewood Cliffs, New Jersey, 1988.

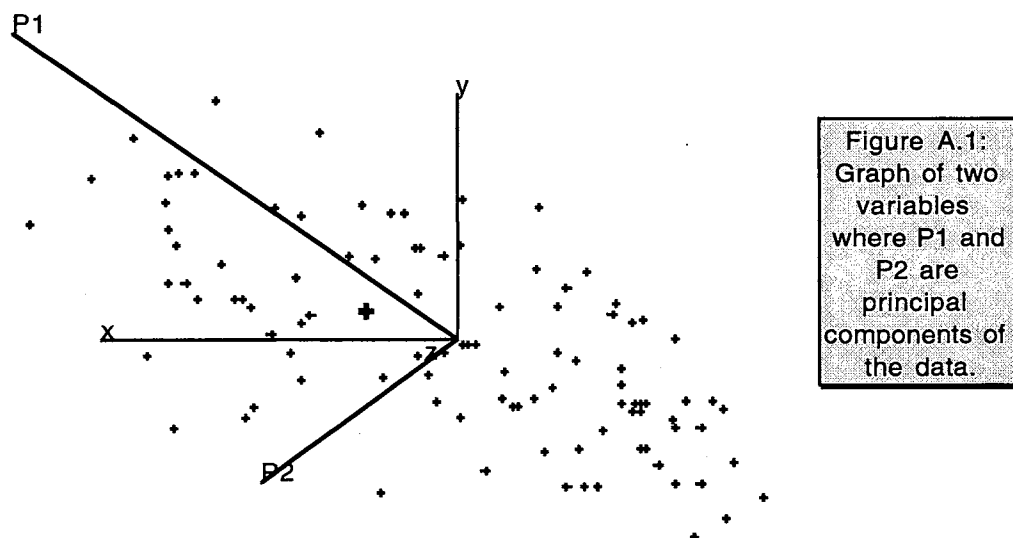
## APPENDIX A

## PRINCIPAL COMPONENT ANALYSIS

The reduction of data to a single numerical value has been a part of science from the beginning. The goal of data reduction is to be able to focus on a smaller set of values that can allow sense to be made of the experiments. The most common form of data reduction for spectroscopic methods is the univariate calibration model also known as Beer's Law (116). The typical scenario of how spectroscopic data are used for the determination of an unknown consists of the acquisition of a series of absorption measurements versus concentration. In many instances, data for the entire spectrum are collected and from this a single datum is extracted at the wavelength of maximum absorbance. The additional information contained in the spectrum remains unused. Absorbance data are plotted as a function of concentration to construct a typical Beer's law working curve. Simple linear regression is used for predicting the concentration of future samples of unknown concentration.

The technological basis for the ability to increase the precision of both enantiomeric purity and chemical purity determination is the direct result of the emergence of the laboratory microcomputer for data acquisition and analysis. This has provided the analytical chemist with the ability to accumulate vast amounts of data in a relatively short period of time. It has become much easier to use the additional data to increase the precision and associated predictive capabilities in chemical determinations as commercial software for this purpose has become more commonplace (29).

The approach of using more than one measurement (variable or wavelength) per sample for spectrochemical analysis results in what is known as an overdetermined system. Although such a system has no unique solution, systematic methods for data reduction can be employed to reduce the number of variables to match the number of samples. If the number of variables is less than the number of samples, then a solution may not exist due to collinearity. This method of multivariate analysis can at times extend the task of the chemist to search for variables that will increase the correlation of the measurements. Also in many cases a model must be made and constructed in such a way that the model has “learned” the correct response to a set of data. This model building takes time and



may ultimately only be useful for a very select set of analytes.

Another approach that is more like the univariate method but uses the spirit of the multivariate method, is termed Principal Component



Analysis (PCA). This method uses vectors of information plotted against one another to derive orthogonal Principal Components (PC) that consist of a length and a direction. Using PCA reduces the dimensionality of a set of data. This means finding a way to picture the structure of the data as completely as possible using as few variables as possible (115).

The actual procedure for determining the PC of a system of observations is simply the iterative process of finding the least squares in orthogonal directions of maximum variance (112). The solutions are reported as eigenvalues (the length) and eigenvectors (the direction) for each PC. If in two dimensions, the resulting vector solutions lie along the major and minor axes of an ellipse, Figure A.1, that would enclose the data. The actual eigenvectors used to correlate the concentration change of one dipeptide chemical impurity added to GA or to characterize the type of insulin have been the z component of the second principal component. The first PC accounts for most of the variability between the sets of data. All the eigenvectors are changing but the value that shows the greatest linear change becomes useful in analytical analysis.

The subtle changes of the eigenvectors correspond to rotation of the ellipsoid major and minor axes which are really the vectors of the PC (111). The actual measure that is linearly correlating (in the case of the dipeptides) with the percent impurity is the growth or decline of the z-axis data in terms of the CD data for the impurity. An easier way to approach this would be to imagine that the ellipsoid generated by the PCs is like a skin covering the data, fitting closest to those maxima and minima that are

protruding. As the composition of the complex changes with the advent of a chemical impurity the parts of the spectra that will be affected the most will be those same maxima and minima. Because the PC is a vector that is defined in three directions even a lateral shift of the spectra will cause one of the eigenvector values to change, indicating that the PC is changing direction. Since the eigenvector values contain directional information for the PCs, the values are insensitive to concentration changes unless the change is so dramatic that a spectra feature is lost due to some new equilibrium that is established. This characteristic makes it ideal for chemical impurity measurements but ineffective in enantiomeric determination which can be compensated with the 2-D algorithm earlier discussed.

VITA

Dennis William Province

Candidate for the Degree of

Doctor of Philosophy

Thesis: ANALYTICAL APPLICATIONS MEASURING THE CHIRALITY OF PROTEINS AND PEPTIDES USING CIRCULAR DICHROISM

Major Field: Chemistry

Biographical:

Personal Data: Born in Pueblo, Colorado, on March 2, 1971, the son of Darryl and Pat Province.

Education: Graduated from Pueblo Central High School, Pueblo, Colorado, June 1989. B.A. in Chemistry and Economics, University of Colorado at Boulder, Boulder, Colorado. Completed requirements for the Doctor of Philosophy in Chemistry at Oklahoma State University in May 1999.

Professional Experience: Teaching Assistant at Oklahoma State University 1994-present. Undergraduate Research Associate for the CCHE (Colorado Commission for Higher Education) working at CIRES (the Center for Interdisciplinary Research in Environmental Sciences), Summer 1993. Undergraduate Research Associate working with a GC/MS project funded by a grant from NASA, Fall of 1993.