INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
STUDIES IN PROTEIN FOLDING: UNDERSTANDING HELIX BACKBONE INTERACTIONS AND THE STRUCTURAL CHARACTERIZATION OF A COMPACT DOMAIN

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

by

Chi-Fon Chang, B.S.

* * * * *

The Ohio State University

1997

Dissertation Committee:

Dr. Micheal H. Zehfus, Adviser
Dr. Charles L. Brooks, Adviser
Dr. James A. Cowan

Approved by

Advisers
Biophysics Graduate Program
ABSTRACT

The protein folding problem is still one of the most important intellectual challenges in biological science. An enormous amount of work has been done to study this problem. In this thesis, the protein folding problem is attacked at two different levels. At the secondary structural level, model helices are manipulated to try to reveal some of the basic interactions that stabilize the helix. At the tertiary structural level, the protein structure is studied in an attempt to isolate and characterize a small subdomain.

Helical peptides, acetyl-WGG(EAAAR)$_4$A-amide and acetyl-WGG(RAAAA)$_4$R-amide, have been chosen as model peptides to study helix backbone interactions. Since the hydrogen bond between amides and carbonyls along the peptide backbone is an important feature of the $\alpha$-helix, experiments were designed to interfere the backbone hydrogen bonds. The backbone hydrogen bonds of model helices have been disrupted in two ways. In the first, the hydrogen bond donor ability of specific peptide bonds has been removed by replacing the NH group with a NCH$_3$ group. In the second modification, the hydrogen bond accepting ability of a residue is eliminated by replacing the CO functional group with a CH$_2$ group. These modifications have been placed into the model peptides so that the same hydrogen bond is removed. CD has been used to follow the helix-coil transition of the model peptides and derivative peptides. Both the
N-methyl modification and the reduced carbonyl modification decrease the helicity of peptides; however, the amount of destabilization is more than expected. The thermodynamic properties of the disruption can be analyzed using different approaches: The Two-State model, Lifson-Roig theory, and the more complicated Dichroic model. Different backbone-backbone interactions may be addressed by comparing the two different backbone modifications.

Understanding the properties that identify domains can also help us study the protein folding problem. The compact unit theory, an algorithm that identifies domains based on their compactness, has been chosen for this study. When applied to staphylococcal nuclease, several compact units were located. The most compact unit, consisting of residues 12-36 of the nuclease sequence, has been synthesized and studied. If this compact unit is a domain, it should retain its native β-β-β structure when isolated from the intact protein. CD and NMR were chosen to study the structural properties for this peptide, and many solvent conditions were used to try to find the structure. However, conditions where the peptide could form its native structure have not been found.
To the Memory of my Father

&

To my Mother and Sisters
I am truly grateful to my adviser, Dr. Micheal H. Zehfus, for all his guidance and advise on my graduate research. His encouragement and understanding during my difficult time are very much appreciated. I also like to thank him for his patience in correcting both my scientific and English mistakes when preparing this dissertation.

I would like to thank my dissertation and examination committee members: Dr. Charles L. Brooks, Dr. James A. Cowan, Dr. Wayne J. Becktel and Dr. Ming-Daw Tsai for their assistance and advise.

Special thanks to Dr. Mark Maciejewski for his friendship and help during my first four years in the laboratory. My sincere appreciation also goes to my friends who show their support and concern. Especially to those who also come from Taiwan for their sharing and assistance during these many years.

Finally, I would like to thank my family. It would be impossible for me to finish my graduate studies without their continuous support and encouragement. If I have made any achievement with the completion of this dissertation, they deserve all the compliments.
VITA

May 1, 1967 ................................................................. Born, Taiwan

1985 – 1989 ................................................................. B.S., Physics,
Fu-Jen Catholic University
Taipei, Taiwan

1989 – 1990 ................................................................. Research Assistant
Institute of Atomic and
Molecular Sciences
Academia Sinica
Taipei, Taiwan

1991 – present ............................................................. Graduate Student, Biophysics
The Ohio State University
Columbus, Ohio, USA

1992 – present ............................................................. Graduate Research Associate
The Ohio State University
Columbus, Ohio, USA

PUBLICATIONS

1. C.-F. Chang and Lian-Pin Hwang, Relaxation Studies of 2IzSz Spin Order in
Liquid Benzene: Selection of Chemical Shift Anisotropy Parameters, Journal of

2. C.-F. Chang and M. H. Zehfus, Effects of N-Methyl Isosteres on Helix Stability,

3. C.-F. Chang and M. H. Zehfus, The effect of N-Methylation on Helical Peptides,
Biopolymers, 40, 609-616, 1996.
FIELDS OF STUDY

Major Field: Biophysics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapters</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
</tbody>
</table>

## Chapters

1. Introduction
   1.1 Effects of backbone modifications .................................................. 2
   1.2 Structural characterization of a compact peptide......................... 4

2. Helix-Coil Theory ............................................................................. 7
   2.1 Background ..................................................................................... 8
   2.1.1 The polypeptide model ................................................................. 8
   2.1.2 The helix model .......................................................................... 9
   2.1.3 Circular dichroism of peptides ................................................... 10
       2.1.3.1 Basic concept of circular dichroism ..................................... 10
       2.1.3.2 The amide chromophore in peptides ..................................... 11
       2.1.3.3 Rotational strength of α-helix .......................................... 12
       2.1.3.4 CD and helix content of helical peptides ......................... 13
   2.2 The Two-State model ..................................................................... 16
   2.3 Original Lifson-Roig theory .......................................................... 17
       2.3.1 Conformational state of a residue ......................................... 18
       2.3.2 Configuration energy, partition function and statistical weights 18
       2.3.3 Matrix method and properties of the partition function .......... 22
7. Structural Characterization of the Most Compact Unit from Staphylococcal Nuclease

7.1 Background information

7.1.1 Compact unit theory

7.1.2 Application of compact unit theory

7.2 Material and methods

7.2.1 Target protein and its compact units

7.2.2 Peptide synthesis and purification

7.2.3 Spectral measurements

7.3 Results

7.3.1 Detergents

7.3.2 Osmolyte solvents

7.3.3 Other solvents

7.4 Summary

Appendices

A. A Maple Program for Simple Uniform Helix Analysis

B. A Maple Program for Modified Uniform Helix Analysis

C. A Fortran Program for Full Helix Analysis

D. A Fortran Program for Dichroic Model

List of References
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 The weightings of center residues for the triplet: Comparison between original Lifson-Roig matrix (Lifson &amp; Roig, 1961) and the modified Lifson-Roig matrix (Doig &amp; Baldwin, 1995)</td>
<td>40</td>
</tr>
<tr>
<td>2.2 Corresponding statistical weights for the center residue of a quintet of residues using Doig et al.'s definition (Doig &amp; Baldwin, 1995) and Rohl et al.'s correction (Rohl et al., 1996)</td>
<td>41</td>
</tr>
<tr>
<td>2.3 The intrinsic parameter library for all residues in water: generated using modified Lifson-Roig model (Rohl et al., 1996)</td>
<td>42</td>
</tr>
<tr>
<td>2.4 All possible peptide conformations for a sequence with five residues and their corresponding decimal integer</td>
<td>43</td>
</tr>
<tr>
<td>3.1 Peptides used to determine intrinsic parameters for Ala (A), Gly (G), Lys (K), Gln (Q), Tyr (Y), acetyl group (X) and amide blocking group (Z)</td>
<td>74</td>
</tr>
<tr>
<td>3.2 Peptides used to determine intrinsic parameters for Trp (W), Glu (E), Arg (R), and Pro (P)</td>
<td>75</td>
</tr>
<tr>
<td>3.3 The intrinsic parameter library generated using modified Dichroic model for some amino acids and blocking groups</td>
<td>76</td>
</tr>
<tr>
<td>4.1 The notation of N-methyl modified peptides</td>
<td>94</td>
</tr>
<tr>
<td>4.2 Purification conditions for N-methyl modified peptides using HPLC</td>
<td>95</td>
</tr>
<tr>
<td>4.3 Mean residue ellipticity ([θ]_obs), mean amide ellipticity ([θ]'_obs), helix content (f_H,obs) and reduction of helix content for N-methyl modified peptides at 0°C</td>
<td>96</td>
</tr>
</tbody>
</table>
4.4 Comparison of helix content for 24p2mxx peptides under different NaCl concentrations................................................................. 97
4.5 Analysis using the Two-State model for N-methyl modification at 0°C. ........ 98
4.6 Simple uniform helix analysis for N-methyl modification at 0°C.................. 99
4.7 Modified uniform helix analysis for N-methyl modification at 0°C............... 100
4.8 Full helix analysis for N-methyl modification at 0°C.................................. 101
4.9 Analysis using Dichroic model for N-methyl modification at 0°C.............. 102
5.1 The notation of the reduced carbonyl peptides........................................... 120
5.2 Mean residue ellipticity([θ]_obs , [θ]'_obs ), helix content (f_obs) and reduction of helix content for the reduced carbonyl peptides at 0°C.............................. 121
5.3 Analysis using the Two-State model for reduced carbonyl modification at 0°C ................................................................. 122
5.4 Simple uniform helix analysis for reduced carbonyl modification at 0°C ...... 123
5.5 Modified uniform helix analysis for reduced carbonyl modification at 0°C .... 124
5.6 Full helix analysis for reduced carbonyl modification at 0°C...................... 125
5.7 Analysis using Dichroic model for reduced carbonyl modification at 0°C .... 126
5.8 Comparison of helix content for N-methyl modification (NCH₃) and reduced carbonyl modification (CH₂)...................................................... 127
6.1 w values (<_w> or w'(i)) and their corresponding ΔAG(kcal/mole) obtained using different models...................................................... 139
7.1 Summary of detergents and conditions used to find the β-sheet structure for Staphylococcal nuclease [12-36]................................................. 147
7.2 Summary of osmolyte solvents used and their resulting CD spectra for Staphylococcal nuclease [12-36]......................................................... 148
7.3 Other solvents used to find the structure for Staphylococcal nuclease [12-36] and their resulting CD spectra. ......................................................... 149
## List of Figures

<table>
<thead>
<tr>
<th>Figures</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 N-methyl modification, where the NH of normal peptide bond has been replaced with an NCH$_3$</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Reduced carbonyl modification, where the CO of normal peptide bond has been replaced with a CH$_2$</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Definition of torsion angles $\phi$, $\varphi$, and $\omega$ for peptide backbone</td>
<td>44</td>
</tr>
<tr>
<td>2.2 Ramachandran plot for poly-L-alanine</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Idealized diagram of the main chain in an $\alpha$-helix, where the arrow indicates the direction from N terminal to C terminal</td>
<td>46</td>
</tr>
<tr>
<td>2.4 Plane-polarized light and elliptically polarized light</td>
<td>47</td>
</tr>
<tr>
<td>2.5 Energy level diagram for amide chromophore</td>
<td>48</td>
</tr>
<tr>
<td>2.6 CD spectra for pure secondary structures</td>
<td>49</td>
</tr>
<tr>
<td>2.7 Residue numbering and statistic weight assignment for different models: a sequence with free amino acid termini</td>
<td>50</td>
</tr>
<tr>
<td>2.8 Residue numbering and statistic weight assignment for different models: a sequence with acetyl blocked N terminal and amide blocked C terminal</td>
<td>51</td>
</tr>
<tr>
<td>3.1 Steps in solid phase synthesis: deprotection, activation, and coupling</td>
<td>77</td>
</tr>
<tr>
<td>3.2 Deportection of the Fmoc group using piperidine</td>
<td>78</td>
</tr>
<tr>
<td>3.3 HBTU activation for Fmoc solid-phase peptide synthesis</td>
<td>79</td>
</tr>
</tbody>
</table>
3.4 Goodness of the fit of experimental data to the Dichroic model predications (correlation coefficient R=0.93). The observed ellipticity of each peptide in Table 3.1 and 3.2 is plotted against its calculated ellipticity using the best-fit parameters in Table 3.3

4.1 Melting curves measured by CD at 222 nm for 24p1m peptides in 10 mM NaCl and 1 mM phosphate buffer at pH 7.

4.2 Melting curves measured by CD at 222 nm for 24p2m peptides in 10 mM NaCl and 1 mM phosphate buffer at pH 7.

4.3 Melting curves measured by CD at 222 nm for 24p2m peptides in 1 M NaCl and 1 mM phosphate buffer at pH 7.

4.4 NaCl concentration dependence of 24p2con peptide at 0°C and pH 7.

4.5 \( R_{lnK} \) vs. \( 1/T \) plot for 24p1m peptides (10 mM NaCl, pH 7) with 95% confidence level linear regression lines.

4.6 \( R_{lnK} \) vs. \( 1/T \) plot for 24p2m peptides (1 M NaCl, pH 7) with 95% confidence level linear regression lines.

4.7 A portion of \(^1\)H NMR spectra for 24p1m21 and 24p2m21 showing the trans and cis N-methyl peaks.

5.1 24p2k07 titration at 298K: \(^1\)H NMR chemical shift data for the modified residue (Ala 7).

5.2 CD spectra for 24p2k peptides at different pH values (1 M NaCl and 0°C), where \([\theta]\) (deg. cm\(^2\)-dmol\(^{-1}\)) is the mean residue ellipticity.

5.3 Melting curves measured by CD at 222 nm for 24p2k peptides in 1 M NaCl and 1 mM phosphate buffer at pH 7.

5.4 \( R_{lnK} \) vs. \( 1/T \) plot for 24p2k peptides (1 M NaCl, pH 7) with 95% confidence level linear regression lines.

7.1 Structure of Staphylococcal nuclease determined by X-ray at 1.7A resolution (Hynes & Fox, 1991).

7.2 Structures for different kinds of detergents: anionic detergents (A, B, C), cationic detergent (D), and nonionic detergents (E,F).
7.3 Structures for osmolyte solvents used in this study. ............................................. 152
7.4 Structures for 2,2,2-trifluoroethanol (TFE), hexafluoroisopropanol (HFIP) and ethanol. .................................................................................................................. 153
CHAPTER 1

INTRODUCTION

Globular protein molecules contain a hierarchy of structure: (1) The primary structure is the linear arrangement of amino acids along a polypeptide chain; (2) The formation of the secondary structure in a local region of the polypeptide chain is then determined by the primary structure; (3) And the tertiary structure is formed by packing secondary structure elements into one or several compact globular units called domains, and the domains pack against each other to complete the globular protein; (4) Finally, individual protein subunits can interact to form useful complexes.

The functional properties of proteins depend on their structures, thus to understand the biological function of proteins, a knowledge of how proteins fold into their three-dimensional structures is necessary. Although an enormous amount of work has been done to try to solve the protein folding problem, this problem is still one of the most important intellectual challenges in biological science. To study the protein folding problem, two different approaches are described in this dissertation.
1.1 Effects of backbone modifications

The first part of this dissertation focuses on the effects of backbone modifications in helical peptides. The α helix is one of the major secondary structures in proteins, thus studying the formation of the α helix by peptides should help us to understand one part of the protein folding problem. Since the backbone/backbone interactions are a major factor that stabilizes a helix, the approach used here involves modification of the peptide backbone to study the formation of the α helix.

In addition to studying helix formation, the backbone modifications are pharmaceutically interesting. Modified peptide bonds are usually not be cleaved by enzymes that degrade peptides, thus the biological lifetime of the backbone modified peptide-based drugs is usually increased. However, before backbone modifications can be applied in this manner, it is necessary to know the effects of these modifications on the peptide's structure.

One of the most important backbone/backbone interactions is the backbone hydrogen bond; yet, the strength of the backbone hydrogen bond is not known. Thus, in this thesis, modifications have been designed to interfere the peptide bond's hydrogen bonding ability, and using these modifications as probes, it is hoped that the strength of the backbone hydrogen bond as well as the other interactions can be determined. The peptide bond (CONH) has been replaced with two different linkages, CONCH₃ (N-methyl) and CH₂NH (reduced carbonyl), that selectively remove the hydrogen bond donor or acceptor ability of this linkage. As shown in Figures 1.1 and 1.2, these
backbone modifications preserve the $C_\alpha-C_\alpha$ distance between amino acids, but change the chemical nature of the peptide bond.

Effects of the backbone modification could be complex because changes could occur in hydrogen bonding ability, geometry, steric hindrance, and charge interactions. However, it is hoped that by carefully analyzing the results from several different modifications, it will be possible to isolate and quantitate the backbone hydrogen bond as well as the other forces involved in helix formation. To analyze the data, helix-coil transition theory has been applied. Chapter 2 reviews the helix-coil theory in detail, and the experimental techniques and data analysis methods are summarized in Chapter 3.

Chapter 4 describes the results of the N-methyl modification (CONCH$_3$) where the hydrogen bond donor (NH) of a peptide bond has been replaced by a bulky NCH$_3$ group (Figure 1.1). This modification was first studied because it is a commonly used modification and easy to incorporate. However, the effects of this modification are complex, since it includes both the removal of a hydrogen bond and the introduction of a bulky methyl group into the helix backbone. It is therefore necessary to compare these results with the other modifications. Chapter 5 introduces the reduced carbonyl modification (CH$_2$NH). This modification removes the hydrogen bond acceptor (CO) from the peptide bond by reducing the CO group to a CH$_2$ group (Figure 1.2). Results are then compared with the N-methyl modification.

After comparing the two different modifications, the results indicate that data obtained from these two modifications is still insufficient to differentiate different backbone/backbone interactions in an $\alpha$ helix. In addition, none of the helix-coil
transition models can analyze the data for our peptide systems. Further experiments that might provide more useful information and some adjustments that might improve the theoretical model are discussed in Chapter 6.

1.2 Structural characterization of a compact peptide

Another way to study the protein folding problem is to understand the properties that identify domains. A domain is defined as a polypeptide chain or a part of a polypeptide chain that can independently fold into a stable tertiary structure (Branden & Tooze, 1991). It is important to locate folding domains, because a domain represents the smallest and simplest folding unit that can be studied. If each small folding unit can be identified, then the protein folding problem may be understood. Several models based on different properties exist for finding domains; however, few experiments have been done to test these models.

In this study the protein staphylococcal nuclease, whose structure is well defined, was chosen to test a model that identifies domains bases on their compactness. The concept of this model is further discussed in Chapter 7. Using the compactness model, eight compact units were identified in staphylococcal nuclease. The most compact unit, consisting of residues 12–36 of the nuclease sequence, has been synthesized and studied here. Unfortunately, the large number of hydrophobic residues in this peptide made it difficult to work with experimentally. Conditions to study this β–β–β structure (Hynes & Fox, 1991) were never found. Chapter 7 summarized the work done on this peptide.
Figure 1.1: N-methyl modification, where the NH of normal peptide bond has been replaced with an $\text{NCH}_3$. 
Figure 1.2: Reduced carbonyl modification, where the CO of normal peptide bond has been replaced with a CH$_2$. 
CHAPTER 2

HELIX-COIL THEORY

The main reason for studying the helix-coil transition in peptides is to understand precisely one part of the protein folding problem. The application of helix-coil transition models is now close to the point where the formation of helix in a peptide can be predicted from its primary sequence.

In the late 1950s and early 1960s, the theory of helix-coil transition was developed into one of the most elegant and important areas of macromolecular science. Schellman provided the first simple thermodynamic treatment of the helix-coil transition in 1955 (Schellman, 1955). A number of papers followed, and all used essentially the same model but adopted different approaches. Recent discussions of the helix-coil transition have focused on two approaches: one by Zimm and Bragg (Zimm & Bragg, 1959), the other by Lifson and Roig (Lifson & Roig, 1961). Both approaches are based on a statistical mechanical model where the contribution of individual residues to the stability of a helical peptide can be determined. The conversion of these two models has been established by Qian and Schellman (Qian & Schellman, 1992). The units in Zimm-Bragg theory are "peptide groups", rather than "residues" in Lifson-Roig theory.
Accordingly, Lifson-Roig model is more straightforward and easier to use for modification problems because it treats the modification as simply the change in intrinsic parameters of the modified residue. Here, the Lifson-Roig model is therefore chosen to study the helix-coil transition.

2.1 Background

To understand the helix-coil theory, it is useful to briefly review the physical structure of an α-helix. Therefore, the peptide model and helix structure will first be reviewed. Then, since circular dichroism has become one of the most widely used spectroscopic methods to follow the helix-coil transition in peptides experimentally, the circular dichroism of peptides is discussed.

2.1.1 The polypeptide model

The conformation of the polypeptide backbone is defined by the three main-chain torsion angles, φ, ϕ and ω. Since peptide units are rigid groups that are linked into a chain by covalent bonds at the Cα atoms, the only freedom they have is rotation around the Cα bonds. φ (ϕ) is defined as the angle of rotation around the N−Cα bond, and the angle around the Cα−C bond from the same Cα atom is ψ (ψ) (Figure 2.1). Except for Proline, ω, the rotation about C′−N bond, may be confined to near 0° or 180° because the peptide bond connecting the adjacent peptide units is known to have a partial double-bond character, and are assumed to be fixed. Therefore, the
conformation of the main chain of the polypeptide is completely determined when the $\phi$ and $\varphi$ angles for each amino acid are defined.

It is important to recognize that not all possible combinations of $\phi$ and $\varphi$ angles are allowed for a residue in a polypeptide chain, because of the steric hindrance between side chain and main chain atoms. Ramachandran and his group first made the calculation of sterically allowed $\phi$, $\varphi$ regions. They constructed plots showing which areas are allowed, and which are not, in steric contour diagrams called Ramachandran plots (Ramachandran & Sasisekharan, 1968). Such plots show separate allowed regions that correspond to conformational angles found for different secondary structures. For example, Figure 2.2 is a Ramachandran plot for poly-L-alanine that shows separate regions for the usual right handed $\alpha$ helices, $\beta$ strands and left handed $\alpha$ helices, respectively.

2.1.2 The helix model

The $\alpha$-helix is one major class of the secondary structure. It is a right handed helix that has 3.6 residues per turn and is stabilized by the hydrogen bond between the carbonyl of the $i-2$ residue and the NH of the $i+2$ residue (Pauling et al., 1951). To bring these two groups, C=O ($i-2$)...NH ($i+2$), into the proper position requires the spatial fixing of the $i-1$, $i$ and $i+1$ residues. Thus, it is assumed that when any pair of ($i-2$, $i+2$) is bounded by a hydrogen bond, the three residues, $i-1$, $i$ and $i+1$ must all be in the $\alpha$ helix conformation (Figure 2.3). According to Ramachandran plots, the $\phi$, $\varphi$
region for the $\alpha$ helix is relatively small. The average $\phi$, $\varphi$ values for a residue in the $\alpha$-helix conformation has been widely used are $(\phi, \varphi) = (-57^\circ, -47^\circ)$ based on the geometry of Pauling et al. (Pauling et al., 1951).

2.1.3 Circular dichroism of peptides

2.1.3.1 Basic concept of circular dichroism

Circular dichroism (CD) is an absorptive phenomenon of optically active chromophores that results from the difference in absorption of left (L) and right (R) circularly polarized light. Only an asymmetric molecule or a molecule in an asymmetric environment can generate the difference in absorption of L and R circularly polarized light. This difference is usually expressed in terms of the absorption coefficients for L and R circularly polarized light, $\varepsilon_L$ and $\varepsilon_R$, as $\Delta \varepsilon = \varepsilon_L - \varepsilon_R$. When the absorption of the L and R circularly polarized components have unequal amplitude, the resulting transmitted light becomes elliptically polarized (Figure 2.4). This polarization can be expressed as ellipticity $\theta$. The ellipticity in degree (deg) is defined as:

$$\theta = 2.303 \times (A_L - A_R) \times 180/(4\pi) \quad (2.1)$$

Where $A$ is the absorbance defined by Beer-Lambert law as:

$$A = \varepsilon \times d \times C_M \quad (2.2)$$

in which $\varepsilon$ is the absorption coefficient (or extinction coefficient) in the unit of $M^{-1} \cdot cm^{-1}$, $d$ is the path-length in cm, and $C_M$ is the concentration in molar (M). However, for
most work, the term molar ellipticity $[\theta]$ is used to express the CD spectrum. The molar ellipticity at a given wavelength $\lambda$ is then defined as:

$$[\theta]_{\lambda} = \frac{M_w \cdot \theta_{\lambda}}{10^{-1} \cdot d \cdot C} \quad (2.3)$$

where $M_w$ is the molecular weight and $C$ is the concentration in gram per milliliter (g/ml).

Experimentally, the CD spectrometers are usually designed to measure $\Delta \varepsilon$ (or $\Delta A$). At wavelength $\lambda$, the relation between $[\theta]_{\lambda}$ and $\Delta \varepsilon_{\lambda}$ can be obtained easily by combining equations (2.1), (2.2) and (2.3), and is given as:

$$[\theta]_{\lambda} = 3298 \cdot \Delta \varepsilon_{\lambda} \quad (2.4)$$

Notice that the unit of $[\theta]_{\lambda}$ is deg.cm$^{-2}$dmol$^{-1}$ which includes a unitless factor of 100.

It is clear now that CD of a molecule can be observed only when the environment in which a transition occurs is asymmetric. Such a molecule or chromophore is then called optically active.

2.1.3.2 The amide chromophore in peptides

The CD of a peptide is primarily the CD of the amide chromophore, which is the most abundant chromophore in peptides and proteins. The amide group has three $\pi$ orbitals ($\pi_0, \pi_0$, and $\pi^*$), and two lone pair electrons on the carbonyl oxygen. The highest energy lone pair ($n$ orbital) is largely localized on the carbonyl oxygen in a nearly pure 2p orbital with its axis in the amide plane and perpendicular to the carbonyl bond. The other lone pair ($n'$) is at a lower energy which has both 2s and 2p characters and has
its axis directed along the carbonyl bond direction. Figure 2.5 shows the energy level diagram for these orbitals (Woody, 1996).

The well-known $\pi_0 \rightarrow \pi^*$ ($\text{NV}_1$) transition is electrically allowed and the transition moment is directed approximately along the NO direction which occurs near 190nm. The higher energy $\pi_0 \rightarrow \pi^*$ ($\text{NV}_2$) transition has not yet been identified. While the $n' \rightarrow \pi^*$ transition has not been identified, the $n \rightarrow \pi^*$ transition has the lowest transition energy and occurs near 220nm. For polypeptides, the $n\pi^*$ transition lies typically between 215nm and 222nm.

The knowledge of the electronic origin of each band system enables one to use quantum mechanics to calculate oscillator strength associated with it. This, in turn, allows one to evaluate the contribution of those absorption bands to the CD spectrum.

2.1.3.3 Rotational strength of $\alpha$-helix

The rotational strength of the transition from the ground state $0$ to excited state $a$, $R_{0a}$, is defined as a dot product of two vectors $R_{0a} = \text{Im}\{u_{0a} \cdot m_{0a}\}$. Where Im refers to the imaginary part of a complex quantity, and $u_{0a}$ and $m_{0a}$ are the electric and magnetic dipole transition moments, respectively. Theoretically, a particular absorption band can be characterized by its rotational strength (Rosenfeld, 1928); therefore, the rotational strength can be easily related to the ellipticity using the formula:

$$R_{0a} = \frac{hc}{4\pi^2N_0} \int \frac{[\theta_a]}{\lambda} d\lambda$$  \hspace{1cm} (2.5),

where $h$ is the plank's constant ($6.626 \times 10^{-27}$ erg sec), $c$ is the velocity of light ($3 \times 10^{10}$ cm/sec$^{-1}$), $N_0$ is Avogadro's number ($6 \times 10^{23}$), and $[\theta_a]$ is the molar ellipticity of the
midpoint of the absorption band which producing the dichroism. The integrals are taken over the entire electronic band, and the unit for \( R_o \) in cgs is \( \text{erg}\cdot\text{cm}^{-3} \) (Schellman & Oriel, 1962).

In 1956, W. Moffitt first developed a theory to calculate the optical rotation of helical polymers (Moffitt, 1956a). By applying the specific but well defined geometry of \( \alpha \)-helix, the optical rotatory dispersion of helical polymers was predicted (Moffitt, 1956a, b; Moffitt & Yang, 1956). Schellman and Oriel further calculated the rotational strengths of helical polypeptides (Schellman & Oriel, 1962). The experimental results of the CD studies by Holtzwarth and Doty (Holzwarth & Doty, 1965) confirmed those theoretical predictions. More recently, the improved calculation of rotational strength has shown reasonable agreement between the theoretical CD curves and experimental curves for \( \alpha \)-helix (Woody, 1968; Manning & Woody, 1991). In addition, the rotational strength was found to vary significantly with the backbone conformation. For example, the geometry of the hydrogen bond in the \( \alpha \)-helical backbone is the most important factor that influences the dependence of the rotational strength on conformation (Vournakis et al., 1968). Accordingly, CD is very sensitive to the various types of secondary structure and becomes a useful technique in structural studies.

### 2.1.3.4 CD and helix content of helical peptides

It can be expected that each type of secondary structure will have its own characteristic CD spectrum (Greenfield & Fasman, 1969). The reference CD spectra for pure secondary structures shown in Figure 2.6 were measured by Brahms and Brahms
The typical CD spectrum of an α-helix shows a double minimum: one at 222, the other at 208-210 nm; and a maximum at 191-193 nm. CD spectroscopy has become one of the most widely used techniques for measuring helix content. However, this measurement does not define the distribution of helical residues within a peptide, but only provides an estimate of the average helix fraction of a peptide in solution (Gans et al., 1991). The equation used to convert the ellipticity ([θ]_obs) into helix content (f_helix) uses ellipticities of the completely helical state ([θ]_H) and completely random coil state ([θ]_C) is (Scholtz et al., 1995):

\[
\text{f}_{\text{helix}} = \frac{[\theta]_{\text{obs}} - [\theta]_C}{[\theta]_H - [\theta]_C}
\]

(2.6).

It has been shown that the rotational strength of transitions for α-helical peptides is chain length dependent (Woody & Tinoco, 1967; Vournakis et al., 1968). Thus the ellipticity of the full helical state ([θ]_H) at wavelength λ for a complete α helix with n amides at 0°C should be expressed as a function of chain length:

\[
[\theta]^o_{\lambda,h} = [\theta]^m_{\lambda,h} (1 - \frac{k_\lambda}{n})
\]

(2.7),

where \(k_\lambda\) is the chain-length dependent factor at that wavelength, \([\theta]^o_{\lambda,h}\) and \([\theta]^m_{\lambda,h}\) are the mean amide ellipticities of the completely helical n-amide peptide and an infinitely long peptide unit at \(\lambda\), respectively (Chen, Y.-H. et al., 1974; Gans et al., 1991). Also notice that \(n\) is the number of amides in the helix, and a peptide with \(x\) amino acids will have \(x-1\) amides if it is unblocked, \(x\) amides if it is blocked at one end, and \(x+1\) amides if it is blocked at both termini (Woody, 1996).
The stronger $\pi\pi^*$ transition that contributes to 190nm bands has been shown to be more sensitive to chain length than the weaker $\pi\pi^*$ transition that governs the 222nm band (Manning & Woody, 1991). Thus, CD at 222nm is often chosen as the index of helix content. Although this transition has a weak absorption, it has a large magnetic dipole moment directed along the carbonyl bond which results the high rotational strength and strong signal on CD spectrum. Chen et al. gave an $\theta_{222,H}^\infty = -39,500$ and $k=2.57$ from computed CD curves (Chen, Y.-H. et al., 1974). According to the experimental CD data from peptides, ranging from 14 to 50 residues in length, Scholtz et al. came out the value $k=2.5$ using $\theta_{222,H}^\infty = -40,000$ and $\theta_{222,C} = +640$. They further included the temperature dependence of the completely helical state $\theta_{H}$ and completely random coil state $\theta_{C}$ at 222nm (Scholtz et al., 1991b). Their temperature dependent equations are:

$$\theta_{H} = -40,000 + 100 \cdot T$$

(2.8)

$$\theta_{C} = +640 - 45 \cdot T$$

(2.9)

More recent theoretical CD studies lead to values of $k$ ranging from 4.6 to 6.3 for the 222nm $\pi\pi^*$ band at 0°C using $\theta_{222,H}^\infty = -40,000$ (Gans et al., 1991; Manning & Woody, 1991). Experimental CD data for a number of small peptides at 222nm and 0°C suggests that $k=4$, $\theta_{222,H}^\infty = -40,000$, and $\theta_{222,C} = +500$ are also reasonable values (Jackson et al., 1991; Lyu et al., 1991a, b). It is interesting to note that this $k$ value
corresponds approximately to the number of carbonyl groups that do not participate in interamide hydrogen bonding at the end of helix.

2.2 The Two-State model

The Two-State model is a simple thermodynamic treatment of the helix-coil transition theory. This model assumes that the helix-coil transition can be described by a simple two-state transition, Coil ↔ Helix. When applied to protein systems this analysis assumes that at any point on a melting curve, only two types of molecules exist, fully native and fully denatured. Here an intermediate melting value simply reflects the population average of these two states. Thus, any observable property X of the transition would be expressed by only two contributing states as:

\[ X = \frac{K}{(1 + K)} X_H + \frac{1}{(1 + K)} X_C \] (2.10),

where \( X_H \) and \( X_C \) represent the values of the observable property corresponding to 100% helix and 100% coil, respectively (Gans et al., 1991). The equilibrium constant \( K \), \( K = \frac{[\text{helix}]}{[\text{coil}]} \), can be determined by

\[ K = \frac{X - X_C}{X_H - X} \] (2.11).

In terms of ellipticity measured by CD, the equilibrium constant can be expressed as

\[ K = \frac{[\theta]_{\text{obs}} - [\theta]_C}{[\theta]_H - [\theta]_{\text{obs}}} \] (2.12).

At any particular temperature the free energy change due to this transition is then
\[ \Delta G = -RT \ln \left( \frac{[\theta]_{\text{obs}} - [\theta]_c}{[\theta]_H - [\theta]_{\text{obs}}} \right) \] (2.13),

where \( R \) is the gas constant 1.987 cal·K⁻¹·mol⁻¹.

This model describes a straightforward method to estimate the helix-coil transition. However, the discrepancies between the \( \Delta H \) of helix-coil transition determined using CD and that determined by calorimetry have suggested that helix-coil transition in small peptides is a non two-state system (Scholtz et al., 1991a).

Furthermore, NMR studies on model peptides have shown that the end of helices are frayed (Rohl & Baldwin, 1994). In conclusion, the Two-State model provides a baseline analysis but is not appropriate for small peptide systems.

### 2.3 Original Lifson-Roig theory

In 1961, Lifson and Roig established a statistical analysis theory to study the helix-coil transition in polypeptides (Lifson & Roig, 1961). They classified the state of a residue using a simple two-state transition, but determined the contribution of each residue to helix-coil transition in a statistical mechanical model that examined the ensemble nature of the transition. This model describes the state of each amino acid in a sequence using statistical weights for coil probabilities \((u')\), helix nucleation \((v')\), and helix elongation \((w')\). These weights are then combined to form a matrix that represents all possible states of each residue. By multiplying these matrices in the order of peptide sequence, a partition function that describes all the possible states of this peptide system
can be generated. From the partition function many average properties of the system, such as helix content, can be derived.

2.3.1 Conformational state of a residue

The Lifson-Roig theory classifies the state of a residue by its location in $(\phi, \varphi)$ space. The number of residues (N) is defined as the number of $\alpha$-carbons ($C_\alpha$) which are flanked by peptide units (CONH) on both sides. If a residue's $\phi$ and $\varphi$ angles (dihedral angles) are those characteristic of the helix (see section 2.1.1), this residue is then considered to be in a helical state "h". If its dihedral angles are in the remainder of the conformational space, the nonhelical region, then it is in the coil state "c". A polypeptide chain, therefore, can be specified as a sequence of h's and c's such as 

....hhhhccchhhcc... , and there are $2^N$ possible conformations for a molecule consisting of N residues. For example, a peptide with five amino acids in the sequence NH$_3^+$-ABCDE-COO$. Only B, C and D have complete peptide units on both sides of their $C_\alpha$ atoms, and can adopt h or c conformation, thus N=3, so this peptide could have 8 possible conformations: hhh, hhc, hch, chh, hcc, chc, cch, ccc.

2.3.2 Configuration energy, partition function and statistical weights

In general, the energy of the intramolecule hydrogen bonds, together with the interaction between the side-chains, tends to stabilize the $\alpha$-helix structure, while the free energy of the partial freedom of internal rotation and the solvent-polymer interactions
tend to favor the coiled conformation. Therefore, the configurational energy \( V^{(N)} \) for a molecule with \( N \) residues is given by:

\[
V^{(N)} = \sum_{i=1}^{N} V^{(1)}_i (\phi_i, \varphi_i) + \sum_{i=2}^{N-1} V^{(3)}_i (\phi_{i-1}, \varphi_{i-1} , \phi_i , \varphi_i , \phi_{i+1}, \varphi_{i+1}) \tag{2.14}
\]

Where \( V^{(1)}_i (\phi_i, \varphi_i) \) represent the part of the energy of the \( i^{th} \) residue which is independent of other residues. They include the potential of hindered rotation, the energy of solvent-peptide interaction or particular interactions. The \( V^{(3)}_i (\phi_{i-1}, \varphi_{i-1} , \phi_i , \varphi_i , \phi_{i+1}, \varphi_{i+1}) \) term thus represents the energy of formation of the helix, which is assumed to be different from zero if, and only if, all three pairs of internal rotation angles \( (\phi_{i-1}, \varphi_{i-1}) \), \( (\phi_i, \varphi_i) \), and \( (\phi_{i+1}, \varphi_{i+1}) \) are in the region of helix conformation. The attribution of \( V^{(3)} \) to the \( i^{th} \) residue is arbitrary, and could actually be attributed to any of the residues between \( i-2 \) and \( i+2 \).

The configurational partition function, which includes all the possible conformation information of the system, can then be written as

\[
Z = \int_{\phi_1 = 0}^{2\pi} \cdots \int_{\phi_n = 0}^{2\pi} \exp(-\beta \cdot V^{(N)}) d\phi_1 \ldots d\phi_n \tag{2.15}
\]

where \( \beta = 1/(kT) \) (\( k \) is the Boltzmann constant, and \( T \) is temperature). Lifson and Roig divided the integration into two regions, helical and nonhelical, and defined a variable index \( \rho \) which stands for either \( h \) or \( c \) state. Therefore,

\[
\int_{\phi = 0}^{2\pi} \int_{\varphi = 0}^{2\pi} d\phi d\varphi = \sum_{\rho \in [h,c]} \int d\phi d\varphi \tag{2.16}
\]

and the partition function can be rewritten as
\[
Z = \sum_{\phi_n = h,c} \ldots \sum_{\phi_n = h,c} \prod_{i=1}^{i=N} \exp(-\beta \cdot V^{(1)}_i - \beta \cdot V^{(3)}_i) \, d\phi_i \, d\phi_1 \ldots d\phi_N \, d\phi_N \quad (2.17)
\]

In order to factorize the expression of partition function, Lifson and Roig further introduced three statistical weights: \( u', v' \) and \( w' \). Consider a residue \( i \), when integrated over the nonhelical region, the helix interaction term \( V^{(3)}_i \) equals to zero for the entire region and does not contribute to the energy. The statistical weight for coil probabilities, \( u' \), is therefore defined as

\[
u_i = \int_{\text{nonhelical}(\phi_i = c)} \exp(-\beta \cdot V^{(1)}_i) \, d\phi_i \, d\phi_i \quad (2.18)
\]

On the other hand, the integration over the helical region does depend on the state of the adjacent peptide units. When both \( i-1 \) and \( i+1 \) peptide units are in the helix region, the helix interaction term \( V^{(3)}_i \) is different from zero, and the statistical weight for helix probabilities is given as

\[
\nu'_i = \int_{\text{helical}(\phi_i = h)} \exp(-\beta \cdot V^{(1)}_i - \beta \cdot V^{(3)}_i) \, d\phi_i \, d\phi_i \quad (2.19)
\]

The \( \nu'_i \) is known as the helix elongation parameter. If either residue \( i-1 \) and \( i+1 \) is nonhelical, the \( V^{(3)}_i \) term vanishes, so we have the helix nucleation parameter given by

\[
\nu_i = \int_{\text{helical}(\phi_i = h)} \exp(-\beta \cdot V^{(1)}_i) \, d\phi_i \, d\phi_i \quad (2.20)
\]

In summary, there are three statistical weights, coil probability \( u' \), helix nucleation probability \( v' \) and helix elongation probability \( w' \), that represent the conditional probabilities of occurrence of their corresponding events. For practical calculations, \( u' \) is set to 1, which is equivalent to fixing the zero point of the configurational energy \( E^{(0)} \).
As a result, three statistical weights are normalized to 1, v and w. In general, v can be considered as the equilibrium constant for the formation of the helical conformation in the random coil and is considerably less than unity because the region for h state is much smaller than the region for c.

The rule to assign these weights to each residue in a sequence is as follows: (1) a residue in the coil state, i.e. c, always contributes the coil probability $u'$ and its weight is normalized to 1; (2) a residue in the helix state, i.e. h, will contribute the statistical weight $v'$, and is given the factor v. However, if its both sides are flanked by other h residues, then this residue contributes the weight $w'$ and is given the factor w.

For example, consider a 15 residue segment with the following disposition of h and c states: hhhccchhhchcc. The contribution of this conformation to the partition function can then be expressed as $v'w'v'uuv'w'v'uuuv'uu$ or $v'w'v'uuv'w'v'uu$. The product of all the statistical weights ($v^5w^3$) is proportional to the probability of occurrence of that specific peptide conformation.

Now the partition function can be expressed as the sum of all possible conformations:

$$Z = \sum_{\rho_1=h,c} \ldots \sum_{\rho_{n-1}=h,c} y_{\rho_1} x_{\rho_1,\rho_2,\rho_3} \ldots x_{\rho_{n-1},\rho_{n-1},\rho_n} y_{\rho_n}$$

(2.21)

where $y_{\rho_1}$ and $y_{\rho_n}$ are either v or 1 according to whether $\rho_2$ and $\rho_{N-1}$ are h or c.

Similarly, $x_{\rho_{n-1},\rho_{n-1},\rho_n}$ can be either 1 or v or w depends on the conditions of $\rho_{N-2}$, $\rho_{N-1}$ and $\rho_N$. 
2.3.3 Matrix method and properties of the partition function

To simplify the calculation of partition function, Lifson and Roig applied the matrix method to their model. The matrix method has been recognized as a perfect tool to evaluate the complicated partition function and is particularly well adapted to the chain problems like the helix-coil transition in polypeptide (Poland & Scheraga, 1970b). Zimm and Bragg first introduced this method to analyze the helix-coil theory in order to take into account interactions between distant segments as well as nearest neighbors (Zimm & Bragg, 1959).

The rule to assign the statistical weights described above has been expressed by a 4x4 matrix that assigns the weights for the center element of all eight possible triples. The statistical weight matrix for residue i is set up as follows:

\[
\begin{pmatrix}
\bar{h} & \bar{c} & \bar{c} & \bar{c} \\
\bar{h} & w_i & v_i & 0 & 0 \\
\bar{c} & 0 & 0 & 1 & 1 \\
\bar{c} & v_i & v_i & 0 & 0 \\
\bar{c} & 0 & 0 & 1 & 1
\end{pmatrix}
\]

(2.22)

In each triplet, weights are assigned to the center residue which is indicated by a upper bars, a \( \bar{c} \) or \( \bar{h} \). When the states of the center residue of the triplet in the rows and columns are different, the weighting of that triplet must be zero.

Since \( c \) is always given the statistical weight of 1 regardless of its neighbors. The 4x4 matrix can then be reduced to a 3x3 matrix as
Thus the partition function for a polypeptide of \( N \) residues based on Lifson-Roig theory can be written as:

\[
Z = (0 \ 0 \ 1) \prod_{i=1}^{N} \mathbf{J}_i \begin{pmatrix}
0 \\
1 \\
0 \\
1
\end{pmatrix}
\] (2.24),

for the 4x4 matrix representation, or

\[
Z = (0 \ 0 \ 1) \prod_{i=1}^{N} \mathbf{J}_i \begin{pmatrix}
0 \\
1 \\
1
\end{pmatrix}
\] (2.25),

for the 3x3 matrix representation. Notice that the number of residues \( N \), as mentioned before, is defined as the number of \( \mathrm{C}_\alpha \)'s that are flanked by peptide units on both sides.

The first residue (\( i=1 \)) then refers to the second amino acid in the sequence when there is no blocking group, and in like manner, the last residue (\( i=N \)) refers to the \((x-1)^{\text{th}}\) amino acid if there are \( x \) amino acids in the sequence (see section 2.3.1 and Figure 2.7). In addition, if the end residue is in the \( \mathrm{h} \) state, it can only have a \( \nu \)-weighting. And if it is in the \( \mathrm{c} \) state, it will have an unit weighting. Therefore, the initial and end vectors in equation (2.24) and (2.25) ensure that the first and last residues cannot have a \( \omega \)-weighting.
From the partition function many average properties of the system can be calculated. For example, the average number of helical hydrogen bonds in the sequence can be expressed as \( \sum_{i=1}^{N} \frac{\partial \ln Z}{\partial \ln w_i} \) (or \( \frac{\partial \ln Z}{\partial \ln w} \) for homopolymer), since each \( w \)-weighting corresponds to a hydrogen bond (Poland & Scheraga, 1970a). If we wish to calculate a fraction quantity instead of an average quantity, the average number \( <n_h> \) must be divided by the maximum number \( N_h \). Therefore, the fraction of hydrogen bonding for an \( N \) residues peptide is:

\[
f_{\text{H-bond}} = \frac{<n_h>}{N_h} = \frac{\sum_{i=1}^{N} \left( \frac{\partial \ln Z}{\partial \ln w_i} \right)}{N-2}
\]  

(2.26).

Notice that the two termini residues cannot have \( w \) weights, thus the maximum possible hydrogen bond number is \( N_h=N-2 \).

2.4 Modified Lifson-Roig model

In original Lifson-Roig theory, the number of residues is defined as the number of \( \alpha \)-carbons which are flanked by peptide units (CONH) on both sides. When the N-terminus is an amino group, the first amino acid is not considered as a residue and makes no contribution to the partition function and, in turn, should make no contribution to overall helicity. However, the experimental results have shown that large differences in helix content are found when the first amino acid is altered in unacylated peptides (Chakrabartty et al., 1993a). Since the original Lifson-Roig theory predicted the
modification of the first amino acid would have no effect on helix content, the original Lifson-Roig model was modified.

2.4.1 Definition for "number of residues" in modified model

Doig et al. compensated the above problem by introducing both N- and C-capping effects into the Lifson-Roig theory (Doig et al., 1994). In this modified theory, it has been assumed that every amino acid makes a contribution to the partition function. Therefore, rather than having only the center three amino acids considered as residues, the peptide NH₃⁺-ABCDE-COO⁻ will have all five amino acids included in the partition function. That is, the term "residue" used in the modified Lifson-Roig theory refers to any amino acid in the sequence, not just the non-terminal amino acids. In addition, since acetylation of the N-terminus and amidation of the C-terminus introduce extra peptide units that can contribute to the partition function, these groups can also be considered as residues. Thus, the peptide with blocking groups Ac-ABCDE-CO-NH₂ is considered to have seven residues, whereas original theory would include only five units.

2.4.2 Incorporation of capping effects into statistical weight matrix

2.4.2.1 n- and c-weights introduced by Doig et al.

To introduce N- and C-capping effects, Doig et al. further modified the triplet definitions given in the statistical weight matrix (Doig et al., 1994). A c-state residue that immediately precedes an h-state residue is defined as the N-cap residue, and given a weighting of n. The C-cap residue, which is the c-state residue immediately after an h-
state residue, is then given a weighting \( c \). A c-state between two h-states is assumed to have the geometric mean of the n- and c- weightings, namely \( \sqrt{nc} \). Other weightings remain the same. The weighting of center residues of the eight possible triplets for original and modified Lifson-Roig matrix are compared in Table 2.1. Notice that if \( n = 1 \) and \( c = 1 \), the modified theory reduced to the original theory.

Now the modified Lifson-Roig matrix for residue \( i \) is written as:

\[
I_i = \begin{pmatrix}
\bar{hh} & \bar{hc} & \bar{ch} & \bar{cc} \\
\bar{hh} & w_i & v_i & 0 & 0 \\
\bar{hc} & 0 & 0 & \sqrt{n_i} & c_i \\
\bar{ch} & v_i & v_i & 0 & 0 \\
\bar{cc} & 0 & 0 & n_i & 1
\end{pmatrix}
\]  

and the partition function for an \( N \)-residue sequence becomes:

\[
Z = (0 \ 0 \ 0 \ 1) \prod_{i=1}^{N} I_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}
\]  

(2.28).

Notice that the N-terminal residue does not have \( \phi \) angle (the \( \text{NH}_2^- - \text{C}_a \text{CO} \) bond or \( \text{CH}_3\text{CO} \) bond for acetyl blocking group is free rotation), and the C-terminal residue does not have a \( \varphi \) angle (the \( \text{C}_a - \text{COO}^- \) bond or \( \text{CONH}_2 \) bond for amide blocking group has free rotation); that is, the end units cannot be in an h state since the h state requires both \( \phi \) and \( \varphi \) dihedral angles to be restricted into a helical geometry. The initial vector in equation (2.28) ensures that the N-terminal residue can only have the weight \( n \) or \( 1 \) depending on the state of the next residue. Likewise, the end vector restricts the C-terminal residue can only have the weight \( c \) or \( 1 \).
2.4.2.2 Revising of capping weights by Rohl et al.

Rohl et al. have noted one discrepancy in the above capping model: helix capping was being included at the ends of segments that were not helical. Take, for example, the sequence cchhcchcc, while this peptide contains some helical residues, it contains no helix because a helix would require the presence of at least three consecutive helical residues (hhh). The capping statistical weights defined in section 2.4.2.1 are applied not only to helical segments but also to coil segments, like the example, and therefore overestimate the amount of helix capping. Rohl et al. modified the description of the capping interaction to eliminate the contribution of the capping weights to coil conformations (Rohl et al., 1996).

In this model, the N-cap residue only occurs when a c-state residue immediately precedes at least three consecutive h-state residues. That is, it is the first residue in a helical segment with a hydrogen-bonded CO. Likewise, the C-cap residue is a c-state residue immediately after at least three consecutive h-state residues. That is, it is the last residue in a helical segment with a hydrogen-bonded NH (Rohl et al., 1996). Consequently, capping interactions are assumed to occur only when a helical segment of at least one hydrogen bond is formed. The rule to assign statistical weights using different capping models is summarized in Table 2.2. To identify the N-cap and C-cap units, a quintet rather than a triplet of residues is required. The statistical weights matrix for residue i in the middle of the quintet has been reduced to a 6x6 matrix K:
Row labels represent the states for residues i−2, i−1, i and i+1, and column labels represent the states for residue i−1, i,i+1 and i+2. Thus the partition function becomes:

\[
Z = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 1 \\
\end{pmatrix} \prod_{i=1}^{N} K_i \tag{2.30}
\]

### 2.4.3 Adding side chain interactions to the modified Lifson-Roig model

A recent statistical study of protein crystal structures has revealed a strong correlation for specific pairs of amino acids spaced at i and i+4 in α-helices. Especially dominant are interactions that involve salt bridges or hydrophobic interactions (Klingler & Brutlag, 1994). To introduce these factors, the Lifson-Roig model has been further modified (Shalongo & Stellwagen, 1995; Stapley et al., 1995).

Stapley et al. incorporated i, i+3 and i, i+4 side chain interactions into the modified Lifson-Roig model by including two additional parameters, p and q, that can be...
considered as equilibrium constants for formation the respective interactions in the helix.

In their definition, the interaction occurs only when the interacting residues and all the intervening residues are in h state. \( p \) is the \( i, i+4 \) interaction parameter, and \( w_{i,p-i-2} \) will be assigned to the central residue in the \( \text{hhh} \) quintet when the first \((i)\) and last \((i+4)\) residues are interacting. Likewise, \( q \), the \( i, i+3 \) interaction parameter will be applied to the central residue of \( \text{hhhc} \) as \( w_{q_{i-2}i} \) when the first \((i)\) and fourth \((i+3)\) residues interact with each other. When both \( i, i+3 \) and \( i, i+4 \) interactions occur, then the central residue of the \( \text{hhhh} \) quintet is assigned the weight \( w_{p_{i-2},2q_{i-2},2i+1} \). In addition, the statistical weight \( r_{i-2,i+1} \) is applied to the central residue \( i \) in the quintets \( \text{chhh} \) and \( \text{chhh} \) due to the interactions occurring between the N-cap residue \( i-2 \) and residue \((i-2)+3 \) (Rohl et al., 1996). Thus, equation (2.29) is modified as:

\[
L_i = \begin{pmatrix}
\text{hhhh} & \text{hhhc} & \tilde{d}_w(c\cdot h) & (c\cdot h)\tilde{d}_w(c\cdot h) & (c\cdot h)\tilde{d}_w(c\cdot h) & (c\cdot h)\tilde{d}_w(c\cdot h)
\end{pmatrix}
\begin{pmatrix}
w_{p_{i-2,2q_{i-2},2i+1}} & w_{c_{i-2},q_{i-2},2i+1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & v & 0 & 0 \\
0 & 0 & 0 & 0 & v & 0 \\
w_{q_{i-2},2q_{i-2},2i+1} & w_{q_{i-2},c_{i-2},f_{i-2},2i+1} & 0 & v & 0 & 0 \\
0 & 0 & v & 0 & 0 & 1 \\
0 & 0 & 0 & v & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\]

(2.31).

When applying the modified Lifson-Roig model, the two terminal residues can exist only in the c state, thus the second and the \((N-1)\)th residues can only have the weight of \( v \) (if in h state) or 1 (if in c state).
The maximum possible hydrogen bond number (i.e. the maximum number of residues that could have \( w \) weighting) is then \( N_h = N - 4 \), where \( N \) (number of residues) refers to the total number of amino acids plus blocking groups, if there are any, in the sequence. The fraction of hydrogen bonding for a peptide becomes:

\[
f_{\text{H-bond}} = \sum_{i=1}^{N} \left( \frac{\partial \ln Z}{\partial \ln w_i} \right) \frac{N - 4}{N}
\]

In addition to the fraction of hydrogen bonding, the probability that a given \( i, i+3 \) or \( i, i+4 \) interaction is present can be determined from the partition function as well. The free energy change due to these interactions are simply \( \Delta G_{(i, i+3)} = -RT \ln q \) or \( \Delta G_{(i, i+4)} = -RT \ln p \).

2.4.4 Application of the Lifson-Roig model

The Lifson-Roig model offers hope for predicting helical structure in an unknown peptide from its primary sequence. In order to predict the helical structure using Lifson-Roig statistics, the intrinsic parameters (statistical weights) first must be determined. This can be done by comparing the experimental data to theoretically calculated results, and using optimization methods to obtain the best-fit parameters.

2.4.4.1 Calculation of ellipticity from helix content

For most work, it has been assumed that small peptides contain only one helical segment (Scholtz et al., 1991b; Chakrabarty et al., 1994; Doig et al., 1994; Doig &
Baldwin, 1995; Stapley et al., 1995), thus by rearranging equation (2.6), the calculated ellipticity at 222 nm has been expressed as:

\[
\theta_{\text{cal}}^n = f_{H,\text{cal}} \cdot \theta_{n}^n + (1 - f_{H,\text{cal}}) \cdot \theta_c
\]

(2.33),

where \([\theta]_n^n\) is the chain-length dependent ellipticity for a completely helical state and can be determined using equation (2.7), \([\theta]_c\) is the ellipticity for a completely random coil state, and \(f_{H,\text{cal}}\) is the calculated helix content which could be derived from partition function. Thus the calculated ellipticity, as a function of intrinsic parameters \((w, v, n \text{ and } c)\) of each component amino acid, can be obtained.

It has also been assumed that all amino acids have the same value of helix nucleation parameter \(v\) because it is a property of the peptide backbone. The experimental results for the kinetics of amide proton exchange in several helical peptides have been used to determine the \(v\) value (Rohl et al., 1992). Thus, \(v\) is given a constant value. By floating the other parameters \((w, n, \text{and } c)\) in the equation for calculated CD (i.e. equation (2.33)) until the sum of the residuals squared between calculated CD values \(([\theta]_{\text{cal}})\) and experimental data \(([\theta]_{\text{obs}})\) at 222nm reach a minimum, the best-fit parameters can be obtained.

2.4.4.2 Results and discussion

Several sets of parameters can be found in literature (Chakrabarty et al., 1994; Doig et al., 1994; Doig & Baldwin, 1995; Stapley et al., 1995; Rohl et al., 1996). The most recent published intrinsic parameter library (Table 2.3) was obtained by Rohl et al. (Rohl et al., 1996). They used \([\theta]_c = +640, [\theta]_n^n = -42,500, \text{ and } k=4\) for unblocked
peptides or k=3 for peptides with C terminal amide blocking groups, to do their calculations. Since the fraction of hydrogen bonding (\( f_{\text{ct-bond}} = \langle n_h \rangle / N_h \)) has been considered a good measure of the helix content when CD at 222nm is used to follow the helix-coil transition (Qian & Schellman, 1992), the calculated helix content (fn,cal) in Rohl et al.'s calculation was assumed to be the same as fn,bond.

In Rohl et al.'s analysis, the helix content was calculated from the partition function using the equation:

\[
fn,\text{cal} = \frac{n_h}{N_h} = \frac{\sum_i \left( \frac{\partial \ln Z}{\partial \ln w_i} \right)}{N - 4}
\]

The ellipticity was calculated based on equations (2.33) and (2.34). After they obtained a best-fit intrinsic parameter library, a computer program implementing the modified Lifson-Roig model written in FORTRAN was provided to predict the helix content for a given sequence (anonymous FTP from cmgm.stanford.edu in the directory /pub/helix/helix2).

However, there are two major problems in this application. First, the chain length dependence of the CD signal has been calculated improperly. When using equation (2.7) to calculate the ellipticity for the completely helical state, the total number of amides in the sequence has been considered as the "n" factor in this equation. This is true only when all the residues in the sequence are in the helical segment (Chen, Y.-H. et al., 1974). When applying the Lifson-Roig theory, there is only one state in which all the residues in the sequence are helical. For all the other states, the number of amides that
are helical is less than the total number of amides in the sequence. For example, consider a 17 residues peptide, if one of its possible states contains only 9 helical residues (such as cccchhhhhhhhhcc) then, instead of applying \([\theta]_H^n = [\theta]_H^o (1 - k_\lambda/16)\) to equation (2.23), \([\theta]_H^n = [\theta]_H^o (1 - k_\lambda/8)\) should be used for that conformation. That is, instead of using the total number of amides in the sequence, the "n" factor in the equation should refer to the number of amides in each helical segment (Chen, Y.-H. et al., 1974), and then sum over the conformational ensemble. This error will be further accentuated by the fact that the CD signal is calculated from the sum of the ensemble of states that have varying chain length (i.e. the "n" factor). Additionally, the possibility of multiple helical segments in a sequence is not appropriately considered in this application as well.

Second, it is assumed that only residues having the w-weighting (i.e. are fully hydrogen bonded) contribute to the CD at 222 nm. However, residues with v-weighting are also helical and should contribute to the ellipticity at 222nm (Doig et al., 1994). Therefore, the assumption that the helix content is equal to the fraction hydrogen bonds and is directly related to CD signal at 222nm must be considered more carefully.

Gans et al. has suggested an algorithm to predict the helix structure (Gans et al., 1991). Their approach examines the chain length dependence of the ellipticity more carefully and also takes into account the distribution of the number of helical segments in a chain. They computed the mean amide ellipticity for a chain with n amides using

\[
[\theta]_{\text{cal}}^n = [\theta]_{\text{cal}}^o \sum_r (\text{Prob}(r) \times (1 - \frac{k_\lambda}{r}))
\]

(2.35),

33
where \( \text{Prob}(r) \) stands for the probability that a segment contains \( r \) amides. More recently, Shalongo and Stellwagen introduced a statistical model called the Dichroic model that not only recognizes the contributions of each helical segments in the conformational ensemble but also considers the source of ellipticity at 222nm in more detail (Shalongo & Stellwagen, 1997).

2.5 Dichroic model

2.5.1 Contribution of a helical ensemble to ellipticity

In the Dichroic model, the mean peptide bond ellipticity for a peptide with \( n \) amide bonds \( ([\theta]_c)^n \) is expressed as

\[
[\theta]_c^n = \sum_k f_k [\theta]_k
\]

where \( k \) is a counter of conformation, \( f_k \) is the fractional concentration (or probability) of the \( k^{th} \) conformation, and \( [\theta]_k \) is the mean peptide bond ellipticity for the \( k^{th} \) conformation. Notice in this model the term "mean peptide bond ellipticity" is used to emphasize the fact that it is the amide bond, not an amino acid residue, that contributes to the CD spectrum.

The contribution of ellipticity at 222nm has been studied theoretically. The principal contributor to \( [\theta]_k \) at 222nm is the \( n\pi^* \) transition in peptide, which is primarily a transition of lone pair electrons in the electronic environment of \( C=O \ldots N \) group. Strong hydrogen bonds, such as the backbone/backbone hydrogen bonds present in the helical conformation, constrain the \( C=O \ldots N \) angle and provide larger separation in
energy level than the weaker backbone/solvent hydrogen bonds in the coil conformation. Therefore, the amide groups involved in backbone hydrogen bonds will contribute more CD signal at 222 nm than the non-hydrogen bonded amide groups. Since a single peptide bond may be involved in two hydrogen bonds, the CD at 222 nm will vary depending on the number of hydrogen bonds formed with a given peptide unit.

Thus, the mean peptide bond ellipticity of a particular conformation k for a peptide molecule ([θ]k) should be expressed as a sum of terms of peptide bonds involved in zero, one or two hydrogen bonds:

\[
[\theta]_k = \frac{b_c[\theta]_c + b_{h1}[\theta]_{h1} + b_{h2}[\theta]_{h2}}{b}
\]  

(2.37),

where \(b_c\), \(b_{h1}\) and \(b_{h2}\) are number of peptide bonds in the coil state (c) (i.e. no hydrogen bond), in the helix state with one backbone/backbone hydrogen bond (h1), and in the helix state containing two backbone/backbone hydrogen bonds (h2), respectively. And \(b\) represents the sum of these three terms, i.e. \(b = b_c + b_{h1} + b_{h2}\), and is equal to the total number of peptide bonds in the molecule. The corresponding \([\theta]_c\), \([\theta]_{h1}\), and \([\theta]_{h2}\) terms are the mean peptide bond ellipticities for a peptide in a completely random coiled conformation, in a completely helical conformation participating in one helical backbone/backbone hydrogen bond, and in a completely helical conformation participating in two helical backbone/backbone hydrogen bond, respectively.

Accordingly, all possible helical segments are considered in this model.
2.5.2 Incorporation into Helix-Coil theory using enumeration method

Using the matrix method to generate the partition function of a system, and then calculating the average properties from the partition function, is a useful approach when the desired properties depend only on a single parameter. For example, when applying the Lifson-Roig theory, the chain length dependent term ([θ]_θ) is assumed to depend only on the total number of amides in a given sequence, and the matrix method is a perfect approach. However, in the Dichroic model, chain length dependent properties like CD have to be considered more carefully. Here, properties do not depend simply on the total number of amides in the peptide, but on the actual distribution of those amides along the sequence. In this situation, where one must examine an entire structural ensemble, and not just one average number, the matrix method is less useful and an enumeration method becomes necessary to be able to evaluate every member of the structural ensemble.

In the enumeration approach, every state of an ensemble is identified and enumerated. This is easily done in a computer using binary notation where 1 represents a helical residue (h) and 0 represents a non-helical residue (c). Here a sequence described using residue conformation ......hhhhccc hhcc ...... is written as ......11100011100...... using the binary notation. As each state is enumerated, an appropriate statistical weights (w, v, n or c) can be assigned based on the conformational state of the residue and its neighbors. The probability of occurrence of a specific peptide conformation (Z_k) is then the product of its residue statistical weights.
2.5.2.1 Partition function and the fractional concentration

In the enumeration method, enumeration of all states is accomplished by simply counting in binary from 0 to a given number. That is, each possible peptide conformation can be represented simply by a decimal integer. For example, a binary notation conformation sequence 01110 is equivalent to decimal integer 14. Therefore, counting from 0 for the completely coiled conformation and ending with $2^{N-1} - 2$ for the perfectly helical conformation will generate all possible conformations for the system. Notice that because the last residue in the sequence cannot be helical (i.e. not 1), only the even integers need to be evaluated. Also, since the first residue cannot be helical as well, we only have $2^{N-2}$ instead of $2^N$ possible conformations. Table 2.4 shows an example for a sequence with five residues. The partition function ($Z$) for the entire ensemble is then the probability sum over all possible peptide conformation ($\sum Z_k$), and the fractional concentration of a certain conformation ($f_k$) can be calculated by

$$f_k = \frac{Z_k}{Z} = \frac{Z_k}{\sum Z_k} \quad (2.38).$$

The enumeration method is not as fast as the matrix method when calculating the partition function; however, it provides the information for each possible conformation separately. This cannot be obtained from matrix method, and is necessary for evaluation total CD in the Dichroic model. In addition, since the structural details of each conformation are identified, it is easy to expand this approach to include additional interactions involved in the helix-coil transition, such as different types of side chain interactions.
2.5.2.2 *Fractional population and calculated ellipticity*

As described above, the enumeration method can generate binary notation for each conformation individually. Thus, by carefully examining the fractional population \((b_C, b_{hi}, b_{hz})\) through the whole sequence for each specific conformation, the intrinsic dichroic signal \(\langle \theta \rangle_k\) for that conformation can be calculated using equation (2.37). In combination with the fractional concentration derived from equation (2.38), the calculated CD for the whole peptide defined by equation (2.36) can be obtained.

2.5.3 *Application of Dichroic model*

The Dichroic model is a better model that describes the helix-coil transition theory more precisely. In order to correlate the Dichroic model to helix-coil transition, the intrinsic parameters for all amino acids must be generated (see Chapter 3, section 3.5.3). After the parameters are obtained, the ellipticity at 222nm for a peptide can then be predicted from the given sequence.

Shalongo et al. assumed \(\langle \theta \rangle_c\) equal to zero for 222nm at 0°C, and used \[\langle \theta \rangle_{h2} = \langle \theta \rangle_{h1} = -40,000 \text{ (deg cm}^2 \text{dmol}^{-1} \text{ per peptide bond)}\] in their calculations. For a perfect helix, there is only one complete helical segment, thus \(f_k = 1, b_C = 0, b_{hi} = 6,\) and \(b_{hz} = b - 6.\) Under these conditions, equation (2.7) should be equivalent to equation (2.36).

By replacing the \(\langle \theta \rangle_k\) in equation (2.36) with equation (2.37), the relation between \(\langle \theta \rangle_{h1}\) and \(\langle \theta \rangle_{h2}\) in terms of \(k_a\) can be obtained: \[\langle \theta \rangle_{h2}/\langle \theta \rangle_{h1} = (6-k_a)/6.\] Thus, when \(k\) is set at 4.32, \(\langle \theta \rangle_{h1} = -11,200 \text{ (deg cm}^2 \text{dmol}^{-1} \text{ per peptide bond)}\) (Shalongo & Stellwagen, 1997).
Notice that in the original Dichroic model, the amide blocking group at the C-terminal end of the peptide is not included in the residue count, while acetyl group at the N-terminal end is (Shalongo & Stellwagen, 1997). According to the original model, an acetyl blocking group adds an extra peptide bond and a pseudo C\(_\alpha\) atom to the sequence, and is thus considered a residue. On the other hand, the amide carboxyl blocked terminal lacks the pseudo C\(_\alpha\) carbon from the next residue, does not have the same electronic environment as other peptide bonds and is not considered a residue.

We believe this is incorrect. The addition of an amide blocking group restricts the \(\phi\) angle of the last amino acid in the sequence and provides the opportunity for the last amino acid to be in a \(\alpha\) state, just like the acetyl group at the other end would restrict the \(\phi\) angle of the first amino acid. Besides, both acetyl and amide blocking groups add extra potential hydrogen bonds to the peptide that can stabilize the helix. Thus, the amide blocking group, like the acetyl blocking group, will be considered as a residue in our calculations. Accordingly, the same definition for the residue number N as described in modified Lifson-Roig model (section 2.4.1) will be used in our enumeration calculation, where every amino acid and blocking group is counted as a residue, and the total number of peptide bonds \(b\) is then equal to \(N-1\).

In addition, the original Dichroic model used the N- and C- capping definition based on Doig et al. (section 2.4.2.1). We, however, will use the modified definition given by Rohl et al. (section 2.4.2.2) to do the analysis. Figures 2.7 and 2.8 summarize the residue numbering and statistic weight assignment for different models.
<table>
<thead>
<tr>
<th>Residue conformation</th>
<th>Statistical weight of center residue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original Lifson-Roig Matrix</td>
</tr>
<tr>
<td>hhh</td>
<td>w</td>
</tr>
<tr>
<td>hhc</td>
<td>v</td>
</tr>
<tr>
<td>hch</td>
<td>l</td>
</tr>
<tr>
<td>hcc</td>
<td>l</td>
</tr>
<tr>
<td>chh</td>
<td>v</td>
</tr>
<tr>
<td>chc</td>
<td>v</td>
</tr>
<tr>
<td>cch</td>
<td>l</td>
</tr>
<tr>
<td>ccc</td>
<td>l</td>
</tr>
</tbody>
</table>

Table 2.1: The weightings of center residues for the triplet: Comparison between original Lifson-Roig matrix (Lifson & Roig, 1961) and the modified Lifson-Roig matrix (Doig & Baldwin, 1995).
<table>
<thead>
<tr>
<th>Residue conformation</th>
<th>Statistical weight for the center residue</th>
<th>Residue conformation</th>
<th>Statistical weight for the center residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>cccccc</td>
<td>1</td>
<td>hccccc</td>
<td>1</td>
</tr>
<tr>
<td>cccch</td>
<td>1</td>
<td>hccch</td>
<td>1</td>
</tr>
<tr>
<td>ccchc</td>
<td>$n_i$</td>
<td>hcchc</td>
<td>$n_i$</td>
</tr>
<tr>
<td>ccchh</td>
<td>$n_i$</td>
<td>hcchh</td>
<td>$n_i$</td>
</tr>
<tr>
<td>cchcc</td>
<td>$v_i$</td>
<td>hchcc</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>cchch</td>
<td>$v_i$</td>
<td>hchch</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>cchhc</td>
<td>$v_i$</td>
<td>hchhc</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>cchhh</td>
<td>$v_i$</td>
<td>hchhh</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>checc</td>
<td>$c_i$</td>
<td>hhecc</td>
<td>$c_i$, $1$</td>
</tr>
<tr>
<td>chech</td>
<td>$c_i$</td>
<td>hhech</td>
<td>$c_i$, $1$</td>
</tr>
<tr>
<td>chche</td>
<td>$\sqrt{n_i, n_i}$</td>
<td>hhche</td>
<td>$\sqrt{n_i, n_i}$</td>
</tr>
<tr>
<td>chchh</td>
<td>$\sqrt{n_i, n_i}$</td>
<td>hhchh</td>
<td>$\sqrt{n_i, n_i}$</td>
</tr>
<tr>
<td>chhcc</td>
<td>$v_i$, $v_i$</td>
<td>hhccc</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>chhch</td>
<td>$v_i$, $v_i$</td>
<td>hhchc</td>
<td>$v_i$, $v_i$</td>
</tr>
<tr>
<td>chhhc</td>
<td>$w_i$, $w_i n_i - 2 c_i - 2$</td>
<td>hhhhc</td>
<td>$w_i$, $w_i c_i - 2$</td>
</tr>
<tr>
<td>chhhhh</td>
<td>$w_i$, $w_i n_i - 2$</td>
<td>hhhhh</td>
<td>$w_i$, $w_i$</td>
</tr>
</tbody>
</table>

Table 2.2: Corresponding statistical weights for the center residue of a quintet of residues using Doig et al.'s definition (Doig & Baldwin, 1995) and Rohl et al.'s correction (Rohl et al., 1996).
<table>
<thead>
<tr>
<th>Residue code</th>
<th>w</th>
<th>v</th>
<th>n</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>1.70</td>
<td>0.036</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>GLU</td>
<td>0.70</td>
<td>0.036</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CYS</td>
<td>0.33</td>
<td>0.036</td>
<td>5.40</td>
<td>1.00</td>
</tr>
<tr>
<td>ASP-</td>
<td>0.38</td>
<td>0.036</td>
<td>6.60</td>
<td>1.00</td>
</tr>
<tr>
<td>GLU-</td>
<td>0.54</td>
<td>0.036</td>
<td>2.06</td>
<td>1.00</td>
</tr>
<tr>
<td>PHE</td>
<td>0.27</td>
<td>0.036</td>
<td>2.06</td>
<td>1.00</td>
</tr>
<tr>
<td>GLY</td>
<td>0.04</td>
<td>0.036</td>
<td>3.90</td>
<td>1.00</td>
</tr>
<tr>
<td>HIS+</td>
<td>0.22</td>
<td>0.036</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ILE</td>
<td>0.46</td>
<td>0.036</td>
<td>1.57</td>
<td>1.00</td>
</tr>
<tr>
<td>HIS</td>
<td>0.36</td>
<td>0.036</td>
<td>2.12</td>
<td>1.00</td>
</tr>
<tr>
<td>LYS+</td>
<td>1.00</td>
<td>0.036</td>
<td>0.72</td>
<td>1.00</td>
</tr>
<tr>
<td>LEU</td>
<td>0.87</td>
<td>0.036</td>
<td>2.06</td>
<td>1.00</td>
</tr>
<tr>
<td>MET</td>
<td>0.65</td>
<td>0.036</td>
<td>1.31</td>
<td>1.00</td>
</tr>
<tr>
<td>ASN</td>
<td>0.29</td>
<td>0.036</td>
<td>6.80</td>
<td>1.00</td>
</tr>
<tr>
<td>ASP</td>
<td>0.40</td>
<td>0.036</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PRO</td>
<td>0.00</td>
<td>0.036</td>
<td>1.35</td>
<td>1.00</td>
</tr>
<tr>
<td>GLN</td>
<td>0.62</td>
<td>0.036</td>
<td>0.12</td>
<td>1.00</td>
</tr>
<tr>
<td>ARG</td>
<td>1.14</td>
<td>0.036</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SER</td>
<td>0.40</td>
<td>0.036</td>
<td>3.90</td>
<td>1.00</td>
</tr>
<tr>
<td>THR</td>
<td>0.18</td>
<td>0.036</td>
<td>2.23</td>
<td>1.00</td>
</tr>
<tr>
<td>LYS</td>
<td>0.00</td>
<td>0.036</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAL</td>
<td>0.25</td>
<td>0.036</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>TRP</td>
<td>0.29</td>
<td>0.036</td>
<td>3.60</td>
<td>1.00</td>
</tr>
<tr>
<td>ACE</td>
<td>0.00</td>
<td>0.036</td>
<td>5.90</td>
<td>1.00</td>
</tr>
<tr>
<td>TYR</td>
<td>0.48</td>
<td>0.036</td>
<td>4.90</td>
<td>1.00</td>
</tr>
<tr>
<td>AMD</td>
<td>0.00</td>
<td>0.036</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 2.3: The intrinsic parameter library for all residues in water: generated using modified Lifson-Roig model (Rohl et al., 1996).
<table>
<thead>
<tr>
<th>Possible conformations using h or c notation</th>
<th>Possible conformations using Binary notation</th>
<th>Corresponding decimal integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccccc</td>
<td>00000</td>
<td>0</td>
</tr>
<tr>
<td>ccchc</td>
<td>00010</td>
<td>2</td>
</tr>
<tr>
<td>cchcc</td>
<td>00100</td>
<td>4</td>
</tr>
<tr>
<td>cchhc</td>
<td>00110</td>
<td>6</td>
</tr>
<tr>
<td>chccc</td>
<td>01000</td>
<td>8</td>
</tr>
<tr>
<td>chchc</td>
<td>01010</td>
<td>10</td>
</tr>
<tr>
<td>chhcc</td>
<td>01100</td>
<td>12</td>
</tr>
<tr>
<td>chhhc</td>
<td>01110</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2.4: All possible peptide conformations for a sequence with five residues and their corresponding decimal integer.
\( \phi_i = 0 \) when the atoms \( C_{i+1} \) and \( C_i \) are trans about the bond \( N_i - C_{\alpha i} \);

\( \varphi_i = 0 \) when the atoms \( N_i \) and \( N_{i-1} \) are trans about the bond \( C_{\alpha i} - C'_{i} \);

\( \omega_i = 0 \) when the atoms \( C_{\alpha i} \) and \( C'_{i-1} \) are trans about the bond \( C'_{i} - N_{i-1} \).


Figure 2.1: Definition of torsion angles \( \phi \), \( \varphi \), and \( \omega \) for peptide backbone.
\[ \phi \] psi

\( \alpha \) : right handed \( \alpha \)-helices
\( \beta \) : \( \beta \) strands
\( L \) : left handed \( \alpha \)-helics

(From Branden, Carl. and Tooze, John, Introduction to Protein Structure, p. 9, 1991 by Garland Publishing, Inc.).

Figure 2.2: Ramachandran plot for poly-L-alanine.
Figure 2.3: Idealized diagram of the main chain in an α-helix, where the arrow indicates the direction from N terminal to C terminal.
A: If the right (E_R) and the left (E_L) waves have the same amplitudes, the result is a plane-polarized light;

B: If the right (E_R) and the left (E_L) waves have different amplitudes, the resulting light is elliptically polarized, that is, the head of the resultant vector will trace the ellipse shown as the dashed line.

(From Freifelder, David, Physical Biochemistry, Application to Biochemistry and Molecular Biology, 2nd Ed., p. 578, 1982 by W.H. Freeman and Company)

Figure 2.4: Plane-polarized light and elliptically polarized light.

Figure 2.5: Energy level diagram for amide chromophore.
solid line: α-helix; long dashed line: antiparallel β-sheet; 
dotted line: β-turn; short dashed line: random coil.


Figure 2.6: CD spectra for pure secondary structures.
Figure 2.7: Residue numbering and statistic weight assignment for different models: a sequence with free amino acid termini.
Amino acid numbering

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th></th>
</tr>
</thead>
</table>

Original LR model residue numbering

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th></th>
</tr>
</thead>
</table>

Modified LR model residue numbering

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
</table>

Original Dichroic model residue numbering

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th></th>
</tr>
</thead>
</table>

Modified Dichroic model residue numbering

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
</table>

Residue conformation code

```
ccchhhcc
```

Original LR model statistic weights

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>v</th>
<th>w</th>
<th>v</th>
<th>l</th>
<th>l</th>
<th></th>
</tr>
</thead>
</table>

Modified LR model and Original Dichroic model statistic weights

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>n</th>
<th>v</th>
<th>w</th>
<th>v</th>
<th>c</th>
<th>l</th>
<th>1</th>
</tr>
</thead>
</table>

Capping modified LR model and modified Dichroic model statistic weights

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>v</th>
<th>w<em>n</em>c</th>
<th>v</th>
<th>l</th>
<th>l</th>
<th>1</th>
</tr>
</thead>
</table>

Figure 2.8: Residue numbering and statistic weight assignment for different models: a sequence with acetyl blocked N terminal and amide blocked C terminal.
CHAPTER 3

TECHNIQUES AND METHODS

3.1 Peptide synthesis

3.1.1 Solid phase peptide synthesis

In solid phase peptide synthesis, the C terminus of the peptide chain is covalently bound to a solid support during synthesis. Three chemical reactions: deprotection, activation and coupling are then repeated for each amino acid that is added to the growing peptide chain (Figure 3.1).

Most peptides were synthesized on an Applied Biosystems 432A Synergy peptide synthesizer. The Synergy system uses Fmoc (9-fluorenylmethyloxycarbonyl) protected amino acids and a lightly cross-linked polystyrene ($C_6H_5CH=CH_2$) resin in the form of tiny beads as its solid support. Some amino acids have additional side-chain protecting groups. The Fmoc group is removed by piperidine in combination with DMF (N,N-dimethylformamide) to make the α-amino group on the peptide resin accessible for chemical reaction with an activated amino acid during the "deprotection" step (Figure 3.2). HBTU (2-(1-H-benzotriazol-1-yl)-1,1,3,3-tetramethyllumonium hexafluorophosphate) in combination with the base DIEA (N,N-diisopropylethylamine)
rapidly converts the carboxyl group to an active ester in "activation" step (Figure 3.3). This active ester will couple to the deprotected α-amino group during "coupling" step. All reactions are carried out in DMF, and at the end of synthesis, the resin is washed with THF (tetrahydrofuran) to displace the primary solvent. Pressurized gas then removes THF, and dries the resin.

3.1.1.1 Amidation and acetylation of peptide termini

Amidation of the C-terminal amino acid for each peptide was accomplished by starting the synthesis with an amide linked resin. In a typical synthesis, 25 umol of amide resin was used. After the synthesis was complete, the peptide's N-terminus was acetylated using the following protocol (Stewart & Young, 1984):

a. Open the peptide synthesis column (resin) and pour the resin into a 50ml conical plastic culture tube;

b. Add acetic anhydride (0.47 ml or 5 mmol), TEA (triethylamine, 0.70 ml or 5 mol) and DMF (15 ml) to the resin and stir on an orbital shaker for 2 hours at room temperature;

c. Centrifuge the resin to the bottom of the tube and vacuum off the mixture;

d. Wash the resin twice with 10 ml DMF, centrifuge the resin and remove DMF.

e. Place the resin into lyopholyzer for 30 minutes to remove excess DMF.

3.1.1.2 Cleavage and deprotection

Synthetic peptides must then be cleaved from the solid support and the side-chain protecting groups removed. Cleavage and deprotection were performed using "Synergy Fmoc Cleavage Procedure" as follows (Applied Biosystems, 1993):

53
a. Prepare a fresh cleavage mixture in an ice/salt bath (around -5°C) which containing 90% TFA (trifluoroacetic acid), 5% thioanisole, 2.5% ethanedithiol, and 2.5% H$_2$O;

b. Add enough cold cleavage mixture to peptides (50-100 mg resin, add 700 ul mixture; 100-200 mg resin, add 1200 ul mixture). Stir the resin and mixture at 0°C for 10-15 minutes then allow the reaction to continue at room temperature for a further 2 hours 45 minutes (Total time = 3 hours);

c. Filter off the resin using a Pasteur pipette plugged with glass wool, and wash the resin with 0.5 ml TFA. Retain the filtrate;

d. Add approximately 0.5 ml of peptide-TFA filtrate to 8 ml of MTBE (methylbutyl ether). Peptide should precipitate out. Centrifuge for 5 minutes at room temperature and decant off the supernatant ether. Re-suspend the pellet in the fresh ether and add more peptide-TFA solution until all the peptide is precipitated;

e. Wash the final peptide pellet in fresh ether at least 4 times, and dissolve the final peptide pellet in an aqueous solution and lyophilize.

Crude peptides were then purified using reverse phase HPLC.

3.1.2 Peptide purification using high performance liquid chromatography (HPLC)

HPLC purification was performed on a Waters HPLC system with the Millennium software package, Waters 501 high pressure pumps, Waters 486 UV/VIS detector, and either a Waters U6K injector or Waters 717 Plus Autosampler. An
Eppendorf CH-30 Column heater was used for temperature control. Water was prepared fresh daily using a Vanguard RGW5 water purification system with organic and ionic exchange filters. All HPLC quality solvents including water were passed through 0.45 um filters prior to use to remove particulate matter and were degassed by vacuum aspiration. A two solvents system was used, and both solvents contain 0.1% TFA. Zorbax C8 or C18 reverse phase columns was used in both analytical (4.6 mm x 25 cm) and semi-preparative (9.4 mm x 25 cm) sizes.

3.2 Identification of peptides

To accurately determinate the amino acid composition and mass of the peptides, amino acid analysis and mass spectroscopy were used.

3.2.1 Amino acid analysis

Amino Acid analysis consists of two steps: hydrolysis of the peptide to the constituent amino acids followed by quantitative analysis of the hydrolysis products. Analyses were performed at the Biochemical Instrumentation Center at The Ohio State University to confirm the composition of each peptide.

3.2.2 Mass spectrometry

Mass spectra were taken on a Kompact MALDI III MALDI-TOF (Matrix Assisted Laser Desorption Ionization Time of Flight) mass spectrometry manufactured by Kratos Analytical.
In these analyses, αCCA (alpha-cyano-4-hydroxycinnamic acid) was chosen as the matrix. Glucagon (Sigma, G7774) and Bovin insulin (Sigma, 15500) were used as standards. Samples containing 10 pmole of peptides gave the best results, therefore the samples were prepared in the following manner:

a. Prepare saturated solution of αCCA in 50:50/AcN:H₂O solution (notice that fresh matrix solution always gives better results);

b. 1 ul of 100 uM peptide stock solution and 1 ul of 100 uM standard solution were added to 8 ul matrix solution;

c. 1 ul of sample was then added to the sample strip, and was air dried to form the crystal.

A two-point calibration using αCCA matrix peak (MW=189.2, M+1=190.2) or sodium peak (MW=23.0) and standard glucagon peak (MW=3482.8, M+1=3483.8) or insulin peak (MW=5733.6, M+1=5734.6) was applied to the spectrum to determine the mass for samples.

3.3 Spectroscopic methods

3.3.1 Ultraviolet / visible (UV/VIS) spectroscopy

All UV/VIS measurements were made using a Uvikon 860 spectrophotometer made by Kontron Instruments. UV quality quartz cells with 1.0 cm path length were used. The spectrophotometer was blanked with the same cuvette as the sample to eliminate absorbance differences between cells. Measurements were taken in double beam mode using a slit of 1 nm and air as the reference.
Absorbances of tryptophan (Trp) at 280 nm and tyrosine (Tyr) at 274 nm were used to determine the peptide concentration. At room temperature and pH 7.0, the molar absorption coefficients for Trp and Tyr are 5559 M\(^{-1}\) cm\(^{-1}\) and 1400 M\(^{-1}\) cm\(^{-1}\), respectively (Mihalyi, 1968). Concentrations were then determined by Beer-Lambert law:

\[
C_M = \frac{A}{\varepsilon \cdot d} \quad (3.1)
\]

where \(A\) is the absorbance, \(\varepsilon\) is the molar absorption coefficient (M\(^{-1}\) cm\(^{-1}\)), \(d\) is the path length (cm), and \(C_M\) is the molar concentration (M).

### 3.3.2 Circular dichroism (CD) spectrometry

CD measurements were made on a computer controlled Jasco J-500A spectropolarimeter. The software used to control the spectropolarimeter was provided by Dr. W. C. Johnson and revised by Dr. M. H. Zehfus. The computer interface also connects to the Neslab M-RS-2 bath and Neslab RS-2 remote sensor, so that the temperature can be directly controlled by computer.

Since the instrument’s high-intensity light source can convert oxygen into ozone, the instrument was continuously purged with pure dry nitrogen at 5 liter per minute to prevent degradation of optics from ozone. This also eliminated the absorption of oxygen at low wavelengths. Scans were taken at 1 nm interval with a 2 nm slit width. Sensitivity was set to achieve a voltage near ±5 volts at the input to the interface board, and time constant (TC) was set appropriately based on the formula: \(TC \cdot \text{scan rate}\)
Temperature control was achieved to within ±0.2°C using a circulating water bath system.

High quality quartz CD cells (by Hellma) with several different pathlengths from 0.01 cm to 1 cm were used, depending on the sample concentration. Measurements were taken from 330 nm to the lowest possible wavelength achievable for the sample. The lowest possible wavelength for a sample was determined by observing the photomultiplier (PM) voltage. When the PM voltage exceeded 700 volts, output of the machine was not accurate.

3.3.2.1 Conversion of CD data from volts to ellipticity

The CD spectropolarimeter is designed to measure $\Delta A$, and the computer interface collects $\Delta A$ value in terms of voltage. For peptides or proteins, CD data is usually expressed as "mean residue ellipticity". To convert the voltage recorded ($V$, in units of volts) to the mean residue ellipticity ($[\theta]$, in units of deg·cm$^2$·dmol$^{-1}$), the following formula is used:

$$[\theta] = \frac{V(\text{volts}) \cdot f \cdot 3298}{d \cdot C_m \cdot n_{AA}}$$

(3.2)

where $d =$ path length (cm), $C_m =$ molar concentration (M), $n_{AA} =$ number of amino acids in the peptide, and $f$ is the correction factor from the calibration of CD machine. Notice that the "residue" here refers to "number of amino acids" in the sequence.

The instrument was calibrated routinely using a 0.1 mm cell containing CSA solution (d-10-camphorsulfonic acid, 26.68 mM). CSA is frequently used for CD calibration. It has a positive band at 290.5 nm and a negative one at 192.5 nm. The $\Delta \varepsilon$
of CSA is 2.36 M\(^{-1}\) cm\(^{-1}\) and -4.72 M\(^{-1}\) cm\(^{-1}\) at 290.5 nm and 192.5 nm, respectively. The ratio of -2.00 for the minimum and maximum bands provides a simple two-point calibration (Chen, G. C. & Yang, 1977). Accordingly, the correction factor \( f \) of the CD machine using CSA calibration at 290.5 nm maybe:

\[
f = \frac{\Delta A_{\text{CSA at 290.5 nm}}}{V_{\text{CSA at 290.5 nm (volts)}} \times \text{sensitivity}(m^\circ)}
\]  

(3.3),

where \(\Delta A_{\text{CSA at 290.5 nm}}\) is the value obtained for the standard (CSA) at given concentration which can be calculated using equation (3.1) as \(\Delta A = 2.36*d*C_M\), and \(V_{\text{CSA at 290.5 nm}}\) is the net voltage recorded by the CD machine at 290.5 nm.

The output data file contains information for the correction factor, average voltages recorded for each wavelength, and standard deviations corresponding to each wavelength. Using the formulas described above, these data were then converted to mean residue ellipticity using SigmaPlot scientific graphing program.

3.3.2.2 Determination of CD cell path lengths using FT-IR spectroscopy

It is important to determine the pathlength of CD cells accurately. Pathlengths (\(d < 0.1\) cm) were determined experimentally using interference fringes from an FT-IR spectrum.

FT-IR spectra of empty cells were taken with no reference cell and a 1.0 cm cell as a blank to compensate absorbance due to the quartz. The IR spectra contained interference fringes between 4400–2900 wave numbers, which were used to calculate the pathlength from the following formula:
where \( d \) = path length in cm, \( \eta \) = refractive index (1.0 for quartz), and \( n \) is the number of interference fringes between two wave numbers \( \nu_1 \) and \( \nu_2 \).

3.3.3 Nuclear magnetic resonance (NMR)

NMR measurements were carried out on either a Bruker Aspect 250 and an Aspect 270 or a DMX-600 spectrometer. Peptides were prepared in \( \text{D}_2\text{O} \), and TMSP (sodium 3-(trimethylsilyl)propionate-2,2,3,3,4,4-d\(_4\)) was chosen as the reference since the signal for the hydrogen atoms in TMSP is assigned to be 0 ppm. Samples were measured in 5 mm NMR tubes. All NMR spectra were processed using either the software package Felix 2.30 or XWIN NMR on a Silicon Graphics Indigo workstation.

3.4 Data analysis

In order to apply the helix-coil transition models, ellipticity at 222 nm for a completely helical state, \([\theta]_H\), and completely random coil state, \([\theta]_C\), are required. It has been shown that \([\theta]_H\) is chain length dependent and can be expressed as:

\[
[\theta]_H = [\theta]_H^n (1 - \frac{k}{n}) \quad \text{or} \quad n[\theta]_H = n[\theta]_H^n - k[\theta]_H^n
\]  

(3.5)

where \( k \) is the chain-length dependent factor at 222 nm, \( n \) is the number of amides in the complete helix, and \([\theta]_H^n\) is the ellipticity for an infinitely long, completely helical peptide (Chen, Y.-H. et al., 1974; Gans et al., 1991). According to equation (3.5), a plot of
n[θ]_H versus n will give a straight line with slope [θ]_H^+ and intercept -k[θ]_H^+. Thus, [θ]_H^+ and k can be obtained experimentally. In addition, [θ]_C does not depend on chain length and can be obtained by denaturing the peptide.

Several different values for k, [θ]_H^+ and [θ]_C at 222 nm have been used in literature (see Chapter 2, section 2.1.3.4). In this work, we will follow the values given by Shalongo and Stellwagen unless otherwise indicated, where k=4.32, [θ]_H^+=−40000, and [θ]_C is assumed to be zero (Shalongo & Stellwagen, 1997). The differences due to using different values are generally not significant.

Notice that the ellipticities for peptides are usually expressed as the "mean residue ellipticity" and calculated by taking the total ellipticity and dividing by the total number of amino acids in the system (equation (3.2)). In theory, however, it is the peptide bond or amide chromophore that contributes to the CD signal at 222 nm, and our peptides contain N- and C-terminal blocking groups that make the number of peptide bonds not equal to the number of amino acids. A block peptide with x amino acids has x+1 amides when its both termini are blocked, and x−1 amides if no blocking groups are present. Thus when applying the helix-coil theory, we will adjust the [θ]_obs values if necessary and the term "mean amide ellipticity" will be used instead. For example, a correction factor x/(x−1) for a peptide without blocking group or x/(x+1) when both termini are blocked will be applied.
3.4.1 Analysis using the Two-State model

This analysis assumes that the thermal dependence of peptides can be described by a simple two-state, Coil ↔ Helix, transition, and the equilibrium constant \( K \) in terms of ellipticity can be expressed as

\[
K = \frac{[\theta]_{\text{obs}} - [\theta]_c}{[\theta]_H - [\theta]_{\text{obs}}} \tag{3.6}
\]

At any particular temperature, the free energy change \( \Delta G = -RT \ln K \) or

\[
\Delta G = -RT \ln \left( \frac{[\theta]_{\text{obs}} - [\theta]_c}{[\theta]_H - [\theta]_{\text{obs}}} \right) \tag{3.7}
\]

where \( T \) is the absolute temperature in Kelvin (K), and \( R \) is the gas constant (1.987 cal·K\(^{-1}\)·mol\(^{-1}\) or 8.314 J·K\(^{-1}\)·mol\(^{-1}\)). Thus by comparing a peptide's CD ([\(\theta\]_obs) with the CD of fully helical ([\(\theta\]_H) and fully coil ([\(\theta\]_C) states, we can obtain \( \Delta G \) for a peptide at any given \( T \). All modified peptides can then compared to the unmodified control peptide to obtain \( \Delta \Delta G \) values for the free energy change due to the modification:

\[
\Delta \Delta G = \Delta G_{\text{modified}} - \Delta G_{\text{control}} \tag{3.8}
\]

When \( \Delta G \) is followed as a function of temperature, it is possible to derive values for \( \Delta H \) and \( \Delta S \) according to the van't Hoff relationship:

\[
\ln K = -\frac{\Delta H}{R} \left( \frac{1}{T} - \frac{1}{T_m} \right) \tag{3.9}
\]

where \( T_m \) and \( \Delta H \) are derived by finding the best fit of these parameters to the experimental melting curves. And \( \Delta S \) at particular temperature \( T \) is
\[ \Delta S = \left( \frac{\Delta H - \Delta G}{T} \right) \]  

(3.10)

3.4.2 Analysis based on Lifson-Roig theory

The Lifson-Roig theory applies statistical mechanics principles to try to more accurately reflect the ensemble of peptide structures (Lifson & Roig, 1961). Models based on Lifson-Roig theory, using the matrix method (see Chapter 2, section 2.3.3) or the enumeration method (see Chapter 2, section 2.5.2), at several levels of sophistication were applied to our peptide systems. The mathematics of the matrix method was performed using Maple V (Waterloo Maple Software) on a personal computer, and the application of the enumeration method was calculated using Fortran programs.

As the Lifson-Roig theory has evolved over the past several years, different definitions have been used for the number of residues in a peptide and the helix content of the peptide (Doig et al., 1994; Rohl et al., 1996) (see Chapter 2 for detail). In this work, we will use Doig et al.'s definition for residue number N consistently, where each amino acid and blocking group in the sequence is considered a residue (see Chapter 2, section 2.4.1) (Doig et al., 1994), and the helix content will be assumed to be the same as the fraction hydrogen bonds, i.e. \( f_{\text{H,cut}} = f_{\text{H-bond}} = \frac{n_h}{N_h} \) (see Chapter 2, section 2.4.4) (Rohl et al., 1996), unless otherwise mentioned.
3.4.2.1 Simple uniform helix analysis

In this analysis each peptide is treated as a homopolymer, and the capping effect is not included. Thus every residue has the same intrinsic parameters and the original 3x3 statistical weight matrix is applied to do the analysis:

\[
\begin{pmatrix}
\tilde{h}h & \tilde{h}c & \tilde{c}(h \cup c) \\
\tilde{h}h & <w> & <v> & 0 \\
\tilde{h}c & 0 & 0 & 1 \\
\tilde{c}(h \cup c) & <v> & <v> & 1
\end{pmatrix}
\]

(3.11)

and the partition function is:

\[
Z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} J^N \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}
\]

(3.12)

In this model it is assumed that all amino acids have the same value for the helix nucleation parameter \(v\), because it is a property of the peptide backbone, and there is evidence that \(v\) is not influenced greatly by the side chain (Scholtz et al., 1991b; Chakrabartty et al., 1994). Thus, the \(<v> = 0.048\) obtained by Rohl et al. using total NH exchange kinetics method is used (Rohl et al., 1992). After the partition function is generated using repetitive multiplication of the \(J\) matrix, the helix content \(f_{\text{Helix}}\) of a peptide becomes a function of \(<w>\) (equation (3.13)). An experimental helix content \(f_{\text{Helix}}\) is calculated from the experimental CD value ([\(\theta\]o) using equation (3.14), and by equating equations (3.13) and (3.14) the average value for the helix elongation \(<w>\) can be determined. A Maple program based on this analysis is included in Appendix A.
The $\langle w \rangle$ parameter can be related to $\Delta G$ term as (Chakrabartty et al., 1994):

$$\Delta G = -RT \ln \langle w \rangle$$

(3.15).

The difference in $\Delta G$ due to the change of $w$ weighting between modified and control peptide is therefore:

$$\Delta \Delta G = -RT \ln \langle w \rangle_{\text{modified}} - (-RT \ln \langle w \rangle_{\text{control}}) = -RT \ln \frac{\langle w \rangle_{\text{modified}}}{\langle w \rangle_{\text{control}}}$$

(3.16).

Notice that this is the energy change for a single residue. It should be reminded that an N-residue peptide contains $N-4$ residues that can have $w$ values (see Chapter 2, section 2.4.3); therefore, the total energy change for the entire peptide is

$$\Delta \Delta G = (N-4)(\Delta G_{\text{modified}} - \Delta G_{\text{control}}) = -(N-4)RT \ln \frac{\langle w \rangle_{\text{modified}}}{\langle w \rangle_{\text{control}}}$$

(3.17).

This simple model is easy to calculate; however, the effect of the modification is not localized to a single position but is averaged over the entire sequence.

### 3.4.2.2 Modified uniform helix analysis

This analysis is similar to the uniform helix analysis, but a single heteroresidue is inserted into the homopolymer host at the site of the modification (Qian, 1993). Instead of calculating a single $\langle w \rangle_{\text{modified}}$ that is applied to all residues in the modified peptide, all non-modified residues in the sequence are given the same $\langle w \rangle$ value as the residue in
control peptide \(<w>_{\text{control}}\), and the \(<w>\) of the modified residue is replaced with a \(w'\). The partition function for the system, where the \(m^{\text{th}}\) residue is modified, can then be expressed as:

\[
Z = \begin{pmatrix} w' & <v> & 0 \\ 0 & 0 & 1 \end{pmatrix} J^{m-1} \begin{pmatrix} 0 \\ <v> \\ <v> \end{pmatrix} J^{N-m} \begin{pmatrix} 1 \\ 1 \end{pmatrix}
\]

(3.18).

The \(<v>\) is kept the same as in simple uniform helix analysis (\(<v>=0.048\)). Using the same algorithm as described in simple uniform helix analysis, the \(w'\) value for the modified residue can be obtained (see Appendix B for the calculation using a Maple program). The resulting \(\Delta \Delta G\) term becomes:

\[
\Delta \Delta G = \Delta G_{\text{modified}} - \Delta G_{\text{control}} = -RT \ln \frac{w'}{<w>_{\text{control}}}
\]

(3.19).

Notice that because the perturbation is localized to a single residue, the N–4 term has been eliminated.

This model is more realistic because it localizes the effect of the modification to a single residue instead of averaging the effect over the entire helix. However, the assumption of a homopolymer host is clearly an oversimplification for our peptide systems.

### 3.4.2.3 Full helix analysis

To more realistically model the peptide's structure, a complete Lifson-Roig treatment may be used. This analysis includes \(w, v,n\) (N-capping), and \(c\) (C-capping) terms that are unique for each amino acid. In addition, the side chain interactions in
terms of p, q and r parameters have also been included in this model (see Chapter 2, section 2.4.3). Thus a 6x6 statistical weight matrix $L$ is used to this analysis:

$$
L_t = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \gamma_t & 0 & 0 & 0 \\
0 & \gamma_t & 0 & 0 & 0 & 1 \\
0 & 0 & \gamma_t & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
$$

(3.20).

And the partition function is

$$
Z = \left(0\ 0\ 0\ 0\ 0\ 1\right) \prod_{i=1}^{N} L_i
$$

(3.21).

This analysis is Rohl et al.'s modified Lifson-Roig model (see Chapter 2, section 2.4 for detail) (Rohl et al., 1996), thus their intrinsic parameter library (Table 2.3) has been applied to our system. When deriving this parameter library, the values $k=3.0$ (for peptides with amide blocking group) or $k=4.0$ (for peptides without C-terminal blocked), $[\theta]_h=45000$, and $[\theta]_c=+640$ were used. As a result, the CD for our peptides at 222 nm is predicted using the following equation:

$$
[\theta]_{\text{cal}} = \bar{f}_{\text{H,cal}} \times [-42500 \times (1 - \frac{3.0}{n}) - 640] + 640
$$

(3.22),

where $n$ is the number of amide in the sequence, and $\bar{f}_{\text{H,cal}}$ is derived from partition function using
Note that the values used in equation (3.22) differ from those given earlier, since this model is used with the Rohl et al.'s intrinsic parameter library, the values given by them are used for consistency.

Rather than having only one variable (\(<w>\) or \(w'\)) in uniform helix models, there are at least 4 variables (\(w_{\text{modified}}\), \(v_{\text{modified}}\), \(\theta_{\text{modified}}\) and \(c_{\text{modified}}\)) for the modified residue in full helix analysis. The Maple program cannot solve these parameters easily, thus the algorithm has been rewritten in a Fortran program using the enumeration method (Appendix C). Parameters for the modified residue are obtained using the optimization method (see section 3.5).

Notice that the \(v\) values are still assumed to be the same for all residues in this analysis. However, because of the inclusion of capping parameters in this analysis, the \(v\) value has been revised from 0.048 to 0.036 (Rohl et al., 1992, 1996). Therefore \(v=0.036\) is applied to all amino acids except the modified residues.

After the best fit \(w_{\text{modified}}\) value is obtained, the \(\Delta \Delta G\) value due to the change of \(w\) weight can be calculated using:

\[
\Delta \Delta G = \Delta G_{\text{modified}} - \Delta G_{\text{regular}} = -RT \ln \frac{w_{\text{modified}}}{w_{\text{regular}}}
\]  

(3.24)
3.4.3 Analysis using Dichroic model

The Dichroic model uses exactly the same statistical assumptions that are used in the Lifson-Roig theory, but instead of using matrices to keep from identifying individual state of the ensemble, this model actually enumerates each individual state. This enumeration is necessary when calculating chain length dependent properties like CD (see Chapter 2, section 2.5).

The fractional concentration of a particular conformation \( k \) is calculated using

\[
f_k = \frac{Z_k}{\sum Z_k}
\]

where \( Z_k \) is the product of the statistical weight of each residue, that is, the probability of the occurrence of the conformation \( k \). And the corresponding mean amide bond ellipticity for this specific conformation is calculated using

\[
[\theta]_k = \frac{b_c[\theta]_c + b_{b_1}[\theta]_{b_1} + b_{b_2}[\theta]_{b_2}}{b} = b_{b_1} \cdot (-11200) + b_{b_2} \cdot (-40000)
\]

\[
N - 1
\]

\( N \): number of residues (number of amino acids plus blocking groups);

\( b_{b_1} \): number of peptide bonds in the helix state with one backbone hydrogen bond;

\( b_{b_2} \): numbers of peptide bonds in the helix state with two backbone hydrogen bonds.

Notice that we use \([\theta]_c=0\), \([\theta]_{b_1}=-11200\), and \([\theta]_{b_2}=-40000\) in this analysis.

CD for an \( N \)-residue peptide at 222 nm is calculated by summing the mean amide bond ellipticity from all possible \( 2^{N-2} \) conformations as:

\[
[\theta]_{cd} = \sum_k f_k \cdot [\theta]_k
\]

(3.27).
3.4.3.1 *Determination of Intrinsic parameters*

There is no intrinsic parameter library available for Dichroic model yet. Before we can use Dichroic model to analyze our data, it is necessary to derive \( w, v, n \) and \( c \) parameters for Ala, Arg, Glu, Trp, Gly, and the acetyl and amide blocking groups. To do this, the data listed on Tables 3.1 and Table 3.2 have been used to find the best-fit parameters for these residues.

In this fitting, the \( v \) value is kept the same for all amino acids at \( v=0.036 \) as determined by Rohl et al. for all residues (Rohl et al., 1992, 1996). The \( [\theta]_{\text{cal}} \) value is determined for each peptide in Table 3.1 and Table 3.2 using equation (3.25)-(3.27). The \( w, n \) and \( c \) parameters are then varied using a simplex optimization method (section 3.5) until the best-fit values are obtained.

Table 3.3 summarizes the best-fit values we obtained, and figure 3.4 shows the goodness of the fit of experimental data to the Dichroic model predictions based on these best-fit values. After the intrinsic parameters are obtained, the statistical weights for the modified residue can be found using the same algorithm. A Fortran program based on Dichroic model and using enumeration method to calculate the ellipticity for a given sequence can be found in Appendix D.

Once parameters are derived for regular amino acids and the modified residue, a \( \Delta\Delta G \) value for the structural change can be obtained using equation (3.24).
3.5 Optimization method

Finding the best-fit intrinsic parameters that minimize the differences between experimental data and theoretical data is the multidimensional minimization process. We need to find the minimum of a function of more than one independent variable using limited data set, which becomes a nonlinear optimization problem. Here we use an optimization method called "simplex method" to solve this problem. The simplex method is based on the comparison of the objective function (see section 3.5.2) values at the \((n+1)\) vertices of a general simplex and moving this simplex towards the optimum point (Jacoby et al., 1972). (A set of \((n+1)\) equidistant points in \(n\)-dimensional space forms a regular simplex, and a general simplex is a simplex without the property that the vertices have equal distance.)

To perform the simplex optimization, we have chosen to use the "AMOEBA" subroutine written by Press et al. (Press et al., 1989). To use AMOEBA, an initial simplex, the objective function, and the convergence criterion must be determined first.

3.5.1 Generation of the initial simplex

For an \(N\)-dimensional problem (\(N\) unknowns), an initial matrix with \(N+1\) rows and \(N\) columns (i.e. an initial simplex) is necessary. Each row of this matrix thus represents a set of \(N\) parameters, and we have \(N+1\) sets of parameters. We use the following steps to generate the initial simplex and assure that its vertices span the full \(N\)-dimensional space:
a. Give an initial estimate vector $x'=(x_1, x_2, x_3, \ldots, x_n)$:

b. Generate the $n+1$ vertices of a general simplex as:

$$
\begin{align*}
\mathbf{x}^1 &= \begin{pmatrix} x_1 & x_2 & \cdots & \cdots & x_n \end{pmatrix} \\
\mathbf{x}^2 &= \begin{pmatrix} x_1 + p_n & x_2 + q_n & x_3 + q_n & \cdots & x_n + q_n \end{pmatrix} \\
\mathbf{x}^3 &= \begin{pmatrix} x_1 + q_n & x_2 + p_n & x_3 + q_n & \cdots & x_n + q_n \end{pmatrix} \\
\mathbf{x}^4 &= \begin{pmatrix} x_1 + q_n & x_2 + q_n & x_3 + p_n & \cdots & x_n + p_n \end{pmatrix} \\
\mathbf{x}^{n-1} &= \begin{pmatrix} x_1 + q_n & x_2 + q_n & x_3 + q_n & \cdots & x_n + p_n \end{pmatrix}
\end{align*}

(3.28).

Where $p_n$ and $q_n$ are calculated using following equations:

$$
p_n = \frac{\sqrt{n+1-n}}{n\sqrt{2}} S
$$

(3.29),

$$
q_n = \frac{\sqrt{n+1-n}}{n\sqrt{2}} S
$$

(3.30),

$S$ is a scaling factor, and we use $S=0.5$ in our calculation.

### 3.5.2 Objective function and convergence criterion

The objective function for our calculations is

$$
Y = \sum_{i=1}^{N_e} ([\theta]_{i,\text{cal}} - [\theta]_{i,\text{obs}})^2
$$

(3.31),

where $N_e$ is the number of data sets. This objective function is the sum of the residuals, the squares of differences between experimental data $[\theta]_{i,\text{obs}}$ and calculated results $[\theta]_{i,\text{cal}}$.

The best fit parameters are obtained when this objective function reaches a minimum.
The AMOEBA routine varies the parameters by looping over the vectors in the simplex, and calculates the highest and the lowest values of the objective function. In our calculation, when difference between the highest value and lowest value satisfies the criteria:

\[
\frac{Y_{\text{highest}} - Y_{\text{lowest}}}{Y_{\text{highest}}} < 10^{-4}
\]

(3.32), the convergence is assumed and the process is terminated. That is, the range of residuals in the data set must be less than 0.01%.
<table>
<thead>
<tr>
<th>No</th>
<th>Sequence</th>
<th>Ellipticity</th>
<th>N</th>
<th>n_{aa}</th>
<th>n_{amide}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AAKAAAAKAAAAAKAAGYZ</td>
<td>12900</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>QAKAAAAKAAAAAKAAGYZ</td>
<td>10000</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>KAKAAAAKAAAAAKAAGYZ</td>
<td>12100</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>XAKAAAAKAAAAAKAAGYZ</td>
<td>19700</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>XYGGKAAAAAAKAAAAKAAAAKZ</td>
<td>24000</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>XYGGKAAAAAAKAAAAKAAAGAKZ</td>
<td>16100</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>XYGGKAAAAAAKAAAAKAAAGAKZ</td>
<td>8900</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>XYGGKAAAAAAKAAAAKAAAAKZ</td>
<td>12900</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>XYGGKAAAAAAKAAAAKAAAAKZ</td>
<td>25400</td>
<td>20</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>XYGGGKAAAAAAKAAAAKAAAAKZ</td>
<td>23500</td>
<td>22</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>XYGGKAAAAAAKAAAQAAKAAAAKZ</td>
<td>17600</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>XYGGKAQAQAAKAAAQAAKQAKZ</td>
<td>10800</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>XKAQAQAAKAAAQAAAKGQYZ</td>
<td>13700</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>KAQAQAAQAAKAAAQAKQGQYZ</td>
<td>7600</td>
<td>20</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>XYGGQAAAAQAAAAQAARAAAAQZ</td>
<td>18000</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>XYGGQAAAAQAQARAAAAQARAAQZ</td>
<td>13100</td>
<td>21</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>17</td>
<td>XYGAQAAAAKAAAAQAAAAKAZ</td>
<td>18200</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>XYGAQAAAAQAAAAQAAAAKAA</td>
<td>18500</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>XYGAQAAAAQAAAAQAAAAQAK</td>
<td>18300</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>20</td>
<td>XYGAQAAAAQAAAAQAAAAQAKK</td>
<td>16300</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>21</td>
<td>XKAQAAAAQAAAAQAAAAQAKZ</td>
<td>17100→18800</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>22</td>
<td>XAAQAAAAQAAAAQAAAAQAYZ</td>
<td>16000→19300</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

N: number of residues, n_{aa}: number of amino acids, and n_{amide}: number of amides in the sequence.

Table 3.1: Peptides used to determine intrinsic parameters for Ala (A), Gly (G), Lys (K), Gln (Q), Tyr (Y), acetyl group (X) and amide blocking group (Z).
<table>
<thead>
<tr>
<th>Sequence for Trp</th>
<th>Ellipticity</th>
<th>N</th>
<th>$n_{aa}$</th>
<th>$n_{amide}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 WAKAAKAAAAKAAAGYZ</td>
<td>18800→17600</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2 XYGAAKAAAAKAAAAKAW</td>
<td>14000→13100</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3 XKAATAKAWAAKAAAAKZ</td>
<td>17000→15600</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence for Glu</th>
<th>Ellipticity</th>
<th>N</th>
<th>$n_{aa}$</th>
<th>$n_{amide}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 EAKAAKAAAAKAAAGYZ</td>
<td>15300</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2 XYGAAKAAAAKAAAAKAE</td>
<td>16200</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3 XAAQAEAAAQAQAAYZ</td>
<td>11700→14200</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>4 XAAQAEAAQAAQAAYZ</td>
<td>12100→14600</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>5 XAAQAAAAQAQAAYZ</td>
<td>9300→11400</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>6 XAAQAAAAQAQAAYZ</td>
<td>9000→11000</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence for Arg</th>
<th>Ellipticity</th>
<th>N</th>
<th>$n_{aa}$</th>
<th>$n_{amide}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 RAKAAKAAAAKAAAGYZ</td>
<td>12900</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2 XYGAAKAAAAKAAAAKAR</td>
<td>18800</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3 XAAQARAQAQAQAAYZ</td>
<td>14400→17200</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>4 XAAQAARAQAQAQAAYZ</td>
<td>13200→15900</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>5 XAAQAAAAQAQAAYZ</td>
<td>16600→19800</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>6 XAAQAAAAQAQAAYZ</td>
<td>15600→18600</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence for Pro</th>
<th>Ellipticity</th>
<th>N</th>
<th>$n_{aa}$</th>
<th>$n_{amide}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 PAKAAKAAAAKAAAGYZ</td>
<td>13800</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2 XYGAAKAAAAKAAAAKAP</td>
<td>13000</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3 XWGGRAAARAPAARAAAAAARZ</td>
<td>11900</td>
<td>26</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

$N$: number of residues, $n_{aa}$: number of amino acids, and $n_{amide}$: number of amides in the sequence.

Table 3.2: Peptides used to determine intrinsic parameters for Trp (W), Glu (E), Arg (R), and Pro (P).
<table>
<thead>
<tr>
<th>Residue code</th>
<th>w</th>
<th>v</th>
<th>n</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ALA</td>
<td>A</td>
<td>1.73</td>
<td>0.036</td>
<td>1.00</td>
</tr>
<tr>
<td>2 GLU</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 CYS</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 ASP-</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 GLU-</td>
<td>E</td>
<td>0.63</td>
<td>0.036</td>
<td>2.95</td>
</tr>
<tr>
<td>6 PHE</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 GLY</td>
<td>G</td>
<td>0.06</td>
<td>0.036</td>
<td>2.87</td>
</tr>
<tr>
<td>8 HIS+</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 ILE</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 HIS</td>
<td>J</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 LYS+</td>
<td>K</td>
<td>1.02</td>
<td>0.036</td>
<td>0.86</td>
</tr>
<tr>
<td>12 LEU</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 MET</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 ASN</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 ASP</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 PRO</td>
<td>P</td>
<td>0.036</td>
<td>1.69</td>
<td>0.00</td>
</tr>
<tr>
<td>17 GLN</td>
<td>Q</td>
<td>0.57</td>
<td>0.036</td>
<td>0.08</td>
</tr>
<tr>
<td>18 ARG</td>
<td>R</td>
<td>1.59</td>
<td>0.036</td>
<td>1.19</td>
</tr>
<tr>
<td>19 SER</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 THR</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 LYS</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 VAL</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 TRP</td>
<td>W</td>
<td>0.15</td>
<td>0.036</td>
<td>5.67</td>
</tr>
<tr>
<td>24 ACE</td>
<td>X</td>
<td>0.00</td>
<td>0.036</td>
<td>6.95</td>
</tr>
<tr>
<td>25 TYR</td>
<td>Y</td>
<td>0.25</td>
<td>0.036</td>
<td>3.26</td>
</tr>
<tr>
<td>26 AMD</td>
<td>Z</td>
<td>0.00</td>
<td>0.036</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3.3: The intrinsic parameter library generated using modified Dichroic model for some amino acids and blocking groups.

76
Figure 3.1: Steps in solid phase synthesis: deprotection, activation, and coupling.
Figure 3.2: Deprotection of the Fmoc group using piperidine.
Figure 3.3: HBTU activation for Fmoc solid-phase peptide synthesis.
Figure 3.4: Goodness of the fit of experimental data to the Dichroic model predictions (correlation coefficient $R=0.93$). The observed ellipticity of each peptide in Table 3.1 and 3.2 is plotted against its calculated ellipticity using the best-fit parameters in Table 3.3.
CHAPTER 4

THE N-METHYL MODIFICATION

4.1 Introduction

An important feature of the $\alpha$-helix is the hydrogen bond between amides and carbonyls along the helix backbone. If the hydrogen in an amide bond is replaced with a methyl group, the ability of the NH group to take part in a hydrogen bond is lost, and a bulky methyl group is inserted into the helix backbone. Determining how much this substitution destabilizes the helix is important, both because it will help us determine the effects of this substitution in biologically relevant peptides (Dive et al., 1992; Ron et al., 1992) and because it can give us information on basic helix backbone interactions. Here we study the effects of the N-methyl modification on the helix by placing this group into two model peptides, one stabilized with Glu$^-$—Arg$^+$ (E$\cdots$R) salt bridges, the other with no salt bridges.

In this study, the melting curve of a peptide's helix-coil transition is followed using circular dichroism (CD). By comparing each substitution with an unmodified control, we try to derive the free energy change ($\Delta\Delta G$) for the destabilization of the substitution itself.
In analyzing the data obtained here, we use the Two-State model, various
modifications of the Lifson-Roig model (Lifson & Roig, 1961; Qian, 1993; Doig et al.,
1994; Rohl et al., 1996), and the Dichroic model (Shalongo & Stellwagen, 1997).

Besides removing a backbone hydrogen bond, the N-methyl modification can
introduce cis-trans isomerization as an additional variable. The ability of the N-methyl-
analine derivative to form a cis peptide bond was first demonstrated by Goodman et al.
in 1974 (Goodman et al., 1974). In this work, 1D NMR was used to monitor cis-trans
isomerization, and some adjustments for this factor will be discussed.

4.2 Material and method

4.2.1 Peptide design

The peptide Ac-WEAAAREAAAAREAAARA-Amide is reported to be
nearly completely helical in 10 mM NaCl at 0°C and pH 7.0 due to its E-R salt bridges
(Merutka et al., 1991). In order to exclude the aromatic side-chain contribution to far-
UV CD signal used to monitor the helix-coil transition, two glycine residues were
inserted in model peptides to separate tryptophan from the helix (Chakrabartty et al.,
1993b), thus our first model peptide (24p1con) has the sequence Ac-
WGGEAAAREAAAAREAAARA-Amide. Our second model peptide (24p2con)
has the sequence Ac-WGGRAAAAAARAAAAARAAAAR-Amide. The second
peptide is similar to the first, but the E-R salt bridges have been eliminated to try to
simplify the data analysis. In labeling these peptides, the mxx suffix indicates the
position of the N-methyl amino acid; thus 24p1m06 or 24p2m06 refers to the peptide.
synthesized with an N-methyl amino acid placed at the sixth amino acid, and 24p1con or 24p2con refers to non-methylated control peptide. Table 4.1 summarizes sequences and notations for these peptides.

4.2.2 Peptide synthesis and purification

Peptides were synthesized on an Applied Biosystems 432A Synergy peptide synthesizer using Fmoc chemistry. Amidation of the C-terminal amino acid was accomplished by starting synthesis with an amide linked resin. N-terminus was acetylated using the acetic anhydride, triethylamine (TEA) and dimethylformamide (DMF) mixture on resin bound, fully protected peptide (see chapter 3, section 3.1.1.1 for details). The N-methyl amino acids were incorporated in the standard Fmoc synthesis using an Fmoc protected N-methylated amino acid (Nova Biochem, Fmoc-MeAla-OH, A12213). Coupling of the next amino acid in the sequence was expected to be slow due to steric hindrance, so this reaction time was routinely doubled to improve yield.

Peptides were cleaved from the resin at 0°C using 90% trifluoroacetic acid (TFA), 5% thioanisole, 2.5% ethanedithiol and 2.5% H₂O, then extracted five times with methylbutyl ether (MTBE) (see chapter 3, section 3.1.1.2 for details).

Peptides were purified by HPLC using Zorbax C8 or C18 reverse phase columns with isocratic methanol/water mixtures containing 0.1% TFA. The conditions used to purify each peptide is summarized in Table 4.2.

The amino acid composition was confirmed by amino acid analysis. Analyses were performed by the Biochemical Instrumentation Center at The Ohio State
University. The mass of each peptide was determined by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry to be within two mass units of the calculated values.

4.2.3 Spectral measurements

Circular dichroism (CD) measurements were made on a computer controlled Jasco J-500A spectropolarimeter using either a 1 mm or 1 cm quartz cell. Temperature control was achieved to within ±0.2°C using a circulating water bath system.

The 24p1 peptides contain ionized Glu\(^{-}\) and Arg\(^{-}\) side chains at pH 7.0, and the attractive charge interaction between these groups is thought to help stabilize the \(\alpha\) helix in these peptides (Merutka et al., 1991). On the other hand, the 24p2 peptides contain only the Arg\(^{-}\) side chain at pH 7.0, and the repulsive charge-charge interaction between these groups is expected to destabilize this helix. These ion pair interaction may be modulated by increasing or decreasing NaCl concentration in the solution (Huyghues-Despointes et al., 1993). Because of this, the 24p1 peptides were studied under low salt conditions (10 mM NaCl) to maximize their favorable charge-charge interactions, while the 24p2 peptides were studied under high salt conditions (1 M NaCl) to minimize its unfavorable charge interactions. The 24p2 peptides were also studied under low salt conditions (10 mM NaCl) to determine how strong this interaction is.

The concentration of the peptides was determined by tryptophan absorbance at 280 nm using \(\varepsilon_{280nm}=5559 \text{ M}^{-1}\cdot\text{cm}^{-1}\) at pH 7.0 (Mihalyi, 1968). Before each CD
measurement, the UV absorbance of the sample was measured on a Uvikon 860 spectrometer with a CD cell holder.

Nuclear magnetic resonance (NMR) measurements were carried out on either a Bruker Aspect 250 or a DMX-600 NMR spectrometer. Peptides were prepared in D₂O, and the residual water peak was suppressed by presaturation during the relaxation delay.

4.3 Results and data analysis

4.3.1 Effects of N-methyl modification

Melting curves measured by CD at 222 nm for the model peptide and N-methyl modified peptides are shown in Figures 4.1, 4.2 and 4.3. The mean residue ellipticity at 222 nm ([θ]_{obs}, in units of deg·cm²·dmol⁻¹) was calculated by taking total ellipticity and dividing by the total number of amino acids (see Chapter 3, section 3.3.2.1, equation (3.2)). Since two glycine residues were inserted between tryptophan residue and the rest of helical sequence, the aromatic contribution to the CD at 222 nm is not significant (Chakrabarty et al., 1993b). Data obtained under both heating and cooling regimes are identical, indicating the reversibility of the thermal transition from 0°C to 80°C.

Table 4.3 summarizes the effect of N-methyl group in terms of the reduction of helix content at 0°C. The conversion of mean residue ellipticity to helix content at 0°C follows equation (3.14), where we used k=4.32, [θ]_H=−40000, and [θ]C=0 to do the calculation (Shalongo & Stellwagen, 1997). Thus the formula used here is:

\[ \text{helix content} = \frac{[\text{mean residue ellipticity}]}{4.32} \]
Notice that we use the mean amide bond ellipticity $[\theta]_{\text{obs}}^{\text{am}}$ in our calculation, because it is the amide bond that contributes to CD at 222 nm. The mean amide bond ellipticity were found from mean residue ellipticity $[\theta]_{\text{obs}}$ using:

$$[\theta]_{\text{obs}}^{\text{am}} = [\theta]_{\text{obs}} \cdot \frac{n_{\text{AA}}}{n_{\text{amide}}}$$  \hspace{1cm} (4.2),

where $n_{\text{AA}}$ is the total number of amino acids, and $n_{\text{amide}}$ is the total number of amides in the sequence.

As seen in Table 4.3, the N-methyl group is a strong helix disrupter and the effect on both peptide models is similar. Even when the N-methyl alanine is placed at position 6, where amide hydrogen bonding is not expected in model peptide, the helix content at 0°C is reduced by 31% for 24p1m06, 20% for 24p2m06 in 10mM NaCl and 27% for 24p2m06 in 1M NaCl. The effect of the methyl group is also position sensitive, with the greatest reduction in helicity occurring when this group is placed in the middle of the helix.

### 4.3.2 Dependence of helix content on NaCl concentration

The change in ellipticity of the 24p2con peptide under different NaCl concentrations at pH 7.0, as shown in Figure 4.4, suggests that the charge interactions of Arg$^+$ in 24p2 peptides can be screened effectively by 1M NaCl. This agree with the result obtained by Huyghues-Despointes et al. (Huyghues-Despointes et al., 1993).
Table 4.4 shows the change in helix content of 24p2 peptides caused by screening charge interactions at pH 7.0 with NaCl. The helix content increases when NaCl concentration increases from 10 mM to 1 M, but does not appear to increase with the addition of more NaCl (Figure 4.4). When applying the helix-coil theory to 24p2 peptides, we used the data obtained in 1M NaCl to simplify the analysis.

4.3.3 Thermodynamic properties of N-methyl backbone modification

To study the effect of the N-methyl modification in terms of energy change, we analyzed the data using helix-coil theory. When applied to helix-coil models, ellipticities for completely helical state and completely random coil state are necessary. Here we used \( k = 4.32 \) and \( [\theta]_h^0 = -40000 \) to calculate the ellipticity for the completely helical state \([\theta]_h\) unless otherwise indicated. The temperature dependence of \([\theta]_h\) values for both peptide models were determined experimentally by studying ellipticity of both control peptides (24p1con and 24p2con) in 70% TFE as a function of temperature. According to our results, the temperature dependence of \([\theta]_h\) for 24p1 peptide in 10mM NaCl can be expressed as:

\[
[\theta]_h = -40000 \left(1 - \frac{4.32}{25}\right) + 222 \cdot \text{Temp(°C)} \tag{4.3},
\]

and for 24p2 peptides in 1M NaCl the equation is:

\[
[\theta]_h = -40000 \left(1 - \frac{4.32}{25}\right) + 196 \cdot \text{Temp(°C)} \tag{4.4}.
\]
These equations are similar but not identical to those given by others (Chen, Y.-H. et al., 1974; Scholtz et al., 1991b), but should be more appropriate for our system. The ellipticity for the completely random coil state \([\theta]_c\) was assumed to be zero at 0°C, and the temperature dependence based on Scholtz et al.'s equation was used (Scholtz et al., 1991b):

\[ [\theta]_c = -45 \times \text{Temp(°C)} \]  
(4.5).

4.3.3.1 Analysis using the Two-State model

Table 4.5 summarizes the free energy change (\(\Delta G\) and \(\Delta AG\)) calculated at 0°C using the Two-State model (see Chapter 3, section 3.4.1 for details). The results suggest that the H to CH\(_3\) substitution destabilizes the helix by about 0.8 to 1.9 Kcal.

When RlnK is plotted vs. 1/T, the temperature dependence of \(\Delta G\) (\(\Delta G = -RT\ln K\)) can be interpreted to yield information on \(\Delta S\) and \(\Delta H\). According to van't Hoff relationship, RlnK vs. 1/T should be linear; however, as shown in Figures 4.5 and 4.6, our data is not. Due to the non-linear nature of our results, the van't Hoff coefficients for our system are very difficult to interpret, and further analysis is not worthwhile.

4.3.3.2 Analysis using Lifson-Roig theory

In this work, different levels of Lifson-Roig analysis have been applied: simple uniform helix analysis, modified uniform helix analysis, and full helix analysis. The \(<w>\) or \(w'\) values were calculated by comparing the experimental data and theoretical predictions (see Chapter 3, section 3.4.2 for details).
Table 4.6 summarizes the results obtained using the simple uniform helix analysis. The ΔΔG values vary from 1.8 to 4.1 Kcal, this is about 2.5 times larger than the values obtained using the Two-State model. When using the more complex modified uniform helix analysis, as can be seen in Table 4.7, only in a few cases are positive w' values determined. The w' term corresponds to the probability that a residue is in a helix, and a negative probability is not real. Accordingly, we cannot determine a ΔG term to associate with the w-weighting for most modified peptides.

Notice that the w term associated with a residue (i) in a helix corresponds to an additional energy term that occurs when the residue is placed within a hydrogen bonded segment. This energy may be attributed both to the hydrogen bond that forms between residues i−2 and i+2, and to any other interactions such as side-chain interactions that occur with the residue i. When the NCH₃ group is placed at residue i, it breaks the hydrogen bond between that residue i and the i−4 residue, which corresponds to the w term of the residue i−2. Thus, to model the perturbation of the N-methyl group on its neighbors and its effect of removing a hydrogen bond, both w'(i) and w'(i−2) values may be modified. If we wish to model unfavorable interaction between that residue and its neighbors, we should modify the w term of the i residue itself. On the other hand, if we wish to model an unfavorable hydrogen bond interaction, we should modify the w of residue i−2.

In our analysis, both alternatives have been investigated. Instead of calculating ΔΔG values, we include additional data columns that show the expected helix content.
that would occur when $w'(i)$ or $w'(i-2)$ are adjusted. As can be seen in Table 4.7, when the $w'(i)$ is set to be zero, the model overestimates the helix content for most peptides. We then adjust the $w'(i-2)$ value. When the $w'(i-2)$ is replaced with the $v$ weighting to reflect the absence of hydrogen bond, i.e. $w'(i-2)=0.048$, the model still overestimates the amount of helix for most peptides. Even when both $w'(i)$ and $w'(i-2)$ are adjusted and set to zero, the model only works for few cases.

When using the full helix analysis (Rohl et al., 1996), a $p$ value of 2.05 for the salt bridge interaction is required in the 24pl peptides to make the model fit the experimental data. This corresponds to a $\Delta G$ for the Glu$^\text{-}$—Arg$^\text{-}$ salt bridge of 390 cal/mole at 0°C, and is similar in energy to the Glu$^\text{-}$—Lys$^\text{-}$ salt bridge (Scholtz et al., 1993). In addition, the $c$ weighting of the modified residue has been set to zero in our calculation, since according to Rohl et al.'s capping definition, the N-methyl amino acid cannot be a C-capping residue because its hydrogen bond donor (NH group) has been removed (Rohl et al., 1996).

In this model we have also challenged the assumption that the helix nucleation parameter $v$ is the same for all residues. N-methylated peptides have unusual backbone conformations that include cis peptide bond linkage. Therefore, changing both $w$ ($w'(i)$ and $w'(i-2)$) and $v$ ($v'(i)$) terms are investigated in full helix analysis. As shown in Table 4.8, when only $w'(i)$ values were adjusted, the model overestimates the ellipticity at 222 nm for most peptides. When only the $w'(i-2)$ values were adjusted, where we replaced the $w'(i-2)$ with the $v$ weighting ($v=0.036$ for full helix model), the model fails. When
both \( w'(i) \) and \( w'(i-2) \) were adjusted, the predicted ellipticities are closer to experimental values. As we start to change the \( v'(i) \) value, the results are improved; however, even for the worst case scenario where \( w'(i) \), \( v'(i) \) and \( w'(i-2) \) are all zero, the analysis still fails for some peptides.

### 4.3.3.3 Application of the Dichroic model

When applying the Dichroic model (see Chapter 3, section 3.4.3), a \( p \) value of 2.96 for the Glu\(^-\)Arg\(^+\) salt bridge interaction is required in the 24p1 peptides to make the predicted ellipticity of 24p1con fit its experimental data. This corresponds to a \( \Delta G \) for the Glu\(^-\)Arg\(^+\) salt bridge of 490 cal/mole at 0°C. As described above, since the N-methyl amino acid cannot have C-capping effect, we also set the c-weighting of the modified residue as zero (\( c'(i)=0.0 \)) in this analysis. All three parameters, \( w(i) \), \( w(i-2) \) and \( v(i) \) were modified using the same approach as in full helix analysis. Unfortunately, the Dichroic model still overestimates the ellipticity for most peptides. As shown in Table 4.9, even when \( w'(i) \), \( v'(i) \) and \( w'(i-2) \) are all zero, only 24p1m11 and 24p2m06 peptides have reasonable predicted values.

### 4.3.4 Isomerization (cis-trans) of the N-methyl peptide bond

One complication in our peptide system is the \textit{cis-trans} isomerization of the N-methyl peptide bond (Goodman et al., 1974). According to the NOE difference experiment for a small model peptide, Tyr-Glu-(N-methylated-Ala)-Ala- NH\(_2\), the \textit{cis} and \textit{trans} peaks for the protons of NCH\(_3\) group can be assigned easily. Figure 4.7 shows
the 1D NMR spectra for some peptides, where we find that the \( \text{cis} : \text{trans} \) ratio in our N-methyl peptides is about 1:4 based on peak intensity.

4.4 Discussion

While the Two-State model is probably not correct for this system, it gives a lower limit on an energy interaction term, and is the simplest model that fits all data (Scholtz et al., 1991a). The Lifson-Roig model is probably more appropriate, but this model only works for few substitutions. The failure of the Lifson-Roig model led us to the realization that the Lifson-Roig model does not properly account for the circular dichroism of the ensemble of the resulting peptide, so the Dichroic model was ultimately adopted to correct this problem. The Dichroic model, however, still does not predict the effect of this modification very well.

When this phenomena was first observed in the 24p1 series of peptides that are stabilized by E→R salt bridges, it was thought that the additional destabilization might be due to the removal of the salt bridge that spanned across the helix where the N-methyl group was located. It was argued that if the N-methyl group broke the helix, it would also break the salt-bridge. This would then leave E and R residues without charge partners, and they, in turn, might disrupt other adjacent salt bridges. This led to the development of the 24p2 series of peptides that do not use salt bridges for helix stabilization. It can be seen that the data from these peptides fit the models just as badly as the 24p1 series of peptides, so disruption of salt bridges is not the only source of the additional destabilization.
The other complication of the N-methyl peptides is the *cis-trans* isomerization. A *cis* peptide linkage cannot be accommodated in a helix, so the existence of this factor could influence our results. We can incorporate the presence of a *cis* peptide bond into the Lifson-Roig model and Dichroic model by setting the helix nucleation parameter of that residue to zero ($v=0$). This, in essence, prevents a *cis* residue from participating in helix formation. It has been shown, however, that even if $v$ is set to zero, which would correspond to 100% *cis* instead of observed 20% *cis*, these models still cannot account for the instability of our modified peptides appropriately.

Another factor that could destabilize the helix is the bulky methyl group. By adopting a different modification that does not include this perturbation, it was hoped that we might obtain better results. Therefore, we tried another modification, one in which the carbonyl (CO) of a specific residue was reduced to a CH$_2$. This modification also removes the ability of the peptide backbone to hydrogen bond; however, the complications due to isomerization and bulky substituting group are eliminated. By comparing different modifications, it is hoped that the details of backbone-backbone interaction in helical peptides may be addressed more clearly.
<table>
<thead>
<tr>
<th>Notation</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p1con</td>
<td>Ac-WGGEAAAREAAAREAAAREAAARA-Amide</td>
</tr>
<tr>
<td>24p1m06</td>
<td>Ac-WGGEAXAREAAREAAAREAAARA-Amide</td>
</tr>
<tr>
<td>24p1m11</td>
<td>Ac-WGGEAAAREAAREAAREAAARA-Amide</td>
</tr>
<tr>
<td>24p1m16</td>
<td>Ac-WGGEAAAREAAREAAREAAREAAREAAARA-Amide</td>
</tr>
<tr>
<td>24p1m21</td>
<td>Ac-WGGEAAAREAAREAAREAAREAAREAAREAAARA-Amide</td>
</tr>
<tr>
<td>24p2con</td>
<td>Ac-WGGRAAARAAAAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2m06</td>
<td>Ac-WGGRAAXAARAAAAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2m11</td>
<td>Ac-WGGRAARAXAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2m16</td>
<td>Ac-WGGRAARAAAAARAXAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2m21</td>
<td>Ac-WGGRAAAARAAAAARAAAAARAXAAR-Amide</td>
</tr>
</tbody>
</table>

X is N-methyl Ala, and the mxx suffix indicates the position of the N-methyl amino acid.

Table 4.1: The notation of N-methyl modified peptides.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Column</th>
<th>Mobile Phase</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Solvent A</td>
</tr>
<tr>
<td>24p1con</td>
<td>C18</td>
<td>AcCN</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p1m06</td>
<td>C8</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p1m11</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p1m16</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p1m21</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p2con</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p2m06</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p2m11</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p2m16</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
<tr>
<td>24p2m21</td>
<td>C18</td>
<td>MeOH</td>
<td>H₂O</td>
</tr>
</tbody>
</table>

Both solvent A and B contain 0.1% TFA, MeOH: Methanol (B&J High purity solvent, cat 230-4), AcCN: Acetonitrile (B&J High purity solvent, cat 015-4). Both C8 (ZORBAX C8, EX 1531) and C18 (ZORBAX ODS, BX 2198) Columns are semi-preparative size (9.4mm x 25cm).

Table 4.2: Purification conditions for N-methyl modified peptides using HPLC.
Peptide & $[\theta]_{\text{obs}}$ & $[\theta]'_{\text{obs}}$ & $f_{H,\text{obs}}$ & Reduction of $f_{H,\text{obs}}$
\hline
24p1con & -28100 & -27000 & 0.82 & \\
24p1m06 & -17500 & -16800 & 0.51 & 31% \\
24p1m11 & -11100 & -10500 & 0.32 & 50% \\
24p1m16 & -8300 & -8000 & 0.24 & 58% \\
24p1m21 & -13900 & -13300 & 0.40 & 42% \\
24p2con & -29800 & -28600 & 0.86 & \\
24p2m06 & -20400 & -19600 & 0.59 & 27% \\
24p2m11 & -5740 & -5510 & 0.17 & 69% \\
24p2m16 & -5420 & -5200 & 0.16 & 70% \\
24p2m21 & -17200 & -16500 & 0.50 & 36% \\

Data for 24p1 peptides were measured in 10 mM NaCl, 1 mM phosphate buffer at pH7, and data for 24p2 peptides were obtained in 1 M NaCl, 1 mM phosphate buffer at pH7. $[\theta]_{\text{obs}} = \text{total ellipticity}/n_{\text{AA}}$(i.e. number of amino acids, $n_{\text{AA}}=24$ for our peptides), $[\theta]'_{\text{obs}} = \text{total ellipticity}/n_{\text{amide}}$(i.e. number of amides, $n_{\text{amide}}=25$ for our peptides), and

$$f_{H,\text{obs}} = \frac{[\theta]_{\text{obs}} - [\theta]_C}{[\theta]_H - [\theta]_C} = \frac{[\theta]'_{\text{obs}}}{-33088} \quad (\text{using } k=4.32, [\theta]_H=-40000, \text{and } [\theta]_C=0).$$

Table 4.3: Mean residue ellipticity ($[\theta]_{\text{obs}}$), mean amide ellipticity ($[\theta]'_{\text{obs}}$), helix content ($f_{H,\text{obs}}$) and reduction of helix content for N-methyl modified peptides at 0°C.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>$f_{H, obs}$, 10mM NaCl</th>
<th>$f_{H, obs}$, 1M NaCl</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p2con</td>
<td>0.68</td>
<td>0.86</td>
<td>0.18</td>
</tr>
<tr>
<td>24p2m06</td>
<td>0.48</td>
<td>0.59</td>
<td>0.11</td>
</tr>
<tr>
<td>24p2m11</td>
<td>0.11</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>24p2m16</td>
<td>0.11</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>24p2m21</td>
<td>0.47</td>
<td>0.50</td>
<td>0.03</td>
</tr>
</tbody>
</table>

24p2 peptides in 10mM NaCl and 1M NaCl (1mM phosphate buffer at pH7 and 0°C), where $f_{H, obs} = \frac{[\theta]_{obs} - [\theta]_c}{[\theta]_H - [\theta]_c} = \frac{[\theta]_{obs}'}{33088}$ (using $k=4.32$, $[\theta]_H^\infty = -40000$, and $[\theta]_c=0$).

Table 4.4: Comparison of helix content for 24p2mxx peptides under different NaCl concentrations.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>$\Delta G$</th>
<th>$\Delta \Delta G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p1con</td>
<td>-810</td>
<td>-</td>
</tr>
<tr>
<td>24p1m06</td>
<td>-20</td>
<td>790</td>
</tr>
<tr>
<td>24p1m11</td>
<td>410</td>
<td>1220</td>
</tr>
<tr>
<td>24p1m16</td>
<td>620</td>
<td>1430</td>
</tr>
<tr>
<td>24p1m21</td>
<td>210</td>
<td>1020</td>
</tr>
<tr>
<td>24p2con</td>
<td>-1000</td>
<td>-</td>
</tr>
<tr>
<td>24p2m06</td>
<td>-200</td>
<td>800</td>
</tr>
<tr>
<td>24p2m11</td>
<td>870</td>
<td>1870</td>
</tr>
<tr>
<td>24p2m16</td>
<td>910</td>
<td>1910</td>
</tr>
<tr>
<td>24p2m21</td>
<td>3</td>
<td>1000</td>
</tr>
</tbody>
</table>

$\Delta G$, $\Delta \Delta G$ (cal/mole)

Table 4.5: Analysis using the Two-State model for N-methyl modification at 0°C.
### Table 4.6: Simple uniform helix analysis for N-methyl modification at 0°C.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>(&lt;w&gt;)</th>
<th>-RT(\ln)&lt;(w&gt;)</th>
<th>(\Delta\Delta G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p1con</td>
<td>1.417</td>
<td>-190</td>
<td>_</td>
</tr>
<tr>
<td>24p1m06</td>
<td>1.221</td>
<td>-110</td>
<td>1780</td>
</tr>
<tr>
<td>24p1m11</td>
<td>1.140</td>
<td>-70</td>
<td>2600</td>
</tr>
<tr>
<td>24p1m16</td>
<td>1.104</td>
<td>-50</td>
<td>2980</td>
</tr>
<tr>
<td>24p1m21</td>
<td>1.177</td>
<td>-90</td>
<td>2210</td>
</tr>
<tr>
<td>24p2con</td>
<td>1.485</td>
<td>-210</td>
<td>_</td>
</tr>
<tr>
<td>24p2m06</td>
<td>1.260</td>
<td>-130</td>
<td>1960</td>
</tr>
<tr>
<td>24p2m11</td>
<td>1.059</td>
<td>-30</td>
<td>4030</td>
</tr>
<tr>
<td>24p2m16</td>
<td>1.053</td>
<td>-30</td>
<td>4100</td>
</tr>
<tr>
<td>24p2m21</td>
<td>1.217</td>
<td>-110</td>
<td>2380</td>
</tr>
</tbody>
</table>

\(\Delta\Delta G(\text{cal/mole})\)

\(\Delta G = (N-4)[ -RT(\ln<\text{<w>}_{\text{modified}} - \ln<\text{<w>}_{\text{control}}})] \)
<table>
<thead>
<tr>
<th>Peptide</th>
<th>i</th>
<th>f_{i,obs}</th>
<th>w'(i)</th>
<th>ΔΔG</th>
<th>f_{i,cal}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>w'(i)=0</td>
<td>w'(i-2)=0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>w'(i-2)=0.048</td>
<td>w'(i-2)=0.0</td>
</tr>
<tr>
<td>24p1con</td>
<td>-</td>
<td>0.82</td>
<td>1.417</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24p1m06</td>
<td>6</td>
<td>0.51</td>
<td>0.004</td>
<td>3200</td>
<td>0.50</td>
</tr>
<tr>
<td>24p1m11</td>
<td>11</td>
<td>0.32</td>
<td>0.007</td>
<td>2900</td>
<td>0.29</td>
</tr>
<tr>
<td>24p1m16</td>
<td>16</td>
<td>0.24</td>
<td>-0.028</td>
<td>....</td>
<td>0.35</td>
</tr>
<tr>
<td>24p1m21</td>
<td>21</td>
<td>0.40</td>
<td>-0.263</td>
<td>....</td>
<td>0.61</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>0.86</td>
<td>1.485</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24p2m06</td>
<td>6</td>
<td>0.59</td>
<td>0.020</td>
<td>2300</td>
<td>0.57</td>
</tr>
<tr>
<td>24p2m11</td>
<td>11</td>
<td>0.17</td>
<td>-0.018</td>
<td>....</td>
<td>0.35</td>
</tr>
<tr>
<td>24p2m16</td>
<td>16</td>
<td>0.16</td>
<td>-0.035</td>
<td>....</td>
<td>0.42</td>
</tr>
<tr>
<td>24p2m21</td>
<td>21</td>
<td>0.50</td>
<td>-0.196</td>
<td>....</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*i*: position of the N-methyl amino acid

ΔΔG(cal/mole), f_{i,cal}=f_{i-bond}=<n_{i}>/N_{i} (see Chapter 2, section 2.4.4)

Models with reasonable helix content (at or below experimental helix content) are indicated in bold

Table 4.7: Modified uniform helix analysis for N-methyl modification at 0°C.
### Table 4.8: Full helix analysis for N-methyl modification at 0°C.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>( i )</th>
<th>Exp. ([\theta]_{222nm} )</th>
<th>Calculated ellipticity at 222nm ( ([\theta]_{222nm, \text{cal}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>w'(i)=0.0</td>
<td>w'(i-2)=0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>w'(i)=0.0</td>
<td>w'(i-2)=0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>w'(i)=0.0</td>
<td>w'(i-2)=0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>w'(i)=0.0</td>
<td>w'(i-2)=0.0</td>
</tr>
<tr>
<td>24p1con</td>
<td>-</td>
<td>-27000</td>
<td>-27000</td>
</tr>
<tr>
<td>24p1m06</td>
<td>6</td>
<td>-16800</td>
<td>-16000</td>
</tr>
<tr>
<td>24p1m11</td>
<td>11</td>
<td>-10500</td>
<td>-6200</td>
</tr>
<tr>
<td>24p1m16</td>
<td>16</td>
<td>-7960</td>
<td>-8400</td>
</tr>
<tr>
<td>24p1m21</td>
<td>21</td>
<td>-13300</td>
<td>-18800</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>-28600</td>
<td>-28600</td>
</tr>
<tr>
<td>24p2m06</td>
<td>6</td>
<td>-19600</td>
<td>-20500</td>
</tr>
<tr>
<td>24p2m11</td>
<td>11</td>
<td>-5500</td>
<td>-9400</td>
</tr>
<tr>
<td>24p2m16</td>
<td>16</td>
<td>-5200</td>
<td>-11400</td>
</tr>
<tr>
<td>24p2m21</td>
<td>21</td>
<td>-16500</td>
<td>-22000</td>
</tr>
</tbody>
</table>

\( i \): position of the N-methyl amino acid

\( p=2.05 \) for 24p1 peptides, C-capping parameter is zero (i.e. \( c'(i)=0.0 \)) for N-methylated amino acid. Models with reasonable helix content (at or below experimental helix content) are indicated in bold. Underline indicates calculations that are within 10% of experimental values.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>$i^*$</th>
<th>Exp. $[\theta]_{222nm}$</th>
<th>Calculated ellipticity at 222nm ($[\theta]_{222nm, cal}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$w'(i)=0.0$</td>
</tr>
<tr>
<td>24p1con</td>
<td>-</td>
<td>-27000</td>
<td>-27000</td>
</tr>
<tr>
<td>24p1m06</td>
<td>6</td>
<td>-16800</td>
<td>-20500</td>
</tr>
<tr>
<td>24p1m11</td>
<td>11</td>
<td>-10500</td>
<td>-12100</td>
</tr>
<tr>
<td>24p1m16</td>
<td>16</td>
<td>-7960</td>
<td>-11500</td>
</tr>
<tr>
<td>24p1m21</td>
<td>21</td>
<td>-13300</td>
<td>-19600</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>-28600</td>
<td>-26900</td>
</tr>
<tr>
<td>24p2m06</td>
<td>6</td>
<td>-19600</td>
<td>-20700</td>
</tr>
<tr>
<td>24p2m11</td>
<td>11</td>
<td>-5500</td>
<td>-12100</td>
</tr>
<tr>
<td>24p2m16</td>
<td>16</td>
<td>-5200</td>
<td>-11800</td>
</tr>
<tr>
<td>24p2m21</td>
<td>21</td>
<td>-16500</td>
<td>-20100</td>
</tr>
</tbody>
</table>

*i* : position of the N-methyl amino acid

$p=2.96$ for 24p1 peptides, C-capping parameter is zero (i.e. $c'(i)=0.0$) for N-methylated amino acid

Models with reasonable helix content (at or below experimental helix content) are indicated in bold. Underline indicates calculations that are within 10% of experimental values.

Table 4.9: Analysis using Dichroic model for N-methyl modification at 0°C.
Figure 4.1: Melting curves measured by CD at 222 nm for 24p1m peptides in 10 mM NaCl and 1 mM phosphate buffer at pH 7.
Figure 4.2: Melting curves measured by CD at 222 nm for 24p2m peptides in 10mM NaCl and 1 mM phosphate buffer at pH 7.
Figure 4.3: Melting curves measured by CD at 222 nm for 24p2m peptides in 1 M NaCl and 1 mM phosphate buffer at pH 7.
Figure 4.4: NaCl concentration dependence of 24p2con peptide at 0°C and pH 7.
Figure 4.5: RlnK vs. 1/T plot for 24p1m peptides (10 mM NaCl, pH 7) with 95% confidence level linear regression lines.
Figure 4.6: $\ln K$ vs. $1/T$ plot for 24p2m peptides (1 M NaCl, pH 7) with 95% confidence level linear regression lines.
Figure 4.7: A portion of $^1$H NMR spectra for 24p1m21 and 24p2m21 showing the trans and cis N-methyl peaks.
5.1 Introduction

The backbone hydrogen bonds of a helix can be disrupted in two ways. On one hand, the hydrogen bond donor ability of a specific peptide bond can be removed by replacing the NH group with an NCH$_3$ group, as described in chapter 4. On the other hand, the hydrogen bond accepting ability of a residue can be eliminated by replacing the CO group with a CH$_2$ group. In this chapter, the modification of the CO group to a CH$_2$ will be studied.

In order to simplify the system, the model peptide without the salt bridge interaction was chosen. The effects of this modifications on α-helix formation were measured by CD and analyzed using the helix-coil transition theory. This should allow us to study the change in energy associated with the replacement of the peptide bond with this modification.

The CH$_2$ modification removes the ability of a residue to form the hydrogen bond without introducing a bulky methyl group into the helix backbone. It also does not have the cis-trans isomerization problem that the N-methyl modification does. Comparing the
differences between the N-methyl modification and reduced carbonyl modification should help to quantitate the backbone-backbone interaction in helical peptides, and may allow one to separate the hydrogen bonding interaction from the other interactions.

5.2 Material and method

5.2.1 Peptide design

Only the 24p2 peptide described in chapter 4, Ac-WGGRAAAARAAAARAAAAARAAAAAR-Amide, has been used in the study of the reduced carbonyl modifications. The position of modification was arranged so that the same hydrogen bond was removed as the N-methyl modification in chapter 4. For example, the hydrogen bond between the seventh Ala (Ala7) and the eleventh Ala (Ala11) in 24p2m11 peptide was removed by replacing the NH of Ala11 with NCH3. The same hydrogen bond would be removed by replacing the CO of Ala7 with CH2 in the 24p2k07 peptide. Here the kxx suffix indicates the position of the reduction of carbonyl group; thus 24p2k07 refers to the reduction of residue 7's CO to CH2. The sequence and notation of these peptides are listed in Table 5.1.

5.2.2 Peptide synthesis and purification

Peptides with the reduced carbonyl were synthesized and purified by Princeton Biomolecules. The amino acid composition was confirmed by amino acid analysis, and the mass of each peptide was determined by MALDI-TOF mass spectrometry to be within two mass units of the calculated values.
5.2.3 CD measurements

CD samples were prepared by diluting aqueous stock solutions of peptide with 1 M NaCl and 1 mM buffer at pH 4 (formic acid buffer), pH 7 (phosphate buffer) or pH 10 (boric acid buffer). The concentration of peptide stock solution was determined by tryptophan absorbance at 280 nm using $\varepsilon_{280\text{nm}} = 5559 \text{ M}^{-1}\text{cm}^{-1}$ at neutral pH (Mihalyi, 1968). The UV data were measured on a Uvikon 860 spectrometer by Kontron Instruments with a CD cell holder.

CD measurements were made on a computer controlled Jasco J-500A spectropolarimeter using 1 cm quartz cell. Temperature control was achieved to within ±0.2°C using a circulating water bath system.

5.2.4 NMR measurements

NMR samples with final concentration around 1.5 mM to 2 mM were prepared in D$_2$O with 50 mM NH$_4$HCO$_3$ buffer or 50 mM formic acid buffer. The pH* values (uncorrected pH meter reading) were adjusted by adding DCl or KOD.

Notice that the pH reading on the pH meter is 0.4 units lower than the true pD value due to the isotope effect on the glass electrode (Glasoe & Long, 1960). However, there is also the deuterium isotope effect on the ionizaiton equilibrium of the ionizable group in the sample that would raise the pKa values 0.5-0.7 units. This tends to offset the error obtained from not correcting the pH meter reading (Glasoe & Long, 1960; Norman et al., 1961). Thus, in this study, we just use the simple pH meter reading.
NMR measurements were carried out on a Bruker DMX-600 NMR spectrometer. For each pH*, an 1D spectrum was acquired at room temperature. The 2D TOSCY was only obtained at certain pH* values to help in doing the assignments. The residual water peak was suppressed by presaturation during the relaxation delay.

Chemical shift data from 1D NMR as a function of pH* values were fitted to equation (5.1) using the non linear regression method. Accordingly, the pK* value can be obtained (Dyson et al., 1991).

\[
\delta = \delta_{HA} - \frac{\delta_{HA} - \delta_A}{1 + 10^{n(pK_a - \text{pH})}}
\]  

(5.1)

Where \(\delta\) represents the chemical shift value, \(\delta_{HA}\) and \(\delta_A\) are the chemical shift values for the low and high pH limits, and \(n\) is the number of protons transferred with the given pK*.

5.3 Results and Data Analysis

5.3.1 Determination of pK* values for the CH2 modified peptides using 1DNMR

Due to the reduction of carbonyl group, the peptide bond becomes \(\text{C}_\alpha\text{CH}_2\text{NH}_\text{C}_{\alpha+1}\). This introduces a secondary amine in the sequence that will be protonated to \(\text{C}_\alpha\text{CH}_2\text{NH}_2\text{C}_{\alpha+1}\) when the peptide solution's pH is lower than its pK* value. To determine the pK* value for CH2 modified peptides, the chemical shift at different pH* values were obtained. Figure 5.1 shows the titration curves for 24p2k07 peptide at 298K, where the chemical shift of the CH3 protons and the CH2 protons near the modification are plotted against the pH*. The results give pK* values of 6.10 (from
The similar pKₐ value (pKₐ ≈ 6.27) for 24p2k12 was observed using the same method. Surprisingly, the pKₐ value for the reduced carbonyl peptides (6.20 ± 0.10) is about 2 units lower than the secondary amine in dipeptide, CH₃NHCCH₃CONH₂, whose pKₐ is about 8.34.

5.3.2 pH dependence of the CH₂ modification

Figures 5.2 shows the CD spectra obtained in 1.0 M NaCl at 0°C under different pH values: pH 4, pH 7 and pH 10. As shown in Figure 5.2, the pH dependence for our peptide system is negligible. Even at pH 4, where the CαiCH₂NH₂⁻Cαi⁻1 should introduce an additional charge on peptide backbone, there is no significant difference in CD spectra. This might due to the high concentration of NaCl in the solution that could screen out the charge interactions (Huyghues-Despointes et al., 1993).

5.3.3 Effects of CH₂ modification

Melting curves measured by CD at 222 nm for the model peptide and reduced carbonyl modified peptides at pH 7 are shown in Figure 5.3. The mean residue ellipticity at 222 nm ([θ]_obs, calculated using equation (3.2)) in units of deg·cm²·dmol⁻¹ are shown. The aromatic contribution to the CD at 222 nm is ignored since two glycine residues are inserted between tryptophan residue and the rest of helical sequence (Chakrabarty et al., 1993b). Data obtained by heating and cooling experiments are identical indicating the reversibility of the thermal transition from 0°C to 80°C.
The effect of this modification in different positions of the sequence is shown in Table 5.2. The reduction of helix content at pH 7 and 0°C was calculated from mean amide bond ellipticity using equation (4.1). It can be seen that the destabilization of CH$_2$ group replacement is significant and position dependent.

5.3.4 Thermodynamic properties of the CH$_2$ backbone modification

When studying the effect of CH$_2$ modification in terms of energy change, we focus on the data obtained at pH 7 and analyzed using helix-coil theory as described in Chapter 3 (see Chapter 3, section 3.4). The temperature dependence of $[\theta]_H$ and $[\theta]_C$ follows equations (4.4) and (4.5) (see Chapter 4, section 4.3.3).

5.3.4.1 Analysis using the Two-State model

Table 5.3 summarizes the free energy change ($\Delta G$ and $\Delta AG$) due to the CO to CH$_2$ substitution at 0°C using the Two-State model (see Chapter 3, section 3.4.1 for details). The results show that the CH$_2$ modifications destabilize the helix by about 1.1 to 1.9 Kcal.

The non-linear relation between $R\ln K$ and $1/T$ can be seen in Figure 5.4, thus, no further van't Hoff analysis was applied.

5.3.4.2 Analysis using Lifson-Roig theory

To do the Lifson-Roig analysis, we have applied three levels of sophistication: simple uniform helix analysis, modified uniform helix analysis and full helix analysis. The $<w>$ or $w'$ values were calculated by comparing the experimental data and theoretical predictions (see Chapter 3, section 3.4.2 for details).
Table 5.4 summarizes the results obtained using the simple uniform helix analysis. The ΔΔG values are in the range of 2.6 to 4.0 Kcal, this is about 2.4 times larger than the values obtained using the Two-State model. When using the more complex modified uniform helix analysis, as can be seen in Table 5.5, no positive w' values obtained.

As mentioned in Chapter 4, the w term associated with residue i in a helix contains both energy that from a hydrogen bond between residues i-2 and i+2, and the energy of any side-chain interactions that may occur. When CH₂ group is placed at residue i, it breaks the hydrogen bond between that residue i and the i+4 residue, which corresponds to the w term of the i+2 residue. Thus, if we wish to model an unfavorable hydrogen bond interaction, we should modify the w of residue i+2. On the other hand, if we wish to model unfavorable interaction between that residue and its neighbors, we should modify the w term of the i residue itself. Both alternatives have been tried in this analysis.

As shown in Table 5.5, when only w'(i) or w'(i+2) is adjusted, the model overestimates the helix content for all peptides. Notice that the w'(i+2) was adjusted to reflect the absence of hydrogen bond, that is, instead of giving the w weighting, a v weighting was assigned (w'(i+2)=0.048). We then modified both w(i) and w(i+2) terms. However, even when both terms are set to zero, the model still overestimates the helix content for 24p2k12 and 24p2k17.

When using the full helix analysis (Rohl et al., 1996), the n weighting of the modified residue is set to zero in our calculation because according to Rohl et al.'s definition, the CH₂ amino acid has lost the ability to act as an N-capping residue (Rohl et
In addition, the CH\textsubscript{2} modified peptides may have unusual backbone conformations, so that modification of the v values are also reasonable. Therefore, changing both w (w'(i) and w'(i+2)) and v (v'(i)) terms are investigated in this analysis. As shown in Table 5.6, when w'(i) values are set to zero, the full helix model overestimates the ellipticity at 222 nm for all modified peptides. When the w term of the i+2 residue was replaced with the v weighting, i.e. w'(i+2)=0.036, the model still fails. When both w'(i) and w'(i+2) were adjusted, the predictions are closer to experimental values, but only when the v'(i) values are also modified do the predictions become reasonable for all peptides.

### 5.3.4.3 Application of the Dichroic model

The Dichroic model was also applied to this peptide system (see Chapter 3, section 3.4.3 for detail). As described above, we set the n-weighting of the modified residue to zero (n'(i)=0.0) because this residue does not have a hydrogen bond accepting CO group, and changing both w (w'(i) and w'(i+2)) and v (v'(i)) terms are investigated in this analysis as well. It can be seen in Table 5.7 that when more terms are adjusted, the predicted values are closer to the experimental data. However, even for the worst case scenario where w'(i), v'(i) and w'(i+2) are all zero, the analysis still fails for the 24p2k07 and 24p2k12 peptides.

### 5.4 Discussion

By comparing the effects of different modifications, it was hoped that the details of the backbone/backbone interactions in helical peptides could be understood. For
instance, it was expected that the CH\textsubscript{2} modification might reflect the helix destabilization that could be assigned to the removal of the hydrogen bond, while the NCH\textsubscript{3} modification would show the additional destabilization that could be assigned to the introduction of the bulky methyl group or to the \textit{cis-trans} isomerization. Table 5.8 shows, however, that no consistent results are obtained between these two modifications. While comparing p2m11 and p2k07, the NCH\textsubscript{3} modification destroys the helix more than CH\textsubscript{2} modification does; however, the p2m16 and p2k12 peptides destroy the helix about the same amount. Moreover, the NCH\textsubscript{3} in p2m21 reduces less helix content than CH\textsubscript{2} in p2k17 does, which is opposite from what was expected.

To help understand this result, we have also inserted proline into our model peptide to see how proline destabilization compares to our backbone modifications. Proline is expected to be similar to the N-methyl modification since there is no hydrogen bond donor, NH group, in a proline residue. We have replaced the N-methyl-Ala at position 11 with a Pro in a model peptide, which removes the same hydrogen bond as the CH\textsubscript{2} modification at Ala 7. The sequence for this modified peptide, 24p2p11, is Ac-WGGRAAARAPAARAAARAAAAAR-NH\textsubscript{2}. While we expected the proline to break the helix like the N-methyl modification, we instead found that proline was a weaker helix breaker (f_{helix}=0.34). Our result shows that the proline substitution destabilizes the helix more than CH\textsubscript{2} modification, but less than the NCH\textsubscript{3} modification.

These results suggest that these backbone modifications not only break the hydrogen bond, but also have other effects that reduce the helix content. For example, the NCH\textsubscript{3} modification introduces a bulky methyl group into the helix, while the CH\textsubscript{2}
modification remove an oxygen that might change the geometry of the peptide bond. The geometry of the planar sp\(^2\) hybridized CO group would be slightly different from the geometry of the sp\(^3\) hybridized CH\(_2\) group. Besides, both NCH\(_3\) modification and Pro substitution involve the cis-trans isomerization that would reduce the helix content as well. It is obvious that it is still difficult to distinguish the different backbone/backbone interactions based on the data obtained here.

We have also tried to analyze the effects of peptide backbone modification using the helix-coil transition theory which should allow us to quantitate the change in total energy associated with these modifications. However, our results show that neither the simple model, like the Two-State model, nor the more complicated models, such as Lifson-Roig model and Dichroic model, can analyze our data appropriately. As can be seen in Tables 5.5, 5.6 and 5.7, only when both the i and i+2 residues are both modified, do we begin to get reasonable results. However, even when all \(w'(i)\), \(w'(i+2)\) and \(v'(i)\) are set to zero, these models can still fail. Therefore, one interpretation for the failures of these models to properly predict the effect of a modification is that the effect of the modification is not confined to only one or two residues. When the unfavorable interactions in multiple residues are incorrectly attributed to one or two residues, the model fails and needs negative \(w\) terms to function. Thus, the models may need to be further modified to mimic the system more closely.
<table>
<thead>
<tr>
<th>Notation</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p2con</td>
<td>Ac-WGGRAAAARAAAAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2k02</td>
<td>Ac-WGGRAAAARAAAAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2k07</td>
<td>Ac-WGGRAAZARAAAAARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2k12</td>
<td>Ac-WGGRAAAARAAZARAAAAARAAAAAR-Amide</td>
</tr>
<tr>
<td>24p2k17</td>
<td>Ac-WGGRAAAARAAAAARAAZARAAAAAR-Amide</td>
</tr>
</tbody>
</table>

*Z* is Ala with reduced carbonyl, *G* is Gly with reduced carbonyl, and the *kxx* suffix indicates the position of the reduced carbonyl amino acid.

Table 5.1: The notation of the reduced carbonyl peptides.
Peptide & $[\theta]_{obs}$ & $[\theta']_{obs}$ & $f_{H,obs}$ & Reduction of $f_{H,obs}$ \\
\hline
24p2con & -29800 & -28600 & 0.86 & - \\
24p2k02 & -15300 & -14700 & 0.44 & 42% \\
24p2k07 & -5800 & -5600 & 0.17 & 70% \\
24p2k12 & -14700 & -14100 & 0.43 & 44% \\
24p2k17 & -14700 & -14100 & 0.43 & 44% \\
\hline

Where $[\theta]_{obs} = \text{total ellipticity/}n_{AA}$ (i.e. number of amino acids, $n_{AA}=24$ for our peptides), $[\theta']_{obs} = \text{total ellipticity/}n_{amide}$ (i.e. number of amides, $n_{amide}=25$ for our peptides), and

$$f_{H,obs} = \frac{[\theta]_{obs} - [\theta]_C}{[\theta]_H - [\theta]_C} = \frac{[\theta]_{obs}}{-33088} \quad (\text{using } k=4.32, [\theta]_H^{\infty}=-40000, \text{and } [\theta]_C=0).$$

Table 5.2: Mean residue ellipticity($[\theta]_{obs}$, $[\theta']_{obs}$), helix content ($f_{H,obs}$) and reduction of helix content for the reduced carbonyl peptides at 0°C.
Table 5.3: Analysis using the Two-State model for reduced carbonyl modification at 0°C.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>ΔG</th>
<th>ΔΔG</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p2con</td>
<td>-1000</td>
<td></td>
</tr>
<tr>
<td>24p2k02</td>
<td>120</td>
<td>1120</td>
</tr>
<tr>
<td>24p2k07</td>
<td>860</td>
<td>1860</td>
</tr>
<tr>
<td>24p2k12</td>
<td>160</td>
<td>1160</td>
</tr>
</tbody>
</table>

ΔG, ΔΔG(cal/mole)
### Table 5.4: Simple uniform helix analysis for reduced carbonyl modification at 0°C.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>&lt;w&gt;</th>
<th>RT ln&lt;w&gt;</th>
<th>ΔΔG</th>
</tr>
</thead>
<tbody>
<tr>
<td>24p2con</td>
<td>1.485</td>
<td>-210</td>
<td>-</td>
</tr>
<tr>
<td>24p2k02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2k07</td>
<td>1.195</td>
<td>-100</td>
<td>2590</td>
</tr>
<tr>
<td>24p2k12</td>
<td>1.061</td>
<td>-30</td>
<td>4010</td>
</tr>
<tr>
<td>24p2k17</td>
<td>1.187</td>
<td>-90</td>
<td>2680</td>
</tr>
</tbody>
</table>

ΔΔG(cal/mole)

ΔΔG = (N−4)[−RT(ln<w>_{modified}−ln<w>_{control})]
<table>
<thead>
<tr>
<th>Peptide</th>
<th>i</th>
<th>f_{i,obs}</th>
<th>w(i)</th>
<th>$\Delta G$</th>
<th>f_{i,cal}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>w(i) = 0</td>
<td>w(i+2) = 0.048</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>0.86</td>
<td>1.485</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24p2k02</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2k07</td>
<td>7</td>
<td>0.44</td>
<td>-0.03</td>
<td>....</td>
<td>0.52</td>
</tr>
<tr>
<td>24p2k12</td>
<td>12</td>
<td>0.17</td>
<td>-0.02</td>
<td>....</td>
<td>0.34</td>
</tr>
<tr>
<td>24p2k17</td>
<td>17</td>
<td>0.43</td>
<td>-0.03</td>
<td>....</td>
<td>0.47</td>
</tr>
</tbody>
</table>

*i: position of the reduced carbonyl amino acid

$\Delta G$ (cal/mole), $f_{i,cal} = f_{i,bond} = \langle n_h \rangle / N_h$ (see Chapter 2, section 2.4.4)
Models with reasonable helix content (at or below experimental helix content) are indicated in bold

Table 5.5: Modified uniform helix analysis for reduced carbonyl modification at 0°C.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>$i$</th>
<th>Exp. $\theta'_{222nm}$</th>
<th>Calculated ellipticity at 222nm ($\theta'_{222nm, ca}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$w'(i)=0.0$</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>-28600</td>
<td>-28600</td>
</tr>
<tr>
<td>24p2k02</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2k07</td>
<td>7</td>
<td>-14700</td>
<td>-17300</td>
</tr>
<tr>
<td>24p2k12</td>
<td>12</td>
<td>-5600</td>
<td>-7900</td>
</tr>
<tr>
<td>24p2k17</td>
<td>17</td>
<td>-14100</td>
<td>-14400</td>
</tr>
</tbody>
</table>

*i*: position of the reduced carbonyl amino acid

N-capping parameter is zero (i.e. $n'(i)=0.0$) for reduced carbonyl amino acid

Models with reasonable helix content (at or below experimental helix content) are indicated in bold. **Underline** indicates calculations that are within 10% of experimental values.

Table 5.6: Full helix analysis for reduced carbonyl modification at 0°C.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>( i )</th>
<th>Exp. ([\Theta]_{222\text{nm}})</th>
<th>Calculated ellipticity at 222nm (([\Theta]_{222\text{nm, cal}}))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( w'(i)=0.0 )</td>
<td>( w'(i+2)=0.036 )</td>
<td>( w'(i)=0.0 )</td>
</tr>
<tr>
<td></td>
<td>( w'(i+2)=0.036 )</td>
<td>( \nu'(i)=0.0 )</td>
<td>( \nu'(i)=0.0 )</td>
</tr>
<tr>
<td>24p2con</td>
<td>-</td>
<td>-28600</td>
<td>-26900</td>
</tr>
<tr>
<td>24p2k02</td>
<td>2</td>
<td>-14700</td>
<td>-18300</td>
</tr>
<tr>
<td>24p2k07</td>
<td>7</td>
<td>-5600</td>
<td>-10400</td>
</tr>
<tr>
<td>24p2k12</td>
<td>12</td>
<td>-14100</td>
<td>-13800</td>
</tr>
<tr>
<td>24p2k17</td>
<td>17</td>
<td>-14700</td>
<td>-18300</td>
</tr>
</tbody>
</table>

\( i \) : position of the reduced carbonyl amino acid

N-capping parameter is zero (i.e. \( n'(i)=0.0 \)) for reduced carbonyl amino acid

The reasonable ellipticities are indicated in bold (less than experimental data) or underline (within 10% error)

Table 5.7: Analysis using Dichroic model for reduced carbonyl modification at 0°C.
The hydrogen bond between residue a→b has been removed.

Table 5.8: Comparison of helix content for N-methyl modification (NCH$_3$) and reduced carbonyl modification (CH$_2$).

<table>
<thead>
<tr>
<th>Modification</th>
<th>$f_{i,obs}$</th>
<th>2→6</th>
<th>7→11</th>
<th>12→16</th>
<th>17→21</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCH$_3$</td>
<td>0.59(p2m06)</td>
<td>0.17(p2m11)</td>
<td>0.16(p2m16)</td>
<td>0.50(p2m21)</td>
<td></td>
</tr>
<tr>
<td>CH$_2$</td>
<td>0.44(p2k07)</td>
<td>0.17(p2k12)</td>
<td>0.43(p2k17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig 5.1: 24p2k07 titration at 298K: $^1$H NMR chemical shift data for the modified residue (Ala7).
Figure 5.2: CD spectra for 24p2k peptides at different pH values (1 M NaCl, 0°C), where \([\theta]\) (deg cm\(^2\) dmol\(^{-1}\)) is the mean residue ellipticity.
Figure 5.3: Melting curves measured by CD at 222 nm for 24p2k peptides in 1 M NaCl and 1 mM phosphate buffer at pH 7.
Figure 5.4: RlnK vs. 1/T plot for 24p2k peptides (1 M NaCl, pH 7) with 95% confidence level linear regression lines.
CHAPTER 6

DISCUSSION AND CONCLUSION

There are two major goals for this study. One is to determine the ability of some modified peptide bonds to be incorporated into an \( \alpha \) helix; the other is to identify and further quantify the different backbone/backbone interactions, especially the strength of the backbone hydrogen bond, by using these modifications as probes. The backbone modifications have been chosen to complement each other: one removes a hydrogen bond donor, \( \text{NH} \) (the N-methyl modification, Chapter 4); the other removes a hydrogen bond acceptor, \( \text{CO} \) (the reduced carbonyl modification, Chapter 5). The primary effect of these modifications is the removal of one single hydrogen bond from the helix backbone.

The ability of these modifications to be accommodated into an \( \alpha \) helix was accomplished by measuring the stability of the modified peptides using CD. In addition, by comparing the data from different types of modifications, it was hoped that we could identify the different backbone interactions along the helix. Unfortunately, the comparison between data sets gives no consistent result.
To quantify the effects of these modifications, several thermodynamic analyses have been applied. However, from our thermodynamic results, neither the simple Two-State analysis nor the most complicated Dichroic model fits the data from our peptide system.

6.1 Effects of backbone modification

As can be seen in Tables 4.3 and 5.2, both the N-methyl group, and the CH$_2$ replacement, act as a helix disrupter. The helix content of all modified peptides are reduced and the amount of reduction depends on the position of the modification.

When comparing the effect between different modification, it was expected that the NCH$_3$ modification would destabilize the helix more than the CH$_2$ modification because the NCH$_3$ modification introduces an additional bulky methyl group into the sequence. However, as shown in Table 5.8, only one pair of the peptides, p2m11 and p2k07, in which the hydrogen between residues 7 and 11 has been removed, shows this tendency. No consistency is found for other pairs of peptides. This indicates that these modifications not only break the backbone hydrogen bond interaction, but also have other destabilizing effects that depend on the modification itself. Thus, the data obtained here still cannot provide enough information to differentiate between the various interactions along the helix backbone.

Accordingly, more modifications of the peptide backbone are necessary. The next change under synthesis is the COCH$_2$ modification. Like the NCH$_3$ group, this modification also removes the ability of a residue to act as a hydrogen bond donor, but
the replacement of NH group with a CH$_2$ group should cause only minor structural change. Thus, the data should be easier to interpret and should help to identify the different backbone/backbone interactions. It will also be interesting to try the multiple modifications in the future work. That is, instead of replacing only one single peptide bond in the model peptide, multiple modifications in the single peptide can be designed. By comparing the difference between single replacement and multiple modification, more interesting and useful information should be obtained.

6.2 Application of helix-coil theory

Several different models based on helix-coil transition theory have been used to study the thermodynamic properties of these modifications. These models tend to reflect the interactions that contribute to the formation of a helical structure, and it was hoped that the thermodynamic properties of these modifications, such as free energy change, could be derived from these analyses.

The free energy change $\Delta G$ for a system can be calculated from equilibrium constant, $K$, using $\Delta G = -RT\ln K$, thus the energy change due to the modifications can be obtained by comparing the $\Delta G$ of the modified peptides to the $\Delta G$ of the unmodified control peptides. That is, $\Delta \Delta G = \Delta G_{\text{modified}} - \Delta G_{\text{control}}$. To obtain this energy property, we first need to find the equilibrium constant of the helix-coil transition for the peptide at particular temperature.
When applying the simple Two-State model, the equilibrium constant for the helix-coil transition is simply: \( K = [\text{helix}] / [\text{coil}] \) (see Chapter 2, section 2.2). However, to apply the more realistic statistical models, the equilibrium constant must be extracted from the statistical weights. That is, the statistical weights of helical segments must be compared with statistical weights of the random coil. Notice that the random coil here is defined as all conformations that do not include helical segments. For example, one single h residue or two adjacent h residues cannot form a helical segment and, therefore, are part of the ensemble of random coil conformation. Thus, in the random coil, a residue can be in either h or c state with either v or 1 weighting.

In Lifson-Roig theory, to nucleate a helix, three consecutive residues must be converted from the random coil conformation to the helical conformation. In addition, residues on two sides of the helix must be fixed in the c states with capping interactions to terminate the helix. Accordingly, the equilibrium constant for nucleating a helix is expressed as: \( K = w v^2 / (1+v)^3 \). The equilibrium constant for adding a residue to an existing helical segment, however, can be simply expressed as: \( K = w / (1+v) \) (Rohl et al., 1996). When applying to our peptide system, if only the \( w \) values have been modified, then the total energy change due to the modification can be calculated using

\[
\Delta \Delta G = - \sum_i RT \ln \left( \frac{K_{\text{modified}}}{K_{\text{control}}} \right) = - \sum_i RT \ln \left( \frac{w'_i}{w_i} \right)
\]

(6.1).

Unfortunately, according to our results, only the simple Two-State model can analyze all peptides and obtain the \( \Delta \Delta G \) values (Tables 4.5 and 5.3); the more detailed models fail for most peptides due to the negative statistical weights (negative \( w' \) term).
It has been shown, however, that the helix-coil transition in small peptides is a non two-state system, thus the Two-State analysis is not appropriate for our peptide system (Scholtz et al., 1991a). While there are only few peptides that can be analyzed using the more detailed models, the results in Table 6.1 indicate that by adding more detail to the analysis, more energy is attributed to the modification. For example, the ΔΔG value for 24p2k07 found using the full helix analysis is roughly triple the value derived from the Two-State model, and about ten to twenty percent larger than the values obtained using the uniform helix models. While the 24p2k17 is the only peptide that can be analyzed using the Dichroic model, its ΔΔG derived from the Dichroic model is roughly a factor of four larger than the Two-State analysis.

It appears that as the models increase in complexity, the energy associated with the single residue modification increase. We believe that the simple models underestimate the energy contribution, and the Dichroic model is the one that describes the helix-coil transition most precisely. However, it only works for one peptide, thus, the model still need to be improved.

6.3 Model improvement

To apply the Dichroic model using Lifson-Roig statistics, the intrinsic parameters (w, v, n and c) first need to be derived. However, the data sets available in literature are not sufficient to assure that the best values have been obtained. In particular, there are only a few data sets available for Glu, Arg, and Trp, yet parameters for these residues are important in analyzing our peptide system. The resulting errors due to the error from
parameters could be very large. Thus, more experimental data are necessary to obtain the best-fit parameters and, therefore, to improve the accuracy of our prediction.

On the other hand, the model itself might need further modification to mimic our system more closely. One adjustment will be to separate the hydrogen bond interaction from the other interactions. Notice that we have focused on the change of w weights, since the w-weighting corresponds to the presence of the backbone hydrogen bond. However, how the w-weighting relates to the removal of one backbone hydrogen bond must be examined more carefully. According to the original Lifson-Roig theory, the w weight of a residue was derived from two energy terms, one of them is independent of the other residues (ex: potential of hindered rotation, the energy of solvent-peptide interaction…...), the other term is related to its neighbor residues (ex: backbone hydrogen bond, side-chain/side-chain interaction which help to stabilized the helix…….). Notice that the w derived parameter are unique for each residue, yet the backbone hydrogen bond strength should be the same for all residues. Also, since our modifications are designed to remove the backbone hydrogen bond, it is reasonable to separate the hydrogen bond interaction from the other interactions that involve in the w term.

In addition, it is also reasonable to expect that different residue will have different v values, especially for residues with unusual backbone conformations, such as proline and our modified residues. Thus, instead of giving all amino acids the same v-weighting, determining the v value for each amino acid might be necessary.
During the past few years, the Lifson-Roig theory has been continuously modified to fit the experimental results better. The Dichroic model is the most recent modification, however, it is still not complete enough. Our study reveals some problems that suggest further modifications and should help to reach the goal that the model can predict helix formation from primary structure.
<table>
<thead>
<tr>
<th>Peptides</th>
<th>Two-State model</th>
<th>Simple uniform helix analysis</th>
<th>Modified uniform helix analysis</th>
<th>Full helix analysis</th>
<th>Dichroic model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔΔG</td>
<td>&lt;w&gt;</td>
<td>ΔΔG</td>
<td>w(i)</td>
<td>ΔΔG</td>
</tr>
<tr>
<td>24p1m06</td>
<td>0.8</td>
<td>1.221</td>
<td>1.8</td>
<td>0.035</td>
<td>3.6</td>
</tr>
<tr>
<td>24p1m11</td>
<td>1.2</td>
<td>1.140</td>
<td>2.6</td>
<td>0.234</td>
<td>2.6</td>
</tr>
<tr>
<td>24p1m16</td>
<td>1.4</td>
<td>1.104</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p1m21</td>
<td>1.0</td>
<td>1.177</td>
<td>2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2m06</td>
<td>0.8</td>
<td>1.260</td>
<td>2.0</td>
<td>0.184</td>
<td>3.0</td>
</tr>
<tr>
<td>24p2m11</td>
<td>1.9</td>
<td>1.059</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2m16</td>
<td>1.9</td>
<td>1.053</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2m21</td>
<td>1.0</td>
<td>1.217</td>
<td>2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2k07</td>
<td>1.1</td>
<td>0.195</td>
<td>2.6</td>
<td>0.206</td>
<td>2.9</td>
</tr>
<tr>
<td>24p2k12</td>
<td>1.9</td>
<td>1.061</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24p2k17</td>
<td>1.2</td>
<td>1.187</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The w'(i) values for modified uniform helix analysis are obtained when w'(i−2) (for 24p1 peptides) or w'(i+2) (for 24p2 peptides) are set to 0.048, and the w'(i) values for full helix analysis and Dichroic model are derived when w'(i−2) (for 24p1 peptides) or w'(i+2) (for 24p2 peptides) are set to 0.036. ΔΔG = −RT[ln(w'(i)/w(i))+ln(w'(i±2)/w(i±2))], where w(i)=w_{Ala}, and w(i−2)=w_{Glu}, w(i+2)=w_{Arg}.

Table 6.1: w values (<w> or w'(i)) and their corresponding ΔΔG (kcal/mole) obtained using different models.
CHAPTER 7

STRUCTURAL CHARACTERIZATION OF THE MOST COMPACT UNIT
FROM STAPHYLOCOCCAL NUCLEASE

One way to approach the protein folding problem is to try to identify domains that can fold independently. If these domains can be isolated and identified, then the higher folding process in which domains are joined together would be simplified. However, the properties that define a domain and its boundaries have not been well established.

There are several different models available that can be used to identify domains, but few experiments have been done to test these models. In this study, the protein staphylococcal nuclease was chosen to test one of these models: the compact unit theory.

7.1 Background information

A domain is defined as any polypeptide chain, or part of a polypeptide chain that can independently fold into a stable tertiary structure (Branden & Tooze, 1991). Based on this generally accepted concept, several algorithms that identify domains have been
developed. These theoretical methods locate domains based on various properties, such as distance maps (Liljas & Rossman, 1974; Go, 1983), clustering (Crippen, 1978), plane cutting (Rose, 1979), minimization of interface area (Wodak & Janin, 1981; Janin & Wodak, 1983), minimization of specific volume (Lesk & Rose, 1981), maximization of solvent exclusion (Rashin, 1981), and compactness (Zehfus & Rose, 1986; Zehfus, 1987, 1993).

It is important to determine which theoretical model is the most appropriate one, because this implies its corresponding properties are more important when identifying domains. When applied to proteins, these models all agree reasonably well on the location of large domains, but differences arise when identifying smaller subdomains. Here, a model based on compactness, the compact unit theory (Zehfus & Rose, 1986), was chosen for testing because its domain locating algorithm finds subdomain-sized units that are very different from those of the other methods.

7.1.1 Compact unit theory

Compact unit theory uses compactness to identify domains in proteins. In this model, compactness is quantified using a normalized measure of surface area called the coefficient of compactness, \( Z \), as

\[
Z = \frac{\text{accessible surface area of segment}}{\text{surface area of sphere of equal volume}} \quad (7.1)
\]

Since a sphere has a minimum surface area for its enclosed volume, \( Z \) is interpreted as a segment's actual area normalized by its minimum possible area. According to equation
A value of $Z=1.0$ refers to a perfect sphere, and the closer the $Z$ value is to 1.0, the more compact the unit is (Zehfs & Rose, 1986).

A domain should be at least as compact as native proteins, thus, any unit with $Z$ values less than the value of native proteins is considered as a compact unit. It has been shown that the average $Z$ value of an globular protein is 1.68 with a standard deviation of 0.12 (Zehfs, 1993). Accordingly, a $Z$ value threshold of 1.55, i.e. one standard deviation below the average compactness of a globular protein, has been used as a reference value to define compact units (Zehfs, 1994).

7.1.2 Application of compact unit theory

If the compact unit theory is a good model for locating domains, then the compact units identified by this model must have domain properties. In particular, each compact unit should retain its native structure in the absence of the rest of the protein. Thus, the following procedures can be use to test this model:

1. Select a model protein whose structure is well defined;
2. Apply the compact unit theory to identify all compact units within the protein;
3. Obtain peptides that correspond to those compact units;
4. Use CD and NMR to study the structure properties for these peptides;
5. Compare solution structure of the isolated peptide with the structure of that peptide in the native protein.

By comparing the unit's structure in native protein with the structure obtained from its corresponding peptide isolated in solution, we can see how closely the two
structures correspond. This information can then be used to either confirm or modify the compact unit theory.

7.2 Material and methods

7.2.1 Target protein and its compact units

Staphylococcal nuclease was chosen for this study because its structure is well defined by three X-ray structures (Cotton et al., 1979; Loll & Lattman, 1989; Hynes & Fox, 1991) and two NMR structures (Torchia et al., 1989; Wang et al., 1990). In addition, there are no disulfides in the protein that may interfere with the analysis (Figure 7.1). When applying the compactness algorithm to Staphylococcal nuclease, eight compact units were found in this protein. The most compact unit, corresponding to the peptide from residues 12 to 36 (LRFTMPQGYMLKVTGDIAKILTA), with the Z value of 1.48 is discussed here. According to the X-ray structure (Hynes & Fox, 1991), this region forms a three strand β sheet (β–β–β) in the native protein (Figure 7.1, B).

7.2.2 Peptide synthesis and purification

Staphylococcal nuclease [12-36] was synthesized on an Applied Biosystem 430A Solid Phase peptide synthesizer using t-Boc chemistry. Notice that the free N- and C-termini introduce two extra charged groups that are not present in the native protein, and could cause a significant structural perturbation in this small peptide. Thus, peptide's N- and C- termini were blocked by acetylation and amidation, respectively. Synthesis was performed by the Biochemical Instrument Center at the Ohio State University.
The peptide was purified by HPLC using Zorbax C18 reverse phase preparative column with a two solvent system of water and acetonitrile ( solvent A: 60% acetonitrile and 40% water with 0.08% TFA; solvent B: 100% water with 0.1% TFA). A gradient from 55% -100% solvent A over 20 minutes was applied to isolate a single peptide peak. The purified peptide was shown to have >95% purity by HPLC, and the amino acid composition was confirmed by amino acid analysis.

7.2.3 Spectral measurements

Circular dichroism (CD) measurements were made on a computer controlled Jasco J-500A spectropolarimeter. The concentration of the peptide was determined by tyrosine absorbance at 274 nm using $e_{274nm}=1400\, \text{M}^{-1}\cdot\text{cm}^{-1}$ (Mihalyi, 1968). Before each CD measurement, the UV absorbance of the sample was measured on a Uvikon 860 spectrometer with a equipped CD cell holder.

NMR measurements were carried out on a Bruker Aspect 250 using the 5 mm sample tube.

7.3 Results

The large number of hydrophobic residues in the peptide gave it poor solubility, thus a wide variety of solution conditions were tried to help to solubilize the peptide. Unfortunately, conditions where the peptide could form the $\beta-\beta-\beta$ structure without aggregation were never found.
7.3.1 Detergents

Several different kinds of detergent were tried (Figure 7.2), including the anionic detergent sodium dodecyl sulfate (SDS), 1-heptanesulfonic acid, sodium salt and deoxycholic acid, sodium salt; the cationic detergent dodecytrimethylammonium bromide; and the nonionic detergent 4 Laurly ether (Brij 30) and 23 Laurly ether (Brij 35).

As can be seen in Table 7.1, CD data showed the β sheet structure for almost all detergents. However, in most cases, the detergent concentrations needed to be much higher than the peptide concentration, and peptide was aggregated. An attempt was made to determine the peptide's structure in SDS and in dodecytrimethylammonium bromide by 1D NMR; however, in both detergents, proton NMR peaks were not observable, presumably because a high molecule weight aggregation had formed and broaden the line widths until they could not be observed.

7.3.2 Osmolyte solvents

Six osmolyte solvents (Figure 7.3) under various solution conditions were tried to find the β sheet structure: Choline chloride, glycerol, 1,2,3-heptanetriol, 2-methyl-2,4-pentanediol (MPD), sucrose and tetraethylammonium chloride. Table 7.2 summarizes their CD results; unfortunately, random coil was the major structure obtained.
7.3.3 Other solvents

The solvents 2,2,2-trifluoroethanol (TFE), hexafluoroisopropanol (HFIP) and ethanol were also tried (Figure 7.4). The peptide was reasonable soluble in these solvents. However, as shown in Table 7.3, all of them gave the typical \( \alpha \) helix CD spectra rather than the expected \( \beta \) sheet structure. This was not surprised since these solvent have been shown to stabilize the helical structure on other peptides (Nelson & Kallenbach, 1986; Fan et al., 1993; Safar et al., 1993; Arunkumar et al., 1997).

7.4 Summary

After the peptide was obtained, the solvent conditions under which the native structure could occur first had to be determined. According to the X-ray structure, Staphylococcal nuclease [12-36] should adopt a three strand \( \beta \)-sheet secondary structure (Figure 7.1, B). Thus, various solvent conditions were tried to find a system that would give us the \( \beta-\beta-\beta \) structure.

Unfortunately, when the CD showed \( \beta \)-sheet structure, peptide was aggregated and undetectable using NMR. When the solvents where the peptide was completely soluble were used, the CD showed either random or helical structure. Different concentrations, pHs and temperatures have been tested for each solvent; however, conditions where the peptide could form its native structure, i.e. the \( \beta-\beta-\beta \) structure, without aggregation were never found. Due to these frustrating results, no further structural analysis has been accomplished for this peptide.
<table>
<thead>
<tr>
<th>Detergent</th>
<th>CD</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>sodium dodecyl sulfate (SDS)</td>
<td>( \beta ) like</td>
<td>- Room temperature, ([\text{detergent}] / [\text{peptide}]) between 2.5 to 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Structure does not depend on pH (pH 4, 7, 10, 12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Structure undetectable from 1D NMR spectrum</td>
</tr>
<tr>
<td>1-heptanesulfonic acid, sodium</td>
<td>( \beta ) like</td>
<td>- Room temperature, as ([\text{detergent}] / [\text{peptide}]) increased, the structure changed from random (ratio (<del>) 2.0) to (\alpha) helix (ratio (</del>) 130) to (\beta) sheet (ratio (~) 260)</td>
</tr>
<tr>
<td>salt</td>
<td></td>
<td>- (\beta) like structure was observed under different pHs (pH 7, 10, 12), but most sable at high pH (pH 12)</td>
</tr>
<tr>
<td>deoxycholic acid, sodium salt</td>
<td>( \beta ) like</td>
<td>- Room temperature, CD is [peptide] dependent, peptide was aggregated \</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- (\beta) like structure was observed under different pHs (pH 4, 7, 10), but most stable at high pH (pH 12)</td>
</tr>
<tr>
<td>dodecyltrimethylammonium bromide</td>
<td>( \beta ) like</td>
<td>- Room temperature, ([\text{detergent}] / [\text{peptide}]) \ (&gt; 1.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- (\beta) like structure was obtained under different pHs (pH 4, 7, 12), but most stable at high pH (pH 12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Structure undetectable in 1D NMR spectrum</td>
</tr>
<tr>
<td>4-Laurly ether (Brij 30)</td>
<td>random</td>
<td>- Room temperature, only tried ([\text{detergent}] / [\text{peptide}]) (~) 2 at pH 7</td>
</tr>
<tr>
<td>23-Lauryl ether (Brij35)</td>
<td>( \beta ) like</td>
<td>- Room temperature, as ([\text{detergent}] / [\text{peptide}]) increased, the structure changed from random (ratio (<del>) 2.0) to (\alpha) helix (ratio (</del>) 10) to (\beta) sheet (ratio (~) 100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Only tried pH 7</td>
</tr>
</tbody>
</table>

[detergent] and [peptide] refer to the concentration of detergent and the concentration of peptide in solution respectively.

Table 7.1: Summary of detergents and conditions used to find the \(\beta\)-sheet structure for Staphylococcal nuclease [12-36].
<table>
<thead>
<tr>
<th>Osmolyte solvent</th>
<th>CD spectra</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>choline chloride</td>
<td>random</td>
<td>-random at pH 4, 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-helix at pH 10</td>
</tr>
<tr>
<td>glycerol</td>
<td>random</td>
<td>-random at pH 4, 7, 10</td>
</tr>
<tr>
<td>1,2,3-heptanetriol</td>
<td>random</td>
<td>-random at pH 2, 7, 12</td>
</tr>
<tr>
<td>2-methyl-2,4-pentanediol (MPD)</td>
<td>helix</td>
<td>-helix at pH 2, 7, 12</td>
</tr>
<tr>
<td>sucrose</td>
<td>random</td>
<td>-random at pH 4, 7, 10</td>
</tr>
<tr>
<td>tetraethylammonium chloride</td>
<td>random</td>
<td>-random at pH 4, 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-helix at pH 10</td>
</tr>
</tbody>
</table>

Table 7.2: Summary of osmolyte solvents used and their resulting CD spectra for Staphylococcal nuclease [12-36].
<table>
<thead>
<tr>
<th>Solvent</th>
<th>CD spectra</th>
<th>Comments</th>
</tr>
</thead>
</table>
| 2,2,2-trifluoroethanol (TFE)        | \(\alpha\) helix | -higher % of solvent, more stable the helix  
|-structure does not depend on pH (pH 4, 7, 10) |
| hexafluoroisopropanol (HFIP)        | \(\alpha\) helix | -higher % of solvent, more stable the helix  
|-structure does not depend on pH (pH 4, 7, 10) |
| ethanol                             | \(\alpha\) helix | -higher % of solvent, more stable the helix  
|-structure does not depend on pH (pH 4, 7, 10) |

Table 7.3: Other solvents used to find the structure for Staphylococcal nuclease [12-36] and their resulting CD spectra.
A: Overall fold of nuclease, where cylinders and arrows represent α-helices and β-sheets respectively

B: Unrolled the β-barrel of nuclease to illustrate the details of the β-sheets regions. Notice that [12-36] region has the β-β-β structure

(From Hynes, Rhomas R. and Fox, Rober O.)

Figure 7.1: Structure of Staphylococcal nuclease determined by X-ray at 1.7Å resolution (Hynes & Fox, 1991).
A. sodium dodecyl sulfate (SDS)  
\[
\text{CH}_3(\text{CH}_2)_{11}\text{SO}_3^- \text{Na}^+
\]

B. 1-heptanesulfonic acid, sodium salt  
\[
\text{CH}_3(\text{CH}_2)_{6}\text{SO}_3^- \text{Na}^+
\]

C. deoxycholic acid, sodium salt  

D. dodecyltrimethylammonium bromide  

E. 4 Laurly ether (Brij 30)  
\[\text{C}_{12}\text{H}_{25}(\text{OCH}_3\text{CH}_2)_{4}\text{OH}\]

F. 23 Laurly ether (Brij 35)  
\[\text{C}_{12}\text{H}_{25}(\text{OCH}_3\text{CH}_2)_{23}\text{OH}\]

Figure 7.2: Structures for different kinds of detergents: anionic detergents (A, B, C), cationic detergent (D), and nonionic detergents (E, F).
Figure 7.3: Structures for osmolyte solvents used in this study.
A. 2,2,2-trifluoroethanol (TFE)

B. hexafluoroisopropanol (HFIP)

C. ethanol

Figure 7.4: Structures for 2,2,2-trifluoroethanol (TFE), hexafluoroisopropanol (HFIP) and ethanol.
A MAPLE PROGRAM FOR SIMPLE UNIFORM HELIX ANALYSIS

This appendix includes a Maple program based on the simple uniform helix analysis (See Chapter 3, section 3.4.2.1). Mathematically, \( \frac{\partial \ln Z}{\partial \ln w_i} \) can be expressed as:

\[
\frac{\partial \ln Z}{\partial \ln w_i} = \frac{1}{Z} \frac{1}{1} \left( \prod_{j=1}^{i-1} \begin{pmatrix} w_j & v_j & 0 \\ 0 & 0 & 1 \\ v_j & v_j & 1 \end{pmatrix} \right) \frac{1}{1} \left( \prod_{j=i+1}^{N} \begin{pmatrix} w_j & v_j & 0 \\ 0 & 0 & 1 \\ v_j & v_j & 1 \end{pmatrix} \right) \frac{1}{1},
\]

thus in this program, \( J \) refers to the original 3x3 statistical weight matrix:

\[
\begin{pmatrix} w_j & v_j & 0 \\ 0 & 0 & 1 \\ v_j & v_j & 1 \end{pmatrix},
\]

and \( Ju \) refers to the matrix:

\[
\begin{pmatrix} w_i & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\]

\( Lend \) and \( Rend \) refer to the right and left end matrix, respectively. Part is the partition function \( Z \), and \( Totw \) is therefore equal to the \( \sum_{i=1}^{N} \left( \frac{\partial \ln Z}{\partial \ln <w>} \right) \) term in equation (3.13).

\[
> \text{with(linalg);} \\
> \# \text{ keep } v \text{ as a constant} \\
> v := 0.048; \\
> J := \text{array}([[w,v,0],[0,0,1],[v,v,1]]); \\
> Ju := \text{array}([[w,0],[0,0],[0,0,0]]); \\
> Lend := \text{array}([[0,0,1]]); \\
> \]

154
> Rend:=array([[0],[1],[1]]);

APPENDIX B

A MAPLE PROGRAM FOR MODIFIED UNIFORM HELIX ANALYSIS

This program is similar to the one in Appendix A, but instead of using the same w term for all residues, the modified residue is given a \( w' \) (see Chapter 3, section 3.4.2.2). Thus, \( X \) and \( X_u \) refer to the matrices for the modified residues, and \( x \) is the \( w' \) value for the modified residues, while \( w \) is the \( <w> \) control. The example given here can be applied to either 24p1m06 or 24p2m06 peptide.

```maple
> with(linalg);
> v:=.048;
> J:=array([[w,v,0],[0,0,1],[v,v,1]]);
> J:=array([[w,0,0],[0,0,0],[0,0,0]]);
> X:=array([[x,v,0],[0,0,1],[v,v,1]]);
> X:=array([[x,0,0],[0,0,0],[0,0,0]]);
> Lend:=array([[0,0,1]]);
> Rend:=array([[0],[1],[1]]);

```

156
>
>Totw:=evalm((Pos1a+Pos2a+Pos3a+Pos4a+Pos5a+Pos6a+Pos7a+Pos8a+Pos9a
>+Pos10a+Pos11a+Pos12a+Pos13a+Pos14a+Pos15a+Pos16a+Pos17a+Pos18a
>+Pos19a+Pos20a+Pos21a+Pos22a+Pos23a+Pos24a+Pos25a+Pos26a)/Part);
>
> # Helix content = <nh>/Nh =<nh>/(N-4), Nh=N-4=26-4=22
> fH:=det(Totw)/22;
>
> # Ellipticity=fH*Ehelix; where Ehelix=40000*(1-4.32/25)
> Ecal:=fH*33088;
>
> # given w the value for control peptide, i.e. the <w>control
> w:= <w>control;
>
> # let Ecal=Eexp to solve x , i.e. w' for modified residue
> fsolve(Ecal=Eexp,x);
APPENDIX C

A FORTRAN PROGRAM FOR FULL HELIX ANALYSIS

This program is based on Rohl et al.'s modified Lifson-Roig analysis (Rohl et al., 1996), and uses the enumeration method to do the calculation (see Chapter 3, section 3.4.2.3).

The input file "sample.seq" should include the total number of residues as well as the sequence of the peptide. For example, to calculate the ellipticity for 24p2con, the file content will be: 26,X,W,G,G,R,A,A,A,A,R,A,A,A,A,R,A,A,A,A,R,Z

Some important variables are defined as:

nres : number of residue, i.e. N

total : partition function, i.e. Z

pcw : percentage of being w, i.e. \( \sum_{i=1}^{N} \frac{\partial \ln Z}{\partial \ln w_i} \)

fH : fraction hydrogen bond or helix content, i.e. \( \frac{pcw}{N-4} \) (equation (3.23))

ellip2 : ellipticity = \( fH \times \left( -42500 \times \left( 1 - \frac{3.0}{(N-2)} \right) - 640 \right) + 640 \) (equation (3.22))
A Program based on Rohl et al.'s "helix2.for" 

Program helix2

dimension vtot(50),wtot(50),isb(2,10)
real fH,ellip2
integer sb
common vtot,wtot,total
common /bridge/isb,vsb
open(unit=1,file='output.txt',status='new')
open(unit=3,file='sample.seq',status='old')

10 do i=1,50
    vtot(i)=0
    wtot(i)=0
10 continue

c call load (subroutine) to load all sequence and assign parameters to each residue

call load(nres)

c call salt (subroutine) to add salt bridge interaction if needed

call salt(sb)

    total=0
    istart=0
    istop=(2**(nres-1))-1

c call evalnew to calculate partition funcion

c as ip is incremented through its range, the bit pattern will generate all possible states

100 do ip=istart,istop,2
    call evalnew(ip,nres,sb)
100 continue

    pcw=0
    pcv=0
    do i=1,nres
        pcw=wtot(i)/total+pcw
        pcv=vtot(i)/total+pcv
    200 continue

    fH=pcw/float(nres-4)
write(1,*)'fH=',fH,pcw/float(nres-4)
    ellip2=fH*(-42500*(1-3.0/float(nres-2))-640)+640
write(1,*)'ellipticity=(Rohl)',ellip2
Subroutine for loading sequence and parameters

Subroutine load(nres)

dimension v(50),w(50),c(50)
real n(50)
character ichar,jchar(30)
common/peptide/v,w,c,n
common vtot,wtot,total

read(3,*) nres,(jchar(j),j=1,nres)

do 20 i=1,nres
ichar=jchar(i)
v(i)=sqrt(0.0013)
if (ichar.eq.'A') then
w(i)=1.70
n(i)=1.00
c(i)=1.00
else if (ichar.eq.'B') then
w(i)=0.70
n(i)=1.00
c(i)=1.00
else if (ichar.eq.'C') then
w(i)=0.33
n(i)=5.4
nc(i)=1.00
else if (ichar.eq.'D') then
w(i)=0.38
n(i)=6.6
nc(i)=1.00
else if (ichar.eq.'E') then
w(i)=0.54
n(i)=2.06
nc(i)=1.00
else if (ichar.eq.'F') then
w(i)=0.27
n(i)=2.06
nc(i)=1.00
else if (ichar.eq.'G') then
w(i)=0.048
\begin{verbatim}
n(i)=3.9
c(i)=1.0
else if (ichar.eq.'H') then
  w(i)=0.22
  n(i)=1.00
  c(i)=1.00
else if (ichar.eq.'T') then
  w(i)=0.46
  n(i)=1.57
  c(i)=1.00
else if (ichar.eq.'T') then
  w(i)=0.36
  n(i)=2.12
  c(i)=1.00
else if (ichar.eq.'K') then
  w(i)=1.00
  n(i)=0.72
  c(i)=1.00
else if (ichar.eq.'L') then
  w(i)=0.87
  n(i)=2.06
  c(i)=1.00
else if (ichar.eq.'M') then
  w(i)=0.65
  n(i)=1.31
  c(i)=1.00
else if (ichar.eq.'N') then
  w(i)=0.29
  n(i)=6.80
  c(i)=1.00
else if (ichar.eq.'O') then
  w(i)=0.40
  n(i)=1.00
  c(i)=1.00
else if (ichar.eq.'P') then
  w(i)=0.001
  n(i)=1.35
  c(i)=1.00
else if (ichar.eq.'Q') then
  w(i)=0.62
  n(i)=0.12
  c(i)=1.00
else if (ichar.eq.'R') then
  w(i)=1.14
  n(i)=1.00
  c(i)=1.00
else if (ichar.eq.'S') then
\end{verbatim}
w(i)=0.40
n(i)=3.90
c(i)=1.00
else if (ichar.eq.'T') then
  w(i)=0.18
  n(i)=2.23
  c(i)=1.00
else if (ichar.eq.'U') then
  w(i)=0.00
  n(i)=0.00
  c(i)=1.00
else if (ichar.eq.'V') then
  w(i)=0.25
  n(i)=0.96
  c(i)=1.00
else if (ichar.eq.'W') then
  w(i)=0.29
  n(i)=3.60
  c(i)=1.00
else if (ichar.eq.'X') then
  w(i)=0.00
  n(i)=5.90
  c(i)=0.00
else if (ichar.eq.'Y') then
  w(i)=0.48
  n(i)=4.90
  c(i)=1.00
else if (ichar.eq.'Z') then
  w(i)=0.00
  n(i)=0.00
  c(i)=1.00
endif
20 continue
return
cend

**************** subroutine salt **********************

Subroutine salt(sb)
  integer sb
  dimension isb(2,10)
  common /bridge/isb,vsb

  if (sb.eq.0) goto 30 ! no salt bridge
  C Max number of salt bridges is 10
  write(6,5)
}

162
5     format('number of salt bridges?')
read(5,*), sb
if (sb.gt.0) then
   write(6,20)
20    format('Strength of salt bridge?')
read(5,*), vsb
   do 25 i=1,sb
      write(6,10)
10   format('Salt bridge from [residue #]?')
read(5,*), isb(1,i)
      write(6,15)
15   format('to? [residue #]')
read(5,*), isb(2,i)
25   continue
   endif
30   return
end

*************** subroutine evalnew to evaluate helicity **************

Subroutine evalnew(ipep,nres,sb)
dimension ires(50),isb(2,10),wtot(50)
dimension v(50),w(50),c(50)
real n(50)
integer sb
common vtot,wtot,total
common/peptide/v,w,c,n
common /bridge/isb,vsb
ip=ipep
prob=1.0
probw=1.0
c ipep is a state of the peptide. We will assume that a
c bit set to 1 is helical and a bit set to 0 is non helical
c what is the probability of this state?
do 11 i=1,50
   ires(i)=0
11   continue
c C terminus must be c-->1, the (N-1) residue could be 1 (if c) or v (if h)
imask=2
ival=iand(ip,imask)
if(ival.eq.2) then ! it is a h state, ie. v
   prob=prob*v(nres-1)
   probw=prob*1
ires(nres-1)=1
endif

C now do the bulk of the peptide
imask=31
do 10 i=2,nres-3
  ival=iand(ip,imask)
  if (ival.eq.31) then
    ires(nres-i)=2
    probw=prob*w(nres-i)
    prob=prob*w(nres-i)
  else if (ival.eq.14) then
    ires(nres-i)=2
    probw=prob*w(nres-i)
    prob=prob*w(nres-i)*n(nres-i-2)*c(nres-i+2)
  else if (ival.eq.15) then
    ires(nres-i)=2
    probw=prob*w(nres-i)
    prob=prob*w(nres-i)*n(nres-i-2)
  else if (ival.eq.30) then
    ires(nres-i)=2
    probw=prob*w(nres-i)
    prob=prob*w(nres-i)*c(nres-i+2)
  else if (ival.eq.4) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
  else if (ival.eq.5) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
  else if (ival.eq.6) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
  else if (ival.eq.7) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
  else if (ival.eq.12) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
  else if (ival.eq.13) then
    prob=prob*v(nres-i)
    probw=prob*w(nres-i)
    ires(nres-i)=1
else if (ival.eq.20) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
else if (ival.eq.21) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
else if (ival.eq.22) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
else if (ival.eq.23) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
else if (ival.eq.28) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
else if (ival.eq.29) then
    prob=prob*v(nres-i) \ v
    probw=prob*1
    ires(nres-i)=1
endif

ip=ishft(ip,-1)
10 continue

C Determine prob of N-1 terminal residue.
C N terminal ie, residue 1 must be C ->1. Residue 2, however, could be 1 (if c) or v (if h)
imask=4
ival=iand(imask,ip)
if (ival.eq.imask) then ! res 2 is h state, ie. v
    ires(2)=1
    prob=prob*v(2)
    probw=prob*1
endif

C check salt bridge
if(sb.gt.0) then
    do 30 ii=1,sb

        imask=1
        imask=ishft(imask,nres-1)
        imask=ishft(imask,1-isb(1,ii))
        do 35 jj=1,2
            ivl=iand(jpep,imask)
            if(ivl.ne.imask) go to 30 ! not a helix bail out
            imask=ishft(imask,-1)

    30 continue

    do 35 jj=1,2
        ivl=iand(jpep,imask)
        if(ivl.ne.imask) go to 30 ! not a helix bail out
        imask=ishft(imask,-1)

35 continue

165
35      continue
C if you get to here all residues between the salt bridges are helical
C multiply in the helix factor
    prob=prob*vsb
30      continue
    endif

    call sumall(ires,nres,prob,probw)

    return
    end

***** Subroutine sumall to get w total and v total *****

Subroutine sumall(ires,nres,prob,probw)
dimension ires(nres),vtot(50),wtot(50)
common vtot,wtot,total

    total=total+prob

    do 20 i=1,nres
       if(ires(i).eq.1) vtot(i)=vtot(i)+prob
       if(ires(i).eq.2) wtot(i)=wtot(i)+probw
20      continue
    return
    end
APPENDIX D

A FORTRAN PROGRAM FOR DICHROIC MODEL

This program based on the Dichroic model (Shalongo & Stellwagen, 1997), but using Rohl et al.'s definition for n and c capping parameters (Rohl et al., 1996) (see Chapter 3, section 3.4.3).

As described in Appendix C, the input file "sample.seq" should include the total number of residue as well as the sequence of the peptide. Notice also the intrinsic parameters (w, v, n and c) were obtained using optimization method (see Chapter 3, section 3.5), and only available for few amino acids (Table 3.3).

Some important variables are defined as:

nres : number of residue, i.e. N

modpep : hydrogen bonding ability for peptide bond
          : for normal peptide bond, modpep=0.0
          : if the peptide without NH, modpep=1.0
          : if the peptide without CO, modpep=-1.0

cdstate : mean amide ellipticity for a specific state, i.e. [θ]k in equation (3.27)

cd : calculated mean amide bond ellipticity (equation (3.27))
*** Dichroic model using new definition of n and c capping *****

program dichroic

C Note: for true L-R this should only be applied to residues
C 2 to n-1;

dimension isb(2,10)
real cd
integer sb
common total
common/bridge/isb,v sb
common/peptide/v,w,c,n, hfac, modpep
open(unit=1, file='output.txt', status='new')
open(unit=3, file='sample.seq', status='old')

cd=0.0

c call load (subroutine) to load all sequence and assign parameters to each residue
C parameters from our own calculation

call load(nres)

c call salt (subroutine) to add salt bridge interaction if needed

call salt(sb)

total=0
istart=0
ctot=0
istop=(2**(nres-1))-1
!
as ip is incremented through its range, the bit pattern will generate all possible states

do 100 ip=istart,istop,2
   cdpep=0
   call evalnew2(ip,nres,sb,cdpep)
   cdtot=cdtot+cdpep
100 continue

   cd=cdtot/total
write(1,210) cd
210 format(' CD based on dichroic:',f8.0)
   close(3)
   close(1)
end
Subroutine for loading sequence and parameters

c parameters from our best fit calculation
c residue without NH, modpep=1.0, residue without CO, modpep=-1.0

Subroutine load(nres)

dimension v(50), w(50), c(50)
real n(50), hfac, vp
integer modpep(50)
character ichar,jchar(30)
common total
common/bridge/isb, vsb
common/peptide/v, w, c, n, hfac, modpep

read(3,*) nres, (jchar(j), j=1, nres)

do 20 i=1, nres
  modpep(i)=0.0
  ichar=jchar(i)
  if (ichar.eq.'A') then
    w(i)=1.73
    n(i)=1.00
    c(i)=1.00
    v(i)=0.036
    modpep(i)=0.0
  else if (ichar.eq.'B') then
    w(i)=?
    n(i)=?
    c(i)=?
    v(i)=0.0036
    modpep(i)=0.0
  else if (ichar.eq.'C') then
    w(i)=?
    n(i)=?
    c(i)=?
    v(i)=0.036
    modpep(i)=0.0
  else if (ichar.eq.'D') then
    w(i)=?
    n(i)=?
    c(i)=?
    v(i)=0.036
    modpep(i)=0.0
  else if (ichar.eq.'E') then
    w(i)=0.63
    n(i)=2.95
    c(i)=0.65

169
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'F') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'G') then
\[ w(i) = 0.06 \]
\[ n(i) = 2.87 \]
\[ c(i) = 0.97 \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'H') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'T') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'T') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'K') then
\[ w(i) = 1.02 \]
\[ n(i) = 0.86 \]
\[ c(i) = 0.75 \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'L') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
\[ v(i) = 0.036 \]
\[ \text{modpep}(i) = 0.0 \]
else if (ichar.eq.'M') then
\[ w(i) = ? \]
\[ n(i) = ? \]
\[ c(i) = ? \]
v(i)=0.036
modalp(i)=0.0
else if (ichar.eq.'N') then
  w(i)=?
  n(i)=?
  c(i)=?
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'O') then
  w(i)=?
  n(i)=?
  c(i)=?
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'P') then
  w(i)=?
  n(i)=1.69
  c(i)=0.00
  v(i)=0.036
  modalp(i)=1.0
else if (ichar.eq.'Q') then
  w(i)=0.57
  n(i)=0.08
  c(i)=1.83
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'R') then
  w(i)=1.59
  n(i)=1.19
  c(i)=2.22
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'S') then
  w(i)=?
  n(i)=?
  c(i)=?
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'T') then
  w(i)=?
  n(i)=?
  c(i)=?
  v(i)=0.036
  modalp(i)=0.0
else if (ichar.eq.'U') then
  w(i)=?
  n(i)=?
c(i)=?
  v(i)=0.036
  modpep(i)=0.0
else if (ichar.eq.'V') then
  w(i)=?
  n(i)=?
  c(i)=?
  v(i)=0.036
  modpep(i)=0.0
else if (ichar.eq.'W') then
  w(i)=0.15
  n(i)=5.67
  c(i)=0.00
  v(i)=0.036
  modpep(i)=0.0
else if (ichar.eq.'X') then
  w(i)=0.0
  n(i)=6.95
  c(i)=0.0
  v(i)=0.036
  modpep(i)=0.0
else if (ichar.eq.'Y') then
  w(i)=0.25
  n(i)=3.26
  c(i)=0.97
  v(i)=0.036
  modpep(i)=0.0
else if (ichar.eq.'Z') then
  w(i)=0.0
  n(i)=0.0
  c(i)=2.74
  v(i)=0.036
  modpep(i)=0.0
endif
20 continue

**c now factor out a h-bond factor and vp value**
c
   hfac=1.00
c   vp=1.00
c   do 60 i=1,nres
c     v(i)=w(i)*vp
   c     w(i)=w(i)*vp*hfac
   c     write(*,*)'out put w,v for res i',w(i),v(i),i
   c60 continue
   return
end
*************** subroutine salt ***************

Subroutine salt(sb)
integer sb
dimension isb(2,10)
common/bridge/isb,vsb

C Max number of salt bridges is 10
write(6,5)
5 format('number of salt bridges?')
read(5,*) sb
if (sb.gt.0) then
write(6,20)
20 format(' Strength of salt bridge?')
read(5,*) vsb
do 25 i=1,sb
write(6,10)
10 format(' Salt bridge from [residue #]?')
read(5,*) isb(1,i)
write(6,15)
15 format(' to? [residue #]?')
read(5,*) isb(2,i)
25 continue
endif
return
end

*************** subroutine salt1, for peptide1 **************

Subroutine salt1(sb)
C designed for peptide 1
integer sb
dimension isb(2,10)
common/bridge/isb,vsb

C Max number of salt bridges is 10
sb=4
if (sb.gt.0) then
C write(6,20)
C20 format(' Strength of salt bridge?')
C read(5,*) vsb
vsb=2.96
isb(1,1)=5
isb(2,1)=9
isb(1,2)=10
isb(2,2)=14
isb(1,3)=15

173
**Subroutine evalnew2 based on new definition of n and c capping**

Subroutine evalnew2(ipep,nres,sb,cdpep)
dimension ires(50),isb(2,10)
dimension v(50),w(50),c(50)
real n(50)
integer modpep(50),hbond(50),sb
common total
common/bridge/isb,vsb
common/peptide/v,w,c,n,hfac,modpep

ip=ipep
prob=1.0
probw=1.0

d o 11 i=1,nres
    ires(i)=0
11 continue

C Determine prob of C-1 terminal residue.
C C terminal ie, residue N must be C — >1, residue N-1, however, could be l (if c) or v (if h)

imask=2
ival=iand(ip,imask)
if(ival.eq.2) then ! it is a h, so given v
    prob=prob*v(nres-1)
    probw=prob*1
    ires(nres-1)=1
endif

C now do the bulk of the peptide
imask=31
do 10 i=2,nres-3
    ival=iand(ip,imask)
if(ival.eq.31) then
    ires(nres-i)=2 ! w
    probw=prob*w(nres-i)
    prob=prob*w(nres-i)
else if (ival.eq.14) then
\begin{verbatim}
ires(nres-i)=2  !wnc
probw=prob*\textit{w(nres-i)}
prob=prob*\textit{w(nres-i)}*\textit{n(nres-i-2)}*c(nres-i+2)
else if (ival.eq.15) then
    ires(nres-i)=2  !wn
    probw=prob*\textit{w(nres-i)}
    prob=prob*\textit{w(nres-i)}*\textit{n(nres-i-2)}
else if (ival.eq.30) then
    ires(nres-i)=2  !wc
    probw=prob*\textit{w(nres-i)}
    prob=prob*\textit{w(nres-i)}*c(nres-i+2)
else if (ival.eq.4) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.5) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.6) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.7) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.12) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.13) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.20) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.21) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.22) then
    prob=prob*\textit{v(nres-i)}  !v
    probw=prob*1
    ires(nres-i)=l
\end{verbatim}

175
else if (ival.eq.23) then
    prob=prob*v(nres-i) ! v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.28) then
    prob=prob*v(nres-i) ! v
    probw=prob*1
    ires(nres-i)=l
else if (ival.eq.29) then
    prob=prob*v(nres-i) ! v
    probw=prob*1
    ires(nres-i)=l
endif
ip=ishft(ip,-l)

C Determine prob of N-1 terminal residue.
C N terminal ie, residue 1 must be C -> 1. Residue 2, however, could be l (if c) or v (if h)
imask=4
ival=iand(imask,ip)
if (ival.eq.imask) then ! res 2 is h, so given v
    ires(2)=l
    prob=prob*v(2)
    probw=prob*1
endif

c check salt bridge
if(sb.gt.0) then
    c we have a salt bridge
    do 30 ii=1,sb
        imask=1
        imask=ishft(imask,nres-1)
        imask=ishft(imask,1-isp(1,ii))
        do 35 jj=isp(1,ii),isp(2,ii)
            ivl=iand(ipep,imask)
            if(ivl.ne.imask) go to 30 ! not a helix bail out
            imask=ishft(imask,-1)
    35 continue
C if you get to here all residues between the salt bridges are helical
C multiply in the helix factor
    prob=prob*vsb
30 continue
endif
cc The CD section Stellwagen

   do 110 i=1,nres
      hbond(i)=0
   110 continue
   ip=ipep
   do 120 i=nres-2,3,-1
      imask=14
      ival=iand(ip,imask)
      C here hbond if i will refer to the hbond between residue i and i+1
      if(ival.eq.imask) then ! three H in a row- an H bond can form
         if(modpep(i+2).ne.1) hbond(i+1)=hbond(i+1)+1
         if(modpep(i-2).ne.-1) hbond(i-2)=hbond(i-2)+1
      endif
      ip=ishft(ip,-1)
   120 continue

   n1hbond=0
   n2hbond=0
   do 130 i=1,nres
      if(hbond(i).eq.1) n1hbond=n1hbond+1
      if(hbond(i).eq.2) n2hbond=n2hbond+1
   130 continue
   npepbond=nres-1

   write(*,*) n1hbond,n2hbond,npepbond

   c assume complete coil=0, and complete helix=-40,000
   cdstate=((n1hbond*11200)+(n2hbond*40000))/npepbond

   c assume complete coil=+640, complete helix=-40,000
   c cdstate=((-(npepbond-n1hbond-n2hbond)*640)+(n1hbond*11200)
   c *(n2hbond*40000))/npepbond
   cdpep=cdstate*prob
   total=total+prob
   return
end
LIST OF REFERENCES


