
Donald Stephen Clark
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John’s Road, Tyler’s Green
High Wycombe, Bucks, England HP10 8HR
CLARK, Donald Stephan, 1949-
CONFORMATIONAL PROPERTIES OF
POLYPEPTIDES: HELIX-COIL
TRANSITION, AND THE CONFORMATIONAL
AND HYDRODYNAMIC PROPERTIES OF
POLY(GAMMA-HYDROXY-L-PROLINE).

The Louisiana State University and
Agricultural and Mechanical College,
Ph.D., 1976
Biophysics, general

Xerox University Microfilms, Ann Arbor, Michigan 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CONFORMATIONAL PROPERTIES OF POLYPEPTIDES:
HELIX-COIL TRANSITION, AND THE CONFORMATIONAL AND
HYDRODYNAMIC PROPERTIES OF POLY(GAMMA-HYDROXY-L-PROLINE)

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biochemistry

by

Donald Stephen Clark
B.S., Louisiana State University, 1972
December, 1976
ACKNOWLEDGEMENT

The firm guidance and unending support demonstrated by Professor W. L. Mattice in his direction of this doctoral research is most humbly appreciated.

Many contributions to the author's career from Dr. G. E. Risinger, Dr. S. H. Chang, Dr. J. A. Bowden, Dr. R. S. Allen, Dr. E. S. Younathan, and especially to D. K. Carpenter are gratefully acknowledged.

The author would also like to acknowledge K. Crossin, A. J. Moser, and C. Robertson for their efforts in helping in the preparation of this dissertation.

Finally, the financial support by the LSU Biochemistry Department and the National Science Foundation is gratefully acknowledged.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENT.</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER 1 General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Matrix Formulation</td>
<td>33</td>
</tr>
<tr>
<td>2.3 End Effects</td>
<td>35</td>
</tr>
<tr>
<td>2.4 Computations</td>
<td>37</td>
</tr>
<tr>
<td>2.5 Results</td>
<td>38</td>
</tr>
<tr>
<td>2.6 Discussion</td>
<td>51</td>
</tr>
<tr>
<td>CHAPTER 3</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>62</td>
</tr>
<tr>
<td>3.2 Experimental</td>
<td>70</td>
</tr>
<tr>
<td>3.3 Computations</td>
<td>74</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>81</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>115</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>135</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>147</td>
</tr>
<tr>
<td>VITA</td>
<td>176</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Structure parameters of three important polypeptide chain conformations</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Statistical weights for the helical and coil states in a polypeptide chain</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>The $\sigma$ and $s$ values of amino acids determined to date in water at $20^\circ C$</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>$\tau$ and $R$ values for the helix-coil transition curves for the Zimm-Bragg treatment, $\gamma(5)$ statistical weight matrix treatment, and $\gamma(10)$ statistical weight matrix treatment.</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>Coefficients of variation, and the percent deviation in $\sigma'$ and $s'$ for the $\gamma(4)$ modification 3 statistical weight matrix and $\sigma = 0.0005$</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>Characteristic constants of the nonbonded interactions</td>
<td>77</td>
</tr>
<tr>
<td>7</td>
<td>Molecular weights of the four poly (γ-hydroxy-L-proline) samples</td>
<td>87</td>
</tr>
<tr>
<td>8</td>
<td>Intrinsic viscosity data of poly (γ-hydroxy-L-proline) samples studied and of poly-L-proline</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>Sedimentation and second virial coefficients for poly (γ-hydroxy-L-proline) samples in water</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>Dimensional properties at 30$^\circ C$ for poly (γ-hydroxy-L-proline) with the pyrrolidine ring puckered by 17$^\circ$</td>
<td>107</td>
</tr>
<tr>
<td>11</td>
<td>Dimensional properties at 30$^\circ C$ for poly (γ-hydroxy-L-proline) with inclusion of a hydrogen bonding potential between the hydroxyl proton of residue $i + 1$ and the carbonyl oxygen of residue $i - 1$.</td>
<td>114</td>
</tr>
<tr>
<td>12</td>
<td>The diffusion coefficients, frictional coefficients, $\beta$, the expansion coefficients, and the characteristic ratios calculated from the experimental results for poly (γ-hydroxy-L-proline)</td>
<td>127b</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Geometrical representation of a portion of a polypeptide chain</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Hydrogen bonding pattern of the α-helix</td>
<td>6</td>
</tr>
<tr>
<td>3.</td>
<td>Experimental correlation between the Zimm-Bragg helix-coil transition parameters $\sigma$ and $s$</td>
<td>32b</td>
</tr>
<tr>
<td>4.</td>
<td>Calculated correlations between $\sigma'$ and $s'$ using the $U(3)$ through $U(10)$ statistical weight matrices to compute the fraction helical content for $\sigma = 0.0005$</td>
<td>39b</td>
</tr>
<tr>
<td>5.</td>
<td>Calculated correlations between $\sigma'$ and $s'$ using the $U(10)$ statistical weight matrix with $\sigma = 0.0001$, 0.001, 0.01, 0.1, and 1.0</td>
<td>42b</td>
</tr>
<tr>
<td>6.</td>
<td>Helix-coil transition curves for the Zimm-Bragg treatment, using the $U(5)$ statistical weight matrix and the $U(10)$ statistical weight matrix with $\sigma = 0.0005$</td>
<td>43b</td>
</tr>
<tr>
<td>7.</td>
<td>Calculated correlations between $\sigma'$ and $s'$ using the $U(5)$ modification 3, $U(5)$ modification 2, $U(5)$, and the $U(5)$ modification 1 statistical weight matrix and $\sigma = 0.0005$.</td>
<td>46b</td>
</tr>
<tr>
<td>8.</td>
<td>Calculated correlations between $\sigma'$ and $s'$ using the $U(4)$ modification 3, $U(4)$ modification 2, and the $U(4)$ statistical weight matrix with $\sigma = 0.0005$</td>
<td>48b</td>
</tr>
<tr>
<td>9.</td>
<td>Helix-coil transition curves for the Zimm-Bragg treatment, and using the $U(4)$ modification 2 statistical weight matrix with $\sigma = 0.005$ and 0.0005</td>
<td>49b</td>
</tr>
<tr>
<td>10.</td>
<td>Eigenvalues of the $U(4)$ statistical weight matrix plotted against $s$ with $\sigma^{1/3} = 0.1$, 0.01, and 0.0</td>
<td>59</td>
</tr>
<tr>
<td>11.</td>
<td>Internal dipeptide geometry of poly ($\gamma$-hydroxy-L-proline) used in the conformational energy calculations</td>
<td>75b</td>
</tr>
</tbody>
</table>
12. The poly (γ-hydroxy-L-proline) chain with all amide bonds in the trans configuration... 80

13. Concentration dependence in the reduced osmotic pressures for the poly (γ-hydroxy-L-proline) samples in water at 30°C ................ 82

14. The logarithm of the fringe number versus the square of the displacement from the center of rotation for the poly(γ-hydroxy-L-proline) sample HP15 ... ....... 84b

15. The logarithm of the fringe number versus the square of the displacement from the center of rotation for the poly (γ-hydroxy-L-proline) sample F2 ................... 85b

16. Concentration dependence of the apparent weight-average molecular weights for the poly (γ-hydroxyl-L-proline) samples measured by equilibrium ultracentrifugation in water at 25°C ....................... 86

17. Concentration dependence of the reduced viscosities for the poly (γ-hydroxy-L-proline) samples at 30°C in water and 1.95M calcium chloride ... ........ 88b

18. Concentration dependence of the reduced viscosities for the poly (γ-hydroxy-L-proline) samples at 30°C in 1.95M, 3.90M, and 5.57M calcium chloride .............. 89

19. Temperature dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) in water .......... 90

20. Molecular weight dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) and poly (L-proline) in water at 30°C ........... 92

21. Intrinsic viscosities of the poly (γ-hydroxy-L-proline) samples at 30°C as a function of calcium chloride concentration ........... 94
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. Intrinsic viscosities of the poly (γ-hydroxy-L-proline) samples at 30°C as a function of calcium chloride activity</td>
<td>95</td>
</tr>
<tr>
<td>23. Molecular weight dependence of the intrinsic viscosities of poly (γ-hydroxy-L-proline) at 30°C in water, 1.95M, 3.90M, and 5.57M calcium chloride</td>
<td>97</td>
</tr>
<tr>
<td>24. Variation of the exponent of the molecular weight dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) with calcium chloride concentration at 30°C</td>
<td>98</td>
</tr>
<tr>
<td>25. The logarithm of the displacement of the sedimenting boundary from the center of rotation versus time for the poly (γ-hydroxy-L-proline) samples F1 and HP15 at 20°C</td>
<td>99b</td>
</tr>
<tr>
<td>26. Concentration dependence of the sedimentation coefficients for the poly (γ-hydroxy-L-proline) samples at 20°C in water</td>
<td>101</td>
</tr>
<tr>
<td>27. Molecular weight dependence of the sedimentation coefficients for poly (γ-hydroxy-L-proline) at 20°C in water</td>
<td>103</td>
</tr>
<tr>
<td>28. Conformational energy maps of the internal dipeptide unit in poly (γ-hydroxy-L-proline) based on a pyrrolidine ring puckering angle of 17°</td>
<td>105</td>
</tr>
<tr>
<td>29. Possible hydrogen bond pattern in the poly-(γ-hydroxy-L-proline) chain</td>
<td>109</td>
</tr>
<tr>
<td>30. Conformational energy maps of the internal dipeptide unit in poly (γ-hydroxy-L-proline) based on pyrrolidine ring puckering angles of 8.5, 17, 25.5, and 45°</td>
<td>110</td>
</tr>
<tr>
<td>31. Characteristic ratios and temperature coefficients for poly (γ-hydroxy-L-proline) based on rigid pyrrolidine rings as a function of ring puckering angle</td>
<td>112</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>32. Characteristic ratio at 30°C for poly (γ-hydroxy-L-proline) based on a pyrrolidine ring puckering angle of 45° as a function of chain length</td>
<td>116</td>
</tr>
<tr>
<td>33. Molecular weight dependence of β for poly (γ-hydroxy-L-proline) at 20°C in water</td>
<td>124</td>
</tr>
<tr>
<td>34. Molecular weight dependence of the characteristic ratio of poly (γ-hydroxy-L-proline) in water at 30°C</td>
<td>127a</td>
</tr>
</tbody>
</table>
ABSTRACT

A correlation between the Zimm-Bragg helix-coil transition parameters \( \sigma \) and \( s \) determined for nine amino acids by the "guest-host" technique of von Dreek, Poland and Scheraga has been found. This correlation can be expressed empirically as

\[ \ln \sigma = 8.6 \ln s - 7.6 \]

which holds over the entire range of \( s \) found experimentally between 0.6 and 1.35. On the basis of matrix methods used to compute helix-coil transition curves, the correlation in part may be explained as arising from end effects caused by "fappiness" in the ends of helical sequences. Matrix methods were inadequate in reproducing the experimental correlation between \( \sigma \) and \( s \) in the range of \( s \) between 1.05 and 1.35. The possible importance of near-range interactions in explaining the correlation between \( \sigma \) and \( s \) are also discussed.

The intrinsic viscosities in water and calcium chloride and the sedimentation coefficients in water for poly (\( \gamma \)-hydroxy-L-proline) have been studied as a function of molecular weight. The molecular weight range covered was 9,000 - 35,000. High molecular weight poly (\( \gamma \)-hydroxy-L-proline) in water has the hydrodynamic behavior of a random coil immersed in a good solvent, as judged from the slopes of the plots \( \log \eta vs. \log M \), \( \log s_0 vs. \log M \), and the size of \( \beta = N_o s_o \eta_0^{1/3} M^{-2/3}(1-v_p)^{-1} \). Here \( N_o \), \( s_o \), \( \eta_0 \), \( M \), \( v \), and \( \rho \) are Avogadro's number, sedimentation coefficient, intrinsic viscosity, solvent viscosity, and...
molecular weight, partial specific volume, and solvent density, respectively. The characteristic ratio, \( \frac{\langle r^2 \rangle_\infty}{n_1} \), is found to be 15.9 ± 1.6, which is not significantly different from the results of 13.7 ± 0.9 obtained for poly (L-proline) by Mattice and Mandelkern under the same conditions. Here \( \langle r^2 \rangle_\infty \) is the unperturbed mean-square end-to-end distance of the polymer chain, \( n_p \) is the number of peptide bonds in the polymer, and \( l_p \) is the distance between adjacent \( \alpha \) carbon atoms. The limiting temperature coefficient for poly (\( \gamma \)-hydroxy-L-proline), obtained as the intercept of \( \frac{d \ln [\eta]}{dT} \) vs. \( 1/M \), is -0.005 deg\(^{-1}\), which is similar to that found for poly (L-proline) of -0.0008 deg\(^{-1}\). In this respect poly (L-proline) and poly (\( \gamma \)-hydroxy-L-proline) are reminiscent of various derivatives of cellulose.

The characteristic ratios, temperature coefficients, as well as the effect of calcium chloride on the hydrodynamic characteristics of poly (\( \gamma \)-hydroxy-L-proline) and poly (L-proline) suggest that both polypeptides possess a similar source of flexibility. Two possibilities for this flexibility are a second energy minimum about the \( \alpha \) carbon—carbonyl carbon bond or cis-trans isomerization about the peptide bond.

Conformational energy maps based on rigid pyrrolidine rings have been computed for the internal dipeptide unit in poly (\( \gamma \)-hydroxy-L-proline) containing planar trans peptide bonds. The conformational energy maps were computed as a function of ring puckering at the \( \gamma \) carbon atom and of rotation about the hydroxyl group. The conformational energy maps exhibit one low-energy region
at $\psi = 325^\circ \pm 40^\circ$ (using the convention in which $\phi, \psi = 0^\circ, 0^\circ$ for the fully extended chain). The calculations support the previous suggestions that a strong hydrogen bond between the hydroxyl proton and the carbonyl oxygen can occur if the pyrrolidine ring is sufficiently puckered at the Y carbon atom, $\chi_2 - \chi_3 > 25^\circ$, and $\psi$ is near $290^\circ$. The characteristic ratios computed from these maps are all higher than the result obtained experimentally for poly (Y-hydroxy-L-proline).
CHAPTER I  GENERAL INTRODUCTION

Proteins comprise the structural building blocks as well as most of the functional machinery of living organisms. They cannot claim the central position held by the nucleic acids as the carriers of heredity, yet certain proteins are responsible for the control of expression of hereditary information from its first transcription from the gene to its final translation into new polypeptide chains. Hundreds of proteins have been identified in the category of enzymes, catalyzing a myriad of complex biochemical reactions. Others are responsible for the basic structural framework of living organisms. In higher animals these structural proteins include collagen of bones, cartilage, and tendons; keratin of hair and nails; elastin of blood vessels and ligaments; and myosin of muscle.

It is now generally accepted that the major portion of the three-dimensional architecture of a protein must remain intact for biological activity. In ribonuclease, for example, amino acid residues from both ends of the polypeptide chain must be in close proximity for enzymatic activity to exist. Also most enzymes have been found to have a cleft or cavity below their surface. The dimensions and environment of this cavity must be rigidly controlled for the specific catalytic activity of the enzyme to remain. Other proteins like collagen and keratin require a specific helical conformation to give them their strength and rigidity demanded by their structural function. It should be reasonable to assume, therefore, that an investigation
into the conformational properties of proteins and model polypeptides should lead to a better understanding of the function of proteins.

Before an earnest attempt could be made in describing the conformations of large polypeptides found in nature, an extensive study on the crystal structure of amino acids and small polypeptides had to be conducted. Such an extensive study was carried out by Pauling and Corey and their collaborators and led to the determination of the fundamental dimensions of the peptide unit. Averaged results (1) based on additional more recent data are shown in Figure 1. The amide group itself, including the C-N bond and the substituent oxygen, amino hydrogen and α-carbon atoms, is found to be planar within experimental error. Normally the α-carbon atoms lie in the trans-conformation about the peptide linkage. This is the sterically favored conformation since it allows the greatest separation for consecutive α-carbon atoms. The distinction does not apply, however, in the case of proline or hydroxyproline. In fact, a polymer of proline has been found with cis-peptide bonds (poly-L-proline, Form I (2)).

In the structure shown in Figure 1, three rotational angles, φ, ψ and ω (3) are depicted. These angles, along with similar angles for the other peptide residues, determine the spatial configuration of the peptide chain. The zero position in φ,ψ and ω correspond to the polypeptide chain in its planar, trans, fully extended conformation as shown in the figure. A positive rotation for these angles is defined as clockwise when viewed down the chain from the amino to carboxyl terminus (3).
Figure 1. Geometrical representation of a portion of a polypeptide chain.
There are actually two conventions in common use in literature today. The convention used throughout this dissertation is the one set up by a subcommission of the IUPAC-IUB Commission on Biochemical Nomenclature in 1966 (3). In 1969 (4) the commission modified the proposals of 1966 to bring them into line with the system of nomenclature current in the fields of organic and polymer chemistry. The new designation of rotational angles may be derived from the old by subtracting 180° from the latter. Because of the large volume of literature published under the 1966 convention and because of the investigators greater familiarity with this convention, the older convention was preferred.

Normally, rotation about the peptide bond, denoted by ω, is of sufficiently high energy that it is usually not considered. Unless otherwise stated the value of ω will be zero, the trans-conformation for the peptide unit.

One of the easiest conformational structures to envision for a polypeptide is a helix. In a helix a set of φ, ψ rotational angles propagates down the polypeptide chain causing a regular spiraled conformation. This set of φ, ψ angles can be per peptide residue or per some integral number of peptide residues. Once this set of φ and ψ's are fixed, the number of residues per turn (n) and the average distance traversed down the helical axis per residue (d) are also defined. A quantity p called the pitch is given by the product of n and d. Every helix can be specified by these parameters.

The importance of various helical conformations for polypeptides was recognized as early as the 1940's (5,6). By 1951 the
first successful proposal of a polypeptide helix was presented by Pauling, Corey and Branson (7). They proposed for L-amino acids a right-handed helix which had 3.7 residues per turn. The term right-handed refers to the polypeptide backbone spiraling in a clockwise direction as viewed down the helical axis. Their helix also had a hydrogen bonding pattern as depicted by the dashed line in Figure 2. They called their structure the now well recognized \(\alpha\)-helix.

When the structure of myoglobin was later determined by Kendrew and his co-workers (8) using X-ray diffraction, a large fraction of the molecule was found to be in this \(\alpha\)-helical conformation. Since that time many polyamino acids have been found to be in the \(\alpha\)-helix in the solid state.

The helical conformation is also frequently retained in non-interacting solvents, such as dimethylformamide, chloroform, trifluoroethanol, and hexamethylphosphoamide. This has been shown by low angle X-ray scattering (9), dipole moment (10-12), UV (13), IR (14-16) and Raman spectroscopy (17-19), as well as hydrodynamic (20-22) and optical rotatory (23-25) studies. Both right-handed and left-handed \(\alpha\)-helices were found experimentally. The sense of the helix depends on the nature of the amino acid residues in the polypeptide chain. For example, homopolymers of L-alanine, L-lysine, L-tyrosine, \(\beta\)-propyl-L-aspartate, \(\gamma\)-benzyl-L-glutamate, \(\gamma\)-methyl-L-glutamate, \(\varepsilon\)-benzyloxycarbonyl-L-lysine, L-glutamic acid, and \(N^5(4\text{-hydroxybutyl})\)-L-glutamine in noninteracting solvents attain a right-handed helical conformation. Polymers of \(\beta\)-methyl,
Figure 2. Hydrogen bonding pattern of the α-helix.
β-benzyl, β-(o-chlorobenzyl)−, and β-(m-chlorobenzyl)−L-aspartate on the other hand form a left-handed helix (23-31).

While the α-helix is an important structural feature of polypeptide systems, it is by no means the only one. The enzyme lysozyme, for example, contains much less α-helix than does myoglobin, but it also contains a section of β-pleated sheet (32,33). The β-pleated sheet is an extended polypeptide chain with long range as opposed to near neighbor hydrogen bonds. That is, the hydrogen bonds are between residues in neighboring polymer chains or within the same peptide chain, but usually greater than three residues apart as is the case of the α-helix.

Polypeptides containing a heteroatom (oxygen or sulfur) in the γ position of the amino acid residue, such as poly-L-serine, poly-L-cysteine, and poly-L-threonine, tend to form a β-structure (26). The dipole-dipole interactions between the heteroatoms and the two adjacent peptide groups reduce the stability of the α-helical conformation but have little effect on the energy of the β-pleated sheet, thus favoring the latter (34). Silk (35-36) is the best example of this pleated sheet structure.

Other examples of helical structures in polypeptides are the α-keratin structure (37), the collagen triple helix (38-39), and the poly-glycine II (40), poly-L-proline (form II) (41-43) or poly (γ-hydroxy-L-proline) A (44) helix. Table 1 gives the ϕ, ϕ rotational angles, the number of residues per turn (n), the translation
per residue (d) and the pitch for a number of these helical structures.

Table 1. Structural parameters of three important polypeptide chain conformations (33).

<table>
<thead>
<tr>
<th></th>
<th>α-helix</th>
<th>β-pleated sheet</th>
<th>polyproline or poly hydroxyproline helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Φ</td>
<td>132°</td>
<td>40°</td>
<td>103°</td>
</tr>
<tr>
<td>Ψ</td>
<td>123°</td>
<td>315°</td>
<td>326°</td>
</tr>
<tr>
<td>n</td>
<td>3.61</td>
<td>2.00</td>
<td>-3.00*</td>
</tr>
<tr>
<td>d (Å)</td>
<td>1.50</td>
<td>3.47</td>
<td>3.12</td>
</tr>
<tr>
<td>p (Å)</td>
<td>5.41</td>
<td>6.95</td>
<td>9.36</td>
</tr>
</tbody>
</table>

* The negative sign denotes a left-handed helix.

The α-keratine structure is simply a modification of the α-helix which has superstructure. It consists of three slightly distorted α-helices twisted left-handedly about each other, not unlike that of a stranded piece of thread (33).

The collagen helix is another example of a twisted triple helix. In this structure the individual threads are left-handed helices. The three threads are twisted about each other in a slightly right-handed twist. Unlike the α-helix or the β-pleated sheet, the collagen triple helix requires that every third residue in the polypeptide chain be a glycine residue. Glycine occupies the interior positions of the triple helix superstructure and is the only amino acid which will allow the close packing requirements of the triple helix.
The left-handed helix of collagen is similar in structure to some synthetic polypeptides in the solid state. Polyglycine II, poly-L-proline (form II) and poly-L-hydroxyproline A all form similar left-handed helices. The amino acid composition of collagen is one-third glycine and approximately one-third proline and hydroxyproline (33). The left-handed helix has three residues per turn and has a translation of approximately 3A per residue. Because it has no side chain to influence the conformation, polyglycine also has the ability to form a right-handed helix in the crystalline state (40). The pyrrolidine rings of proline and hydroxyproline limit the rotation of $\phi$ to angles near 120°. Steric effects (45-56) caused by the hydrogen atoms attached to the $\delta$ methylene carbon atom and preceding pyrrolidine ring hydrogen atoms prohibit the formation of the $\alpha$-helical conformation, besides the fact that there are no peptide hydrogen atoms for hydrogen bonding. The restrictions cited above result in a fairly rigid helical structure not only for polyproline or polyhydroxyproline, but also for polypeptides containing these amino acids (57).

The structure of the helices just described have been worked out from solid-state studies. However, a helix is not the only conformation available to polypeptides in solution or in the solid-state. Most polypeptides and proteins can adopt a myriad of disordered conformations, called the random coil (58), under certain solvent conditions. Also, an ordered (native)-disordered transition can be
induced by changing solvent composition or temperature or both
(57, 59-61). Theoretically, methods for computing the statistical
conformations of macromolecules have been derived (58), and thermo-
dynamic parameters determining conformational transitions have been
deduced (62,63). These investigations aim at providing a better
understanding of structure and function of biological macromolecules.

With the advent of applying classical statistical mechanics to
chain molecules in the random-coiled conformation (58, 64), average
properties of polypeptide chains--average fraction helical content,
mean-square unperturbed end-to-end distance, mean-square unperturbed
radius of gyration, mean-square unperturbed dipole moment and mean-
square optical anisotropy--can be calculated. In any statistical
mechanical treatment a quantity known as the conformational partition
function, Q, must be calculated. Theoretically for polymer chains
with N residues, this quantity is given by

\[ Q(N) = \sum_{i=1}^{N} \exp \left( \frac{-\varepsilon_i}{kT} \right) dq \]  

(1)

All possible polymer conformations
\( \varepsilon_i \) is the free energy of a given residue \( i \) when the polymer is in
a specific conformation. For a polypeptide a specific conformation
corresponds to a specific set of \( \phi, \psi \) rotational angles. The integral
is over all possible polymer conformations. If we assume that the
polymer conformations are discrete, we can replace the integral by a
summation.

In order to get a feel for the types of parameters which go into
the calculation of the partition function, consider a homopolypeptide with discrete polymer conformations and under unperturbed conditions. Unperturbed conditions are those in which the polymer is free from external constraints, such as those caused by long-range interactions or by external forces (58). The conformational partition function for this homopolymer can be written as

\[ Q(N) = \sum_{\phi} \sum_{\psi} \prod_{i=1}^{N} \exp \left( -E(\phi_i, \psi_i)/RT \right) \]

all possible angles

Here \( E(\phi_i, \psi_i) \) becomes the conformational energy of residue \( i \) which includes such energy potentials as van der Waals non-bonded interactions, tortional potentials about \( \phi_i \) and \( \psi_i \), coulombic interactions and hydrogen bonding potentials (58, 65, 66). The summations are over all possible \( \phi, \psi \) rotational angles. If no unusual near-range interactions occur, then the energy contour, energy plotted against \( \phi \) and \( \psi \) for residue \( i \) and any other residue in the chain will be similar and is easily calculated. By the term no near-range interactions is meant that no unfavorable atomic contacts or strong electrostatic interactions exist between residue \( i \) and any other residue in the polypeptide chain except its nearest neighbors.

Energy contours of this type are called dipeptide conformational energy maps (58, 65, 66). These conformational energy maps are usually calculated using approximately 10° increments in \( \phi \) and \( \psi \). Such dipeptide conformational energy maps have been computed for glycine (67-70), D,L-alanine and other \( \alpha \)-amino acid residues in which \( R \) is a longer,
unbranched side chain of the kind $-\text{CH}_2\text{R (67-71)}$; $D,L$-valine (72); $D,L$-phenylalanine (72); $L$-serine (69, 73, 74); $L$-proline (48, 53, 55) and $\gamma$-hydroxy-$L$-proline (56).

In order to test the validity of a conformational energy map, average dimensions, such as those cited above, for the unperturbed random coil can be calculated and compared with experimentally measurable quantities. The most frequently used such experimental quantity is the characteristic ratio (58), defined (75) for polypeptides as $<r^2>_o/n^1_p^2$, where $<r^2>_o$ is the unperturbed mean-square end-to-end distance for a polypeptide containing $n_p$ virtual bonds of length $l_p$.

The calculation of experimental characteristic ratios can be carried out by using Equations 3-5, a procedure first utilized by Brant and Flory (75-77). The characteristic ratio is obtained from measured values of the intrinsic viscosity, molecular weights, and the second virial coefficient. $[\eta]$ is the intrinsic viscosity in a $\theta$ solvent, a solvent in which the expansion coefficient is one, $M_o$ is the molecular weight of an amino acid residue, $M_v$ is the viscosity average molecular weight, $\alpha$ is the expansion coefficient, $N_o$ is Avogadro's number, $A_2$ is the second virial coefficient and $\phi$ is a universal constant usually equal to 0.0021 with $[\eta]$ in deciliters/gram and distance in angstroms.

$$<r^2>_o/n^1_p^2 = ([\eta]_\theta/\alpha M_v^{1/2})^{2/3} M_o l_p^{-2}$$

$$[\eta]_\theta = [\eta]/\alpha^3$$

$$A_2 M/[\eta] = 2^{5/2} \pi N_o (27 \phi)^{-1} \ln [1 + 0.5 \pi^{1/2} (\alpha - 1)]$$
Direct evaluation of the experimental characteristic ratio is rather difficult. Usually poor solvents are used to produce the unperturbed state of synthetic polymers in solution (78). However, in such solvents polypeptides tend to assume ordered conformations or precipitate. Therefore, good solvents are employed instead, and the dimensions derived from experiment are corrected to the unperturbed state using the second virial coefficient and the expansion coefficient, \( \alpha \) (67,75-77,79-82).

The experimental values of characteristic ratios for homopolymers consisting of amino acids with different side chains, such as poly-L-glutamic acid and poly-L-lysine, are found similar to the corresponding value calculated for poly-L-alanine (67,83), namely close to 9. It appears that the nature of the side chain of an amino acid residue does not markedly affect the mean-square end-to-end distance of an unperturbed poly-L-amino acid chain provided the side chain has the structure \(-\text{CH}_2-\text{R}\). However, the characteristic ratios of copolymers, such as copoly(L-Ala-D-Ala), are considerably lower than that of poly-L-alanine (83,84).

As indicated above, the conformation of polyamino acids in solution can be changed from one ordered structure to another or from an ordered to an unordered conformation (58,59,61). This conformational transition can be brought about by altering temperature or solvent composition or both. Investigations of the conformational transitions of macromolecules has been extensively done experimentally and theoretically (58,59,61).
One of the most widely studied conformational transitions in polypeptides is the helix-coil transition (61). This transition was found to be highly cooperative (25, 85) and lead to the development of numerous theories to describe the transition (61). The first of such theories was proposed by Schellman in 1955 (86). This model suggested that the polypeptide molecule in solution exists as either an α-helix or a completely random coil. Later theories took into consideration the possible presence of both helical and coiled segments within a single polypeptide chain (61-63, 87-90). All of these theories, though with different mathematical treatment, use similar model systems.

The theoretical prediction that both coiled and helical segments coexist in a single polypeptide chain has been experimentally verified. X-ray studies of the helix-coil transition of poly-3,5-dibromotyrosine in dimethylformamide-trifluoroacetic acid mixture revealed that, at various stages of the transition, the polypeptide chains contained α-helical segments separated by coil regions (91). The helix-coil transition of other polypeptides has been followed by measuring changes in optical rotatory properties, viscosity, sedimentation velocity, and light scattering (92-95).

One of the simplest, most widely used and now standard theories of the helix-coil transition is the Zimm-Brigg theory (62). In this theory only two parameters (σ and s) are required to fit experimental transition curves within the accuracy of measurements (62).
σ is a statistical parameter which has been interpreted as being an entropy term involved with initiation of helical sequences in the polypeptide chain. The statistical parameter s has been associated with the free energy change per residue of the helix-coil transition.

In an attempt to have a quantitative measure of the relative stabilities of the α-helix and random-coil conformations of each naturally occurring amino acid in water, Scheraga and collaborators have measured σ and s for several polyamino acids in water (92,96-102) using a "guest-host" modification (103) of the Zimm-Bragg theory. The following three requirements must be met to make determination of σ and s experimentally feasible. First, the polyamino acid must be water soluble. Second, it must be able to adopt an α-helical conformation in water. Third, it must be able to undergo a helix-coil transition in the temperature range between 0 and 100°C. None of the homopolymers of the naturally occurring amino acids satisfies all three of these requirements. The "guest-host" modification involves the study of random copolymers in which the desired amino acid, the "guest", is incorporated at random into a nonionic homopolyamino acid, the "host", which does meet all three requirements. Out of a study of the experimentally determined values of σ and s to date, a correlation between σ and s seems prevalent.

This dissertation is divided into two sections. In section one, Chapter 2, the modified procedure for obtaining experimental values of σ and s for naturally occurring amino acids is outlined. The correlation
between $\sigma$ and $s$ is presented. And an attempt is made to explain this correlation by examining the role played by those amino acid residues which initiate helical sequences. In section two, Chapter 3, the similarities and differences in physical properties as determined to date between $\gamma$-hydroxy-L-proline and L-proline are discussed. Hydrodynamic properties of poly($\gamma$-hydroxy-L-proline) are determined and compared with similar properties of poly-L-proline. Conformational energies are computed for the internal dipeptide unit of poly($\gamma$-hydroxy-L-proline) based on a rigid pyrrolidine ring, and characteristic ratios are computed using these energy maps. These ratios are compared with characteristic ratios calculated from conformational energy maps of poly($\gamma$-hydroxy-L-proline) based on flexible pyrrolidine rings. Finally, these computed results are compared with the experimentally determined quantities of both poly($\gamma$-hydroxy-L-proline) and poly-L-proline.
CHAPTER 2 HELIX-COIL TRANSITION

2.1 Introduction

In the general model of the helix-coil transition of polypeptides, two states per residue have been assumed, a helical state, $h$, and a coiled state, $c$. The helical state is defined for an amino acid residue when the dihedral angles, $\phi$ and $\psi$, of this residue are in a narrow range about the $\phi, \psi$ angles found for the $\alpha$-helix (Table 1). A coiled state has been chosen in which the residue angles are any other angles than those for the helical state.

Equation 1 can be rewritten in terms of the free energy for each structure $G\{h,c\}$, where $\{h,c\}$ indicates a given sequence of $h$'s and $c$'s, as

$$Q(N) = \sum \exp \left[ -G\{h,c\} / RT \right],$$

where the summation is over all the possible allowed combinations of $h$'s and $c$'s.

In general it is extremely difficult to know very much about $G\{h,c\}$. But if one assumes that the free energies of successive sequences of $c$'s and $h$'s are independent of each other, $Q(N)$ can be computed with few other assumptions. Using this assumption, the free energy of a sequence of $h$'s or $c$'s depends only on its length and not on its position in the polypeptide chain or on the length of neighboring sequences. The statistical weight of a sequence of $h$'s or $c$'s can be written as

$$u_i = \exp \left[ -G_i(c)/RT \right], \text{ coiled sequence of length } i,$$

$$v_j = \exp \left[ -G_j(h)/RT \right], \text{ helical sequence of length } j.$$

(7a)  

(7b)
In terms of these statistical weights Equation 6 takes the form

\[ Q(N) = \sum_{i,j} u_i v_j \]  

(8)

In the theories of Lifson and Roig (63), and Zimm and Bragg (62) only nearest neighbor interactions were considered in their evaluation of \( Q(N) \). The total number of different statistical weights are reduced to four. In order to give their model more physical meaning, the assumption was also made that the minimum number of residues required for a helical sequence was \( j=3 \). This number was based on the fact that an intramolecular hydrogen bond can be formed only after three successive residues have been locked in to the \( \alpha \)-helix \( \phi, \psi \) (Figure 2). The consideration of the hydrogen bond lead to the assignment of the statistical weights as shown in Table 2.

Table 2. Statistical weights for the helical and coil states in a polypeptide chain (61).

<table>
<thead>
<tr>
<th>Residue State</th>
<th>Statistical Weight for residue ( i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>h or c h or c c</td>
<td>( u/u = 1 )</td>
</tr>
<tr>
<td>c h or c h</td>
<td>( v/v = \sigma^{1/2} )</td>
</tr>
<tr>
<td>h or c c h</td>
<td>( v/v = \sigma^{1/2} )</td>
</tr>
<tr>
<td>h h h</td>
<td>( w/w = s )</td>
</tr>
</tbody>
</table>

a. Lifson-Roig notation (63)
b. Zimm-Bragg notation (62)
c. Hydrogen bond possible (Figure 2)

The \( u_i \) and \( v_j \) of Equation 7 now becomes

\[ u_i = u^i = 1 \]  

(9a)

\[ v_j = v^j w^{j-2} = \sigma^{j-2} \quad \text{for } j>2 \]  

(9b)
In terms of the Zimm-Bragg notation the partition function $Q(N)$ becomes

$$Q(N) = 1 + \frac{(N-2)/2}{\sum_{\ell} \ell \sigma^2 \sum_{k=1}^{N-\ell-2} \frac{(k-1)! (N-k-2)! s^k}{\ell! (\ell-1)! (k-\ell)! (N-k-\ell-2)!}}$$

(10)

where $(N-2)/2$ is the largest integer less than $(N-2)/2$. The fraction of possible hydrogen bonds $\theta(N)$ formed (62) can be directly determined from $Q(N)$ by

$$\theta(N) = \frac{(1/N-3) (\partial \ln Q(N))}{(\partial \ln s)}$$

(11)

In a similar manner the average number of unbroken helical sections $V(N)$ (62), and the average number of amino acid residues per unbroken helical sections $L(N)$ (63) is given by

$$V(N) = \frac{(\partial \ln Q)}{(\partial \ln \sigma)}$$

$$L(N) = \frac{(N-3) \theta(N)}{V(N)}$$

(12)

(13)

When $s$ (the ratio $w/u$) is greater, less than, or equal to unity, helix is, respectively, favored, unfavored, and of equal probability, with respect to coil for infinitely large polypeptide chains. The value of $s$ for various naturally occurring amino acids in water varies from 0.6 to 1.4 (92,96-102,104). The ends of a helical sequence have a relatively low probability of occurrence, because the units at the ends of a helical sequence contribute neither a hydrogen bond nor entropy of internal rotation (coil state). Consequently, one may expect that $\sigma=(v/u)^2$ will be less than unity. This characteristic parameter of the helix-coil transitions reported in literature for various polyamino acids varies in the range: $1 \times 10^{-5}$ to $1 \times 10^{-2}$ (92,96-102, 104). For the transition of poly-L-alanine in water
\( \sigma = 1.44 \times 10^{-4} \) (104). The elongation of an existing helical sequence is thus favored over the initiation of a new helical sequence. This makes \( \alpha \)-helix to coil transition in polypeptides a cooperative phenomenon.

Several methods have been used in the past for evaluating complicated partition functions, e.g., the maximum term (105,106) and steepest descent (106,107) methods, but the method of greatest utility in chain statistics has been the matrix method (58, 108-110). In this method the partition function can be written as

\[
Q(N) = \mathcal{J}^* \mathcal{U}^N \mathcal{J}
\]  

(14)

where \( \mathcal{J} \) and \( \mathcal{J}^* \) are column and row matrices, respectively, which take into account the ends of the polymer chain. \( \mathcal{U} \) is a square matrix whose elements are the statistical weights for the various residue states in the polymer. In terms of the Zimm-Bragg notation \( \mathcal{U} \) is a 2 x 2 matrix with the following elements

\[
\mathcal{U} = \begin{bmatrix} h & c \\ c & h \end{bmatrix}
\]

(15)

\( \mathcal{J} \) and \( \mathcal{J}^* \) become respectively

\[
\mathcal{J} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

\[
\mathcal{J}^* = \begin{bmatrix} 0 & 1 \end{bmatrix}
\]

(16)

and Equation 14 can be written as

\[
Q(N) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & N-3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\]

(17)

\( N \) has been replaced by \( N-3 \) to try and take into account the fact that three residues are required to form a hydrogen bond.
The fraction of hydrogen bonds can be calculated using Equation 11 and replacing \( \partial \ln Q(N) / \partial \ln s \) by (111)

\[
\frac{\partial \ln Q(N)}{\partial \ln s} = Q(N)^{-1} J^* (E \ Q) \ \hat{U}^N \left( \begin{array}{c} O_N \\ E \end{array} \right) \ J
\]  

where \( E \) and \( Q \) are the identity and zero matrices, respectively, and have the same order as \( U \). \( \hat{U} \) is a super matrix and is given by

\[
\hat{U} = \left( \begin{array}{cc} U & U' \\ O & U \end{array} \right)
\]  

where \( U' = s (\partial U / \partial s) \).

In terms of the Zimm-Bragg model Equations 11 and 18 become

\[
Q(N) = (N-3)^{-1} (Q(N))^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \\
\times \begin{bmatrix} s & 1 & s & 0 \\ 0 & s & 1 & 0 \\ 0 & s & 0 & 1 \end{bmatrix}^{N-3} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}
\]

The advantages of the matrix method in calculating \( Q(N) \) and \( \theta(N) \) are numerous. With this method direct evaluation of \( Q(N) \), and hence \( \theta(N) \), can be made in contrast to the maximum term or the steepest descent methods. Also, this method is completely general in terms of the number of different residue states in the polymer that can be accommodated. If, for example, four residue states were required instead of two as in the Zimm-Bragg treatment, then the statistical weight matrix \( U \) would become a 4x4 matrix instead of the Zimm-Bragg's 2x2 matrix. With the advent of computers the matrix method is readily programable, and the computations of \( \theta(N) \) can be performed on the order of seconds.
As mentioned earlier in the general introduction, determination of \( \sigma \) and \( s \) for homopolymers of the naturally occurring amino acids in water is not experimentally feasible. Since the Zimm-Bragg treatment of the helix-coil transition is only applicable in the study of homopolymers, Scheraga and co-workers have modified this treatment to accommodate random copolymers as well, the "guest-host" technique (103). The random copolymers are usually readily made, and they can allow for the experimental determination of \( \sigma \) and \( s \). In this technique, the copolymers are composed of a nonionic homopolyamino acid, the "host" which is interdispersed with the desired amino acid, the "guest", hence the name "guest-host".

For a specific-sequence copolymer, e.g., ABABBAAX, the partition function, \( Q\{A, B\} \), would be

\[
Q\{A, B\} = \mathcal{J}^*_{\{A, B\}} \mathcal{J}
\]

(20)

In terms of the Zimm-Bragg notation

\[
\mathcal{U}_A = \begin{pmatrix} s_A & 1 \\ \sigma_A s_A & 1 \end{pmatrix} \tag{21a}
\]

and

\[
\mathcal{U}_B = \begin{pmatrix} s_B & 1 \\ \sigma_B s_B & 1 \end{pmatrix} \tag{21b}
\]

where the subscripts \( A \) and \( B \) denote the two different types of residues in the copolymer. However, in a solution of a random copolymer prepared from a mixture of \( A \) and \( B \) monomers, all molecules would not have the same random sequence, but would be made of an ensemble
of \( N \) molecules of varying sequence of relatively uniform chain length \( N \). The partition function for this ensemble is given by

\[
Q = \frac{(A,B)}{\{A,B\}} \sum_{A,B} Q(A,B)^N = \frac{\{A,B\}^N}{\{A,B\}^N} \sum_{A,B} P(A,B)
\]

(22)

where \( Q(A,B) \) is the number of molecules having a particular sequence \( \{A,B\} \) (with partition function \( Q(A,B) \) and \( P(A,B) \) is its probability of occurrence. In the "guest-host" treatment \( P(A,B) \) is regarded as that for a random-sequence copolymer with no nearest-neighbor correlations.

The fractional helical content is then given by

\[
\theta(N) = N^{-1} \sum_{A,B} P(A,B) \left( \partial ln Q(A,B) \right) / \left( \partial ln s \right)
\]

(23a)

where \( \partial / \partial ln s = \partial / \partial ln s_A + \partial / \partial ln s_B \) (23b)

Poland and Scheraga (61,111) have developed a treatment for obtaining \( \theta(N) \) which involves a hierarchy of approximations; the lowest order approximations reduce to earlier theories of Lifson (112), Lifson and Allegra (113), and Allegra (114), while the result is exact in the limit of infinite order. They called their treatment the Lifson-Allegra-Poland-Scheraga (LAPS) hierarchy. The LAPS hierarchy is based on the observation that \( \theta(N) \) of Equation 23 can be equally well computed from another partition function, \( Q^* \), which is not equal to \( Q \) of Equation 22, but defined as

\[
Q^* = \sum_{A,B} P(A,B) \frac{Q(A,B)}{C(A,B)} Q(A,B) = 1
\]

(24)

where \( Q(A,B) \) is a differentiable function of \( s_A \) and \( s_B \), and \( C(A,B) \) is a number which is the numerical value of \( Q(A,B) \). The last equality of Equation 24 follows from the facts that \( C(A,B) = Q(A,B) \) numerically
and \( P(A,B) \) is a normalized probability. The fractional helical content is then given by

\[
\Theta(N) = N^{-1} \frac{\partial \ln Q^+}{\partial \ln s} = N^{-1} \frac{\partial Q}{\partial \ln s}
\]

\[
= N^{-1} \frac{\partial}{\partial \ln s} \frac{P(A,B)}{C(A,B)}
\]

(25)

Since the last terms on the right-hand sides of Equations 23 and 25 are identical, the values of \( \Theta(N) \) computed from both equations are the same.

The advantage of introducing \( Q^+ \) is that, unlike \( Q \) of Equation 22, it can be generated with the aid of a matrix-like quantity, namely

\[
Q^+ = \begin{pmatrix} \frac{1}{J} & \frac{1}{J} \\ \frac{1}{J} & \frac{1}{J} \end{pmatrix}
\]

where

\[
U^+ = \begin{pmatrix} \frac{\alpha' \sigma_A}{u_A} + \frac{\sigma_B}{u_B} & \frac{\alpha'}{u_A} + \frac{1}{u_B} \\ \frac{\alpha' \sigma_A}{u_A} + \frac{\sigma_B}{u_B} & \frac{\alpha'}{u_A} + \frac{1}{u_B} \end{pmatrix}
\]

(27)

and \( u_A \) and \( u_B \) are given by Equation 21. In Equation 26, it is to be understood that the indicated operations are carried out so that the multiplication of factors in the "numerator" and "denominator" are performed independently and in parallel until finally the matrix products accumulating in the denominators are transformed into scalar.
quantities by operation with the end vectors \( J \) and \( J^* \). Upon matrix multiplication in this manner, the matrices \( U^A \) and \( U^B \) lead to the quantity \( C \{A, B\} \) and the factor \( \alpha' \)(for the case of a random distribution considered here) generate the quantities \( P \{ A, B \} \), where

\[
P \{ A, B \} = P_A^N A^N B^N = (1 - P_A)^N \left( \frac{P_A}{1 - P_A} \right)^N
\]

and \( P_A \) and \( P_B \) are the a priori probabilities of occurrence of \( A \) and \( B \) (\( P_A + P_B = 1 \)) and \( N_A \) and \( N_B \) are the average numbers of \( A \) and \( B \) units per chain (\( N_A + N_B = N \)), respectively. Since the factor \( (1 - P_A)^N \) is a constant for all \( \{A, B\} \), which cancels in the calculation of \( \Theta(N) \), it is dropped and assigned to each \( A \) unit the factor

\[
\alpha' = \frac{P_A}{1 - P_A}
\]

With \( \alpha' \) defined in this manner, \( P_A \) is taken to be the experimentally determined fraction of \( A \) in the copolymer, \( f_A \).

For large \( N \) it is much too difficult to evaluate \( Q^+ \) explicitly by Equation 26. To circumvent this difficulty the series of approximations for \( Q^+ \) represented by the LAPS hierarchy was devised. In these approximations, the matrix products in the denominators of Equation 27, obtained after \( m' \) explicit matrix multiplications have been performed, are replaced by scalar quantities, such that

\[
Q^+ = \gamma^* U^{+N}_m J = \gamma^* \left( U^{+m'}_m N/m' \right) \gamma J
\]

where \( U^{+}_m \) is now a proper matrix. For infinite chains, the scalars in the denominators of \( U^{+}_m \) are the largest eigenvalues (\( \lambda_A, \lambda_B; \lambda_{AA}, \lambda_{AB}, \lambda_{BB}; \lambda_{AAAA}, \lambda_{ABAA}, \lambda_{AABB}, \lambda_{ABB}, \lambda_{ABAB}, \lambda_{BBBB}, \) respectively).
of the matrix products in the denominators of $U_{\mu}^{+m'}$ for orders of approximation $m'=1, 2, \text{ and } 4$, respectively. In the succession of the approximations from $m'=0$ (lowest) to $m'=N$ (exact), $C_{\{A,B\}}$ varies from 1 to $Q_{\{A,B\}}$ numerically.

To accelerate the rate of convergence of the approximate to the exact result a device introduced by Allegra (114) was used. The value of $\alpha'$ in Equation 27 was taken as an adjustable parameter to satisfy the condition

$$f_A = (1/N) \left( \partial \ln Q_{m'}^+ / \partial \ln \alpha' \right)$$

where $f_A$ is the experimental value of the fraction of A units in the copolymer. In the limit, as $m' \to N$, $\alpha'$ will approach the value given by Equation 29.

For the two lowest orders ($m'=0$ and 1) of the LAPS hierarchy for random copolymers of finite chain length $N$, the partition function $Q_{m'}^+$ can be represented, in terms of the following modification of the 2 x 2 matrix of Equation 27

$$U_{\mu}^+ = \begin{pmatrix} (\alpha_A s_A + s_B) & (\alpha' + 1) \\ (\alpha_A s_A + \alpha_B s_B) & (\alpha' + 1) \end{pmatrix}$$

For the case of $m'=0$, $\alpha'$ is given by Equation 29. For the case of $m=1$, $\alpha'$ is considered to be adjustable. Any normalization factors from the "denominators" of Equation 27 have been omitted, since these only alter the value of $\alpha'$, which is adjusted anyway to satisfy Equation 31. Thus, for $m'=0$ and 1, $Q_{m'}^+$ can easily be written in terms of the eigenvalues of matrix $U_{\mu}^+$ (Equation 32) as

$$Q_{m'}^+ = J^* U_{\mu}^{+N} J = C_1^{A_1} + C_2^{A_2}$$

(33a)
where \( J^* = (0,1) \) and \( J = \text{col} (1,1) \) and

\[
\lambda_{1,2} = (1/2) \left\{ (\alpha' + 1) + (\alpha's_A + s_B) \pm \sqrt{((\alpha' + 1) + (\alpha's_A + s_B))^2 - 4 (\alpha' + 1) ((\alpha's_A + s_B) - (\alpha's_A + s_B)\}} \right\}^{1/2}
\]

(33b)

The coefficients \( C_1 \) and \( C_2 \) are given by

\[
C_{1,2} = \frac{(\alpha's_A + s_B) \lambda_{1,2}}{[\lambda_{1,2} - (\alpha' + 1)]^2 + (\alpha' + 1) (\alpha's_A + s_B)}
\]

(33c)

The fraction helical content is then given explicitly by Equation 34,

\[
\Theta(N) = \frac{1}{(1/N)} \left( \frac{d \ln Q^+}{d \ln s} \right)
\]

(34a)

where \( (3\ln \lambda_{1,2} / \partial \ln s) \) and \( (3C_{1,2} / \partial \ln s) \) are given by

\[
\frac{3\ln \lambda_{1,2}}{\partial \ln s} = \frac{1}{2\lambda_{1,2}} \left[ (\alpha's_A + s_B) \pm ((\alpha's_A + s_B) [((\alpha's_A + s_B) - (\alpha' + 1)]
\]

\[
+ 2 (\alpha's_A + s_B)(\alpha' + 1) \} / \left[ \right. (\alpha's_A + s_B) - (\alpha' + 1)^2 + 4 (\alpha' + 1)
\]

\[
\times (\alpha's_A + s_B)^{1/2}
\]

(34b)

\[
\frac{3C_{1,2}}{\partial \ln s} = C_{1,2} \left[ 1 - \frac{(\alpha' + 1)}{\lambda_{1,2}} \right] \frac{C_{1,2}}{\partial \ln s} + \frac{\partial \ln \lambda_{1,2}}{\partial \ln s} - 2 \frac{c_{1,2}^{3/2} (\lambda_{1,2} - (\alpha' + 1)]}{(\alpha's_A + s_B)}
\]

(34c)
The theoretical fraction of $A$, $F_A$, given by Equation 31, can be obtained explicitly in a similar manner as $\theta(N)$ by replacing $\partial/\partial \ln s$ by $\partial/\partial \ln \alpha'$ in Equation 34. Using this equation with the modification just mentioned, $\alpha'$ is adjusted to the value necessary to have the theoretical fraction of $A$, $F_A$, become equal to the experimental fraction of $A$, $f_A$.

The computer programs for the LAPS hierarchy of orders 0 and 1, and higher are available through the ASIS National Auxiliary Publication Service (115).

It was found theoretically (103) that in order for the low order ($m'=0$ and 1) approximations of the LAPS hierarchy to hold, $\sigma$ and $s$ for the two constituents of the copolymer should not differ appreciably from each other. Experimentally, however, for all the naturally occurring amino acids in water investigated to date using the LAPS hierarchy (92, 96-102), the first-order approximation ($m'=1$) was found to be not only quite adequate, but gave in most cases the same results as the exact treatment ($m'=N$). The values of $\sigma$ (taken as temperature independent) in these studies ranged from $1 \times 10^{-5}$ for glycine to $1 \times 10^{-2}$ for glutamic acid, and $s$ at $20^\circ C$ ranged from 0.6 to 1.4 for glycine and glutamic acid, respectively. The "host" polypeptide used in these studies was either poly[N\textsuperscript{5}-(3'-hydroxypropyl)-L-glutamine] or poly [N\textsuperscript{5}-(4'-hydroxybutyl)-L-glutamine] which have $\sigma$ and $s$ values at $20^\circ C$ of $\sigma = 2.8 \times 10^{-4}$ or $6.8 \times 10^{-4}$ and $s = 0.97$ or 1.02, respectively.

The assumption, used in the LAPS hierarchy, that the statistical weights for the two constituents are independent of the nature of their
neighbors is strictly the simplest assumption that could be made. However, the results, where proteins were considered (116-118), suggest that the assumption of independence is a very good first-order approximation. To test this assumption further with synthetic polypeptides, Scheraga and co-workers (92) made copolymers of poly $[N^5-(3'-\text{hydroxypropyl})-\text{L-glutamine}](\text{PHPG})$ and poly $[N^5-(4'-\text{hydroxybutyl})-\text{L-glutamine}](\text{PHBG})$ with different guest/host ratios. If the assumption is invalid, then different $\sigma$ and $s$ values would be found for the different guest/host ratios. Even though the PHPG content ranged from 6% to 82%, no differences in the experimentally determined $\sigma$ and $s$ values for the two amino acids could be detected. This may have been due to the close similarity in structure of the two side chains although their transition temperatures differ by 40°C.

The assumption that the copolymers are random is not critical. This has been demonstrated by Lehman and McTague (119) who have compared random and Markoffian distributions for the same $f_A$ and found little differences. Also, Poland and Scheraga (120,121) have computed detailed helical probability profiles for a given random sequence for various degrees of coarse graining of the sequence. Coarse graining is a technique in which the polypeptide is seen to consist of blocks of A and B units, which are each M units long. The A and B blocks are then each considered as a pseudounit, $A'$ and $B'$ respectively, thus effectively reducing the length of the molecule from N to $N/M$ units. The degree of coarse graining was found to make little difference if the grain size was much less than the correlation length ($\propto \sigma^{1/2}$).
The LAPS hierarchy, even the low order approximations, appears to be quite adequate in the study of the helix-coil transition of naturally occurring amino acids. It has now been thoroughly tested (92, 96-102), and it appears probable that this method will become standard for the evaluation of $\sigma$ and $s$ of random copolymers.

The value of $\sigma$ and $s$ at 20°C in water for the amino acids determined to date by Scheraga and co-workers (92, 96-102) are shown in Table 3. The table also gives the LAPS hierarchy approximation used in the calculations. In all calculations $\sigma$ was taken as temperature independent. A general trend is apparent from the table, that is, as the value of $s$ increases the value of $\sigma$ also increases. This trend is better illustrated in Figure 3 as a plot of $-\ln \sigma$ vs. $\ln s$ for the ten amino acids shown in Table 3. The error bars shown in the figure

Table 3. $\sigma$ and $s$ values of amino acids determined to date in water at 20°C (92,96-102)

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>$\sigma \times 10^4$ a</th>
<th>$s$</th>
<th>$m^b$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>0.1</td>
<td>0.59</td>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>Ser</td>
<td>0.75</td>
<td>0.76</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>Val</td>
<td>1</td>
<td>0.93</td>
<td>1</td>
<td>101</td>
</tr>
<tr>
<td>Glu, pH 8.0 c</td>
<td>6</td>
<td>0.97</td>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>PHPG d</td>
<td>2.8</td>
<td>0.97</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>PHBG e</td>
<td>6.8</td>
<td>1.02</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Ala</td>
<td>8.0</td>
<td>1.07</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Phe</td>
<td>18</td>
<td>1.08</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Leu</td>
<td>33</td>
<td>1.14</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>Glu, pH 2.3</td>
<td>100</td>
<td>1.35</td>
<td>1</td>
<td>102</td>
</tr>
</tbody>
</table>

a. temperature independent values
b. LAPS hierarchy approximation used (see text)
c. in 0.1 N KCl
d. poly [N\text{5}-(3'-hydroxypropyl)-L-glutamine]
e. poly [N\text{5}-(4'-hydroxybutyl)-L-glutamine]

are an estimate of the errors in the experimental parameters measured.

The dashed arrows for serine, valine and glutamic acid at pH 8.0
are used to illustrate that values of \( \sigma \) much less than the values
reported in Table 1, can fit the experimental data just as adequately
as the reported values. The straight line drawn through the points
has a slope of -8.6 and an intercept of 7.6. The standard deviation
and the correlation coefficient of this line is 0.67 and 0.95,
respectively. Thus, a correlation between \( \sigma \) and \( s \) can be given
empirically by

\[
\ln \sigma = 8.6 \ln s - 7.6 
\]  

(35)

In an attempt to show what might contribute to the correlation
found experimentally between \( \sigma \) and \( s \), greater attention was paid to
the ends of helical sequences by expanding the size of matrix \( \Psi \)
in Equation 14. This larger \( \Psi \) matrix was then used to calculate \( \Theta(N) \) by
Equation 18 as a function of \( N \) and \( s \) for a given value of \( \sigma \). These
sets of calculated \( \Theta(N) \) were then fit by a least-squares fitting
routine with a new set of \( \sigma \) and \( s \)'s to the standard Zimm-Bragg \( \Theta(N) \)
equation (Equation 19). The new set of fitted \( \sigma \) and \( s \)'s were then
plotted as \(-\ln \sigma \) vs. \( \ln s \) to see if a correlation between \( \sigma \) and \( s \)
could be detected.
Figure 3. Experimental correlation between the Zimm-Bragg helix-coil transition parameters $\sigma$ and $s$ at 20°C in water for nine amino acids. The amino acids, from left to right, are glycine, serine, valine, glutamic acid at pH 8.0, PHPG, PHBG, alanine, phenylalanine, leucine, and glutamic acid at pH 2.3.
2.2 Matrix Formulations

Let us first consider the nearest neighbors to the left and to the right of residue \( i \) in the polymer chain. All the possible combinations, of which there are eight, of three residue sequences each having two states can be accommodated by a \( 4 \times 4 \) statistical weight matrix \( \mathcal{U} \). This matrix can be represented as

\[
\begin{pmatrix}
i + 1 & h & c & c \\
i - 1 & h & h & c \\
h & h & s & o^{1/2} & o \\
h & c & o & o & 1 & 1 \\
c & h & o^{1/2} & o^{1/2} & o & o \\
c & c & o & o & 1 & 1
\end{pmatrix}
\]

Let us define a number \( m \) as the number of near-neighbor residues to residue \( i \) considered in matrix \( \mathcal{U} \), in this case \( m = 2 \). Using the definitions of helical and coiled states as outlined in the introduction, the statistical weights for this matrix and the following larger matrices have been chosen as follows:

a) A statistical weight of \( s \) or \( s^{-1/m} \) is assigned to residue \( i \) if residue \( i \) is in the helical state. \( s \) is assigned if all other near neighbor residues considered in \( \mathcal{U} \) are in the helical state; otherwise, \( s^{-1/m} \) is assigned.

b) A statistical weight of 1 is assigned to residue \( i \) if residue \( i \) is in the coiled state.
Since there are only eight possible combinations of c's and h's in a three residue sequence, the suggestion of using a 3 x 3 matrix instead of the 4 x 4 matrix should be considered. A procedure for reducing a matrix size has been illustrated by Poland and Scheraga (122). They state that two conditions are required in order for a contraction of a matrix to take place without loss of information. First, there must be a redundancy in assigning the statistical weight to a conformation, that is, the total correlation implied in the matrix size is not required for the assignment of the statistical weight. Second, the contraction must be able to be made simultaneously in corresponding positions in the index rows and index columns, without losing the information to assign correctly all the statistical weights involved in the contraction. Thus, while the matrix supplies us the following possibilities

\[
\begin{array}{ccc}
  i-1 & i & i+1 \\
  c & c & h \\
  c & c & c \\
\end{array}
\]

the single possibility

\[
\begin{array}{ccc}
  i-1 & i & i+1 \\
  c & c & hUc \\
\end{array}
\]

would be adequate to assign the statistical weight of 1 to residue i. The notation hUc means "either h or c". The reduced 4 x 4 matrix, a 3 x 3 matrix, now becomes

\[
\gamma(3) = \begin{bmatrix}
  i+1 & h & c & hUc \\
  i-1 & i & h & h & c \\
  h & h & s & \sigma^{1/2} & o \\
  h & c & o & o & l \\
  c & hUc & \sigma^{1/2} & \sigma^{1/2} & 1 \\
\end{bmatrix}
\]

(37)
This $J^3$ matrix is the same 3 x 3 statistical weight matrix used by Lifson and Roig (63).

2.3 End Effects

We now have to determine the correct procedure for taking account of the polymer chain ends. For this three residue case we can not use the 3 x 3 statistical weight matrix $J^3$ for residues 1 and $N$ because nothing either precedes or follows these residues, respectively. It is only possible for residue 1 to initiate a helical sequence. There are a maximum of four possible different statistical weights that could be assigned to residue 1, however, we have only two different weights to chose from, $\sigma^{1/2}$ and 1. The statistical weight matrix for residue 1, $U_1^3$, must conform to the 3 x 3 matrix, $U_2^3$, of residue 2. $U_1^3$ is, therefore, a row matrix with three elements.

$$
U_1^3 = \begin{pmatrix}
2 & h & c & hUc \\
1 & h & h & c \\
hUc & \sigma^{1/2} & \sigma^{1/2} & 1
\end{pmatrix}
$$

(38)

Residue $N$ is not followed by any other residue, and so it too can have only a maximum of four different statistical weights. As before, there are only two different weights to chose from, $\sigma^{1/2}$ and 1. The conforming problem with $U_{N-1}^3$ of residue $N-1$ requires that $U_N^3$ be a column matrix with three elements, and thus

$$
U_N^3 = \begin{pmatrix}
N-1 & N & hUc \\
h & h & \sigma^{1/2} \\
h & c & 1 \\
c & hUc & \sigma^{1/2} + 1
\end{pmatrix}
$$

(39)
The last element in the column, $\sigma^{1/2} + 1$, arises from the fact that both states for residue $N$, $h$ and $c$, must be accounted for when residue $N - 1$ is in the coiled state, $c$.

It can be seen by inspection that the end statistical weight matrices, $U_1(3)$ and $U_N(3)$, can be replaced by

$$U_{1} (3) = (\sigma^{1/2}, 1) = \begin{pmatrix} s & \sigma^{1/2} & o \\ o & o & 1 \\ \sigma^{1/2} & \sigma^{1/2} & 1 \end{pmatrix} = J^{*} U(3)$$

and

$$U_{N} (3) = \text{col} (\sigma^{1/2}, 1, \sigma^{1/2} + 1) = \begin{pmatrix} s & \sigma^{1/2} & o \\ o & o & 1 \\ \sigma^{1/2} & \sigma^{1/2} & 1 \end{pmatrix} = U_0(3) J$$

In general the end effects for this nearest-neighbor case as well as the subsequent larger near-neighbor cases can be accounted for by choosing the $J^*$ and $J$ matrices, whose elements are 0 and 1, appropriately. $J^*$, a row matrix of order $m + 1$, is chosen, such that, the row of $U(m + 1)$ which has all residues preceding $i$ in the coiled state is extracted when premultiplied by $J^*$. $J$, a column matrix of order $m + 1$, is chosen, so that, the columns of $U(m + 1)$ which have residue $i + 1$ in the coiled state are extracted and added together when postmultiplied by $J$.

The reduced statistical weight matrices, $U(m + 1)$, of higher order along with $m$, $J^*$ and $J$ are given in Appendix A.
2.4 Computations

Equation 14 was used to compute the partition function, Q(N), using the appropriate \( \Psi(m+1) \), \( J^* \) and \( J \) matrices for the various near-neighbor models. In Equation 11, \( N - 3 \) was replaced by \( N - m \) and used to calculate \( \Theta(N) \). \((\partial \ln Q(N) / \partial \ln s) \) of Equation 11 was computed using Equation 18.

In order to conserve computer time in these calculations, the method of matrix squaring was used. The generator matrix \( \hat{\Psi} \) in Equation 18 was multiplied by itself and the product set equal to a new matrix \( \hat{\Psi} \). This \( \hat{\Psi} \) matrix was then set equal to \( \hat{\Psi} \), and the cycle was repeated until a chosen value of \( N \) was reached. Each cycle incremented \( N \) by \( 2N_{\text{old}} \). The product matrix \( \hat{\Psi} \) was then premultiplied by \( J^* (\hat{\Psi}, \hat{\Theta}) \) and set equal to \( \hat{\Theta} \). \( Q(N) \) was extracted by postmultiplying \( \hat{\Theta} \) by \( \text{col} (\hat{\Psi}, \hat{\Theta}) \). \((\partial \ln Q(N) / \partial \ln s) \) was extracted by postmultiplying \( \hat{\Theta} \) by \( \text{col} (Q, \Theta) \).

Because of the matrix squaring technique used, \( N \) was chosen to be 32, 64, 128, 256, 512 and 1024 for each value of \( s \). Polymers with degrees of polymerization, or \( N \), in this range have usually been used to determine experimental values of \( \sigma \) and \( s \). Although experimental values of \( s \) vary from 0.6 to 1.4, it was necessary to increase the range of \( s \) from 0.2 to 6.0. The reason for this is because the fraction helix, \( \Theta(N) \), calculated becomes less sensitive to \( s \) as the size of \( \Psi(m+1) \) increases. The value of \( \sigma \) used in the calculations varied from 1.0 to 0.0001.
A set of $\theta(N)$ vs. $N$ was calculated using given values of $\sigma$ and $s$. A new $\sigma$ and $s$ was computed from this set by a least-square fitting technique, described in appendix B, to the Zimm-Bragg equation (62) for $\theta(N)$ given as

$$\theta(N) = \left[ \frac{s}{(N-3)} \right] \left\{ [ (N-2) \frac{\lambda_0}{\lambda_o} + (\lambda_0 - 1)/ (\lambda_o - s) ] \lambda_o^{N-2} \right. \nonumber$$

$$X (\lambda_o - s) + [ (N-2) \frac{\lambda_1}{\lambda_1} + (1 - \lambda_1) / (s - \lambda_1) ] \lambda_1^{N-2} \nonumber$$

$$\left. X (s - \lambda_1) \right\} / [\lambda_o^{N-2} (\lambda_o - s) + \lambda_1^{N-2} (s - \lambda_1)] - \left[ \frac{s}{(N-3)} \right] \nonumber$$

$$X \left[ (\lambda_o - \lambda_1) / (\lambda_o - \lambda_1) \right] \nonumber$$

(42a)

where $\lambda_{o,1} = 1/2 \{ (1 + s) \pm [ (1 - s)^2 + 4 \sigma s]^{1/2} \} \nonumber$

(42b)

and $\lambda'_{o,1} = (\partial \lambda_{o,1} / \partial \sigma) \nonumber$

(42c)

The initial $s$ used in $\Psi(m+1)$ was incremented, and another set of $\theta(N)$ vs. $N$ was computed holding the initial $\sigma$ constant. Additional $\sigma$ and $s$'s were fit to this new data. The procedure was continued until the $s$ of $\Psi(m+1)$ had varied from 0.2 to 6.0. The new set of fitted $\sigma$ and $s$'s was then plotted as $-\ln \sigma$ vs. $\ln s$. The $\sigma$ of $\Psi(m+1)$ was incremented, and a new plot was calculated in a similar manner. Plots of this type were calculated for each successively larger statistical weight matrix.

2.5 Results

Figure 4 shows the results of the fitting calculations, with $\sigma = 0.0005$, as plots of $-\ln \sigma'$ vs. $\ln s'$ for the eight expanded matrices examined. The primes on $\sigma'$ and $s'$ are to distinguish the fitted parameters from the statistical weight parameters in the expanded matrices,
Figure 4. Calculated correlations between $\sigma'$ and $s'$ using the $\mathcal{A}^{1/10}(\text{arrow})$ through $\mathcal{A}^{1/3}$ statistical weight matrices to compute the fraction helical content for $\sigma = 0.0005$. The horizontal line is the Zimm-Bragg treatment.
\[ U(10) \]
\[ U(9) \]
\[ U(4) \]
\[ U(3) \]
\[ Z.B. \]
$V(m + 1)$. The dashed lines for the $V(9)$ and $V(10)$ statistical weight matrices in the figure are the anticipated shapes of the curves in the region of $\ln s'$ between 0.02 and 0.15. The difficulty in fitting the curves in this region is due to a very large value of $(3F/3s')$, see Equation B-7, Appendix B. Large values of $(3F/3s')$ result in underflows when solving the set of simultaneous equations of Equation B-6, Appendix B.

All the plots in Figure 4 show similar curvature and shape. Also, two general types of behavior are observable in these plots. First for $s'$ small, $s'$ seems to reach an asymptotic limit, $s'$\_min. This minimum value for $s'$ increases with increasing matrix size. Second, at large values of $s'$, $s' > 1.3$ or $\ln s' > 0.26$, $\sigma'$ seems to reach an asymptotic limit, $\sigma'_{\,\max}$. This limiting value for $\sigma'$ appears to be a function of the matrix size. $\sigma'_{\,\max}$ increase slightly with increasing matrix size. Besides this limiting $\sigma'_{\,\max}$ value, all the curves show a maximum $\sigma'_{\,\max}$ value. This $\sigma'_{\,\max}$ value is slightly smaller than $\sigma'_{\,\max}$ As with $\sigma'_{\,\max}$, $\sigma'_{\,\max}$ increases with increasing matrix size.

Increasing the importance of the ends of helical sequences in the polymer chain, increasing $V(\,m + 1)$, appears to increase the correlation between $\sigma'$ and $s'$. This correlation, dependence of $\sigma'$ on $s'$, increases with increasing matrix size up to a matrix size of $7 \times 7$, $V(7)$. Increasing the matrix size after this point does not appear to affect the dependence of $\sigma'$ on $s'$ in the region of $\ln s'$ between -0.20 and 0.13. As mentioned earlier, both $s'_{\,\min}$ and $\sigma'_{\,\max}$ are still matrix size dependent.
Figure 5 shows plots of $-\ln \sigma'$ vs. $\ln s'$ for the $U(10)$ matrix as a function of $\sigma$. As might be expected the dependence of $\sigma'$ and $s'$ decreases with increasing $\sigma$ in the region of $\ln s'$ between $-0.3$ and $0.2$. It is interesting to note that $\sigma'_{\text{max}}$ is unaffected by $\sigma'$, however, is dependent on $\sigma$ and increases with decreasing $\sigma'$. 

A comparison of the plots of $\theta$ vs. $\ln s$ between the Zimm-Bragg treatment and the expanded matrix treatment is shown in Figure 6 for $\sigma=0.0005$ and $N=1024$. The most noticeable effect of the new treatment is a shift in the helix-coil transition from $\ln s'=0.0$ for the Zimm-Bragg treatment to higher values of $\ln s'$ for the expanded matrix treatment. The transition occurs at $\ln s = 0.14$ for the $U(5)$ matrix and at $\ln s = 0.37$ for the $U(10)$ matrix.

Because of the greater attention being paid to the ends of helical sequences, the helix-coil transition sharpens as the matrix size increases. This can be seen from Figure 6. The figure also gives some clues as to why a dependence of $\sigma'$ and $s'$ might arise from the expanded matrix treatment. By defining quantities $r$ and $R$ as

$$ r = \frac{\ln s(0.95) - \ln s(0.5)}{\ln s(0.05) - \ln s(0.5)} $$

and

$$ R = \frac{r \text{ expanded matrix}}{r \text{ Zimm-Bragg}} $$

where $\ln s(a)$ is the value of $\ln s$ which gives a value of $a$ for $\theta$, a qualitative estimate of the dependence of $\sigma$ on $s$ can be shown. Table 4 gives these quantities for the curves shown in Figure 6.
Figure 5. Calculated correlations between $\sigma^*$ and $s^*$ using the U(10) statistical weight matrix with $\sigma = 0.0001$ (arrow), 0.001, 0.01, 0.1, and 1.0.
Figure 6. Helix-coil transition curves for the Zimm-Bragg treatment, using the $\mathcal{W}(5)$ statistical weight matrix, and the $\mathcal{W}(10)$ statistical weight matrix with $\sigma = 0.0005$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 4. r and R values for the curves shown in Figure 6 (see text)

<table>
<thead>
<tr>
<th>Matrix</th>
<th>r</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z.B.</td>
<td>1.236</td>
<td>1</td>
</tr>
<tr>
<td>ß(5)</td>
<td>1.433</td>
<td>1.16</td>
</tr>
<tr>
<td>ß(10)</td>
<td>2.36</td>
<td>1.91</td>
</tr>
</tbody>
</table>

As mentioned earlier the curve for the ß(5) matrix in Figure 4 shows less dependence of $\sigma'$ on $s'$ than does the curve for the ß(10) matrix. The values of R in Table 4 increase from 1 for the Zimm-Bragg treatment to 1.16 for the ß(5) matrix to 1.9 for the ß(10) matrix. Thus, a qualitative statement about R seems reasonable, that is, as R becomes increasingly greater than 1, the dependence of $\sigma'$ on $s'$ increases.

After comparing the calculated fitted curves of Figure 4 with the experimental correlation curve of Figure 3, it was decided that the curves for the ß(4) matrix and the ß(5) matrix seem to fit the experimental correlation curve best. In order to obtain a better fit with the experimental curve some modifications to the statistical weights chosen for these matrices were examined.

The first modification tested using the ß(5) matrix was the replacement of some of the $\sigma^{1/4}$ statistical weight in ß(5) by $\sigma^{1/4} s$. This replacement occurred whenever residues i, i-1 and i-2 were all in the helical state regardless of the states of residues i-3 or i+1. The resultant ß(5) mod 1 is as follows...
The $\hat{Y}$ matrix of $\hat{U}$ remained as defined by Equation 18 in the introduction. The effect of this modification is shown in Figure 7 for the curve labeled $\hat{U}(5)_{\text{Mod 1}}$. This modification causes a decrease in the correlation between $\sigma$ and $\sigma'$. It also gives a slight negative correlation between $\sigma$ and $\sigma'$ in the region of $\ln s'$ between $-0.5$ to $0.0$.

The next modification attempted was to keep $\hat{U}(5)$ as defined by modification 1, Equation 45, but to replace $\sigma'$ with $\sigma^1/4$.

The result of this second modification is shown in Figure 7 labeled $\hat{U}(5)_{\text{Mod 2}}$. This second modification causes an increase in the correlation between $\sigma$ and $\sigma'$ and is in the direction desired.

A third modification was examined. In this case all the $\sigma$ in $\hat{U}(5)$ were replaced by $\sigma^1/4$. The $\hat{Y}$ matrix was kept as defined in modification 2, Equation 46. A further increase in the correlation between $\sigma$ and $\sigma'$ is observed as shown in Figure 7, curve labeled $\hat{U}(5)_{\text{Mod 3}}$. 

\[
\hat{U}(5)_{\text{Mod 1}} = \begin{pmatrix}
s & \sigma^1/4 & o & o & o \\
o & o & o & o & 1 \\
\sigma^1/4 & \sigma^1/4 & o & o & o \\
o & \sigma^1/4 & \sigma^1/4 & o & o \\
o & \sigma^1/4 & o & \sigma^1/4 & 1 \\
\end{pmatrix} \quad (45)
\]
Figure 7. Calculated correlations between $\sigma^*$ and $s^*$ using the $\Psi(5)$ modification 3 (arrow), $\Psi(5)$ modification 2, $\Psi(5)$, and the $\Psi(5)$ modification 1 statistical weight matrix and $\sigma = 0.0005$. The dashed line is the empirical correlation found in Figure 3.
Modifications 2 and 3 were also employed on the $V(4)$ matrix giving the modified $V$ matrices as:

For modification 2, $V(4)_{\text{Mod 2}}$:

$$V(4)_{\text{Mod 2}} = \begin{pmatrix}
    s & \sigma^{1/3} & 0 & 0 \\
    0 & 0 & 0 & 1 \\
    \sigma^{1/3} & \sigma^{1/3} & 0 & 0 \\
    0 & \sigma^{1/3} & \sigma^{1/3} & 1
\end{pmatrix} \quad \text{(47)}$$

And for modification 3, $V(4)_{\text{Mod 3}}$:

$$V(4)_{\text{Mod 3}} = \begin{pmatrix}
    s & \sigma^{1/3} & 0 & 0 \\
    0 & 0 & 0 & 1 \\
    \sigma^{1/3} & \sigma^{1/3} & 0 & 0 \\
    0 & \sigma^{1/3} & \sigma^{1/3} & 1
\end{pmatrix} \quad \text{(48)}$$

The results of these modifications are shown in Figure 8. As with the $V(5)$ matrix, the modifications cause an increase in the correlation between $\sigma^*$ and $s^*$ with modification 3 having the greater effect.

Figure 9 is a plot of $\theta$ vs. $\ln s$ for the Zimm-Bragg matrix treatment and the $V(4)_{\text{Mod 2}}$ matrix treatment. The graph shows the effect of increasing $\sigma$ from 0.0005 to 0.005 in both matrix treatments. As with the Zimm-Bragg matrix treatment, the $V(4)_{\text{Mod 2}}$ matrix treatment shows increasing cooperativity with decreasing $\sigma$. Unlike the Zimm-Bragg treatment, however, the $V(4)_{\text{Mod 2}}$ matrix treatment has a shift in the helix-coil transition. The value of $\ln s(0.5)$ shifts from 0.080 for $\sigma=0.0005$ to 0.172 for $\sigma=0.005$. Similar effects were observed for the other expanded matrices.

It is now of interest to look at the ability of the Zimm-Bragg matrix treatment to fix the expanded matrix treatment. One measure of
Figure 8. Calculated correlations between $\sigma'$ and $s'$ using the $\Psi(4)$ modification 3, $\Psi(4)$ modification 2, and the $\Psi(4)$ statistical weight matrix with $\sigma = 0.0005$. The dashed line is the empirical correlation found in Figure 3.
Figure 9. Helix-coil transition curves for the Zimm-Bragg treatment, and using the $\xi(4)$ modification 2 statistical weight matrix with $\sigma = 0.005$ and 0.0005.
of the goodness of fit is the unbiased estimate of the standard deviation, \( \hat{\sigma} \), Equation B-10, Appendix B. \( \hat{\sigma} \) does not allow, however, the direct comparison of one point on the \(-\ln \sigma^\prime \) vs. \( \ln s^\prime \) plot with another. A better comparison can be achieved by looking at a dimensionless quantity called the coefficient of variation (128), \( \nu \), defined by

\[

\nu = \frac{\sigma}{\bar{y}}
\]

where \( \bar{y} \) is the mean value of the observed quantity \( y \), in this case, \( \theta \).

Table 5 gives the coefficients of variation, \( \nu \), and the percent deviations in \( \sigma^\prime \) and \( s^\prime \) for several points along the \(-\ln \sigma^\prime \) vs. \( \ln s^\prime \) curve for the \( H(4) \) Mod 3 matrix. The deviations in \( \sigma^\prime \) and \( s^\prime \) were calculated using Equation B-11 of Appendix B.

<table>
<thead>
<tr>
<th>( s )</th>
<th>( s^\prime )</th>
<th>% deviation^a in ( s^\prime )</th>
<th>( \sigma^\prime \times 10^4 )</th>
<th>% deviation^a in ( \sigma^\prime )</th>
<th>( \nu \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.617</td>
<td>0.74</td>
<td>0.0027</td>
<td>3.08</td>
<td>2.03</td>
</tr>
<tr>
<td>0.4</td>
<td>0.657</td>
<td>0.44</td>
<td>\textbf{0.051}</td>
<td>2.04</td>
<td>1.51</td>
</tr>
<tr>
<td>0.6</td>
<td>0.717</td>
<td>0.34</td>
<td>0.300</td>
<td>1.99</td>
<td>1.93</td>
</tr>
<tr>
<td>0.8</td>
<td>0.803</td>
<td>0.21</td>
<td>1.11</td>
<td>1.85</td>
<td>2.87</td>
</tr>
<tr>
<td>1.0</td>
<td>0.931</td>
<td>0.02</td>
<td>2.92</td>
<td>0.47</td>
<td>1.51</td>
</tr>
<tr>
<td>1.2</td>
<td>1.088</td>
<td>0.07</td>
<td>5.29</td>
<td>2.67</td>
<td>5.58</td>
</tr>
<tr>
<td>1.4</td>
<td>1.220</td>
<td>0.37</td>
<td>9.16</td>
<td>13.54</td>
<td>9.00</td>
</tr>
<tr>
<td>1.6</td>
<td>1.307</td>
<td>0.39</td>
<td>9.16</td>
<td>19.97</td>
<td>5.94</td>
</tr>
<tr>
<td>1.8</td>
<td>1.362</td>
<td>0.37</td>
<td>7.99</td>
<td>21.66</td>
<td>3.92</td>
</tr>
<tr>
<td>2.0</td>
<td>1.404</td>
<td>0.35</td>
<td>7.11</td>
<td>22.96</td>
<td>9.92</td>
</tr>
<tr>
<td>2.2</td>
<td>1.437</td>
<td>0.35</td>
<td>6.52</td>
<td>24.24</td>
<td>2.39</td>
</tr>
<tr>
<td>2.4</td>
<td>1.465</td>
<td>0.35</td>
<td>6.10</td>
<td>25.52</td>
<td>2.08</td>
</tr>
</tbody>
</table>

^a. All deviations are one standard deviation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
It can be seen from Table 5 that the ability to fit $s'$ is excellent over the entire range of the curve. The ability to fit $\sigma'$, on the other hand, is position dependent on the $-\ln \sigma'$ vs. $\ln s'$ curve. $\sigma'$ has a good fit in the region of the $-\ln \sigma'$ vs. $\ln s'$ plot near $s' = 1$, $\ln s' = 0$. A fair fit of $\sigma'$ is achieved near $s'_{\text{min}}$, and a poor fit is obtained in the region of $s' > 1.3$, $\ln s' > 0.26$. The coefficients of variation, $v$, indicate that the calculated $\theta$ using the fitted parameters is within two to three significant figures of the "observed" $\theta$. Similar variations in the calculated parameters were observed for the other matrices tested.

2.6 Discussion

When the expanded statistical weight matrices, $V(m+1)$, were first perceived, only two different statistical weights for residue $i$ in the helical state were considered. This simple model envisions, like previous models (61-63), one statistical weight, $\sigma^{1/m}$, as purely an entropy term and the other, $s$, as a free energy term. Thus, $\sigma^{1/m}$ and $s$ could be written as

$$\sigma^{1/m} = e^{\Delta S_{\text{res}}/R}$$

(50)

and

$$s = e^{-\Delta G_{\text{res}}/RT}$$

(51)

$\Delta S_{\text{res}}$ is the entropy change when a residue goes from the coiled state to the helical state. $\Delta G_{\text{res}}$ is the free energy change of a residue going from the coiled to the helical state when at least $m$ of its neighbors are already in the helical state. The entropy
change, $\Delta S_{\text{res}}$, should be negative based on the definitions of a coiled and helical state, see introduction above, and thus, $\sigma^{1/m}$ should be expected to be less than one.

A very rough estimate of $\Delta S_{\text{res}}$ for an unperturbed poly-peptide chain may be calculated from the dipeptide conformational energy map of L-alanine (67). The conformational entropy for the unperturbed random coil of poly(L-alanine) has previously been calculated by Brant, Miller and Flory (67) using the equation (133)

$$S = R \ln Q + \langle E \rangle / T$$

where $Q$ is the conformational partition function and $\langle E \rangle$ is the average energy which are given by

$$Q = \sum_{\phi} \sum_{\psi} \exp \left[ -E(\phi,\psi) / RT \right]$$

$$\langle E \rangle = Q^{-1} \sum_{\phi} \sum_{\psi} E(\phi,\psi) \exp \left[ -E(\phi,\psi) / RT \right]$$

$\phi$ and $\psi$ were incremented by $10^\circ$ intervals over the entire conformational map. The entropy for the coiled state, $S_c$, at 20°C was calculated to be 10 eu.

The entropy for the helical state, $S_h$, was calculated in a similar fashion. The $\phi$, $\psi$ angles, in this case, were limited to a narrow range around the $\phi$, $\psi$ angles for the $\alpha$-helix, namely $\phi = 110^\circ - 130^\circ$, $\psi = 120^\circ - 140^\circ$. As before the $\phi$, $\psi$ angles were incremented by $10^\circ$ intervals. $S_h$ at 20°C was then calculated to be 4 eu. The entropy change for this transition, $\Delta S_{\text{res}}$, is therefore, $\Delta S_{\text{res}} = S_h - S_c = -6$ eu. If this unperturbed model was completely correct,
$\sigma^{1/m}$ would have a value of 0.05. However, unperturbed conditions cannot be expected to hold for the helix-coil transition, since solvent interactions play an important role in the transition (61), e.g., hydrogen bonding with the polypeptide backbone. These calculations are only meant as first-order approximations for the magnitude of $\sigma^{1/m}$.

$\sigma$ was raised to the $1/m$th power in order to try and have some relationship to the Zimm-Bragg $\sigma$. For example, the sequence cchhhccc would be assigned a weight of $\sigma$ by both the $U(3)$ matrix treatment and the Zimm-Bragg treatment.

$\Delta G_{\text{res}}$ could be written in terms of the residue enthalpy change, $\Delta H_{\text{res}}$, and the residue entropy change, $\Delta S_{\text{res}}$, for the transition as

$$\Delta G_{\text{res}} = \Delta H_{\text{res}} - T\Delta S_{\text{res}}.$$  \hfill (53)

$\Delta H_{\text{res}}$ is usually considered to include such stabilizing forces as hydrogen bonds (61-63) and hydrophobic interactions (61, 129-132), and hence, is expected to be solvent and temperature dependent. Thus, for this simple model only $s$ should be sensitive to environmental conditions, and could be less than, equal to or greater than one depending on these conditions.

By using Equations 11 and 18 with the expanded $U(m + 1)$ matrix, only those residues which are in the interior of a helical sequence are considered to be helical. For example, using the $U(4)$ matrix, the sequence cchhhhhccc would be considered to be only 2/10 helical instead of 1/2 helical. Or putting it another way, only 2/10 of
the residues in the chain would be "observed" as being helical. Most helix-coil transition studies of polypeptides are done using optical rotatory methods. It should be reasonable to assume that if the ends of helical sequences are "floppy", then they will not have the same optical rotation as the more rigid residues in the interior of the helical sequence. Hence, these end residues would not be "observed" to be as helical as the interior residues.

It is evident from Figure 4 that some correlation between $\sigma'$ and $s'$ could be obtained even in the case of the $\Psi(3)$ matrix. However, none of the curves in Figure 4 fit the experimental points of Figure 3 over the entire range of $\ln s'$ considered. The curves computed using either the $\Psi(4)$ or $\Psi(5)$ matrix fit the experimental points, error bars included, in the region of $\ln s'$ between -0.5 and 0.0. Beyond $\ln s'$ greater than zero, none of the curves were able to fit the experimental data.

In an attempt to obtain a better fit using the $\Psi(4)$ and $\Psi(5)$ matrices, the modifications described in the results were examined. By changing some of the statistical weights, as in modifications 1 and 2, in $\Psi(4)$ and $\Psi(5)$ from $\sigma^{1/m}$ to $\sigma^{1/m_s}$, in effect what was done was to make some particular combinations of c's and h's environmental sensitive, i.e., solvent composition and temperature sensitive. In particular whenever a three helical sequence occurred, i.e., cchhhccc, the first two helical residues were assigned the usual weight of $\sigma^{1/4}$, in the case of $\Psi(5)$, and the third was assigned a weight of $\sigma^{1/4_s}$.
Or for this sequence, the weight of $\sigma^{3/4}$s would be assigned instead, as before, just a weight of $\sigma^{3/4}$. This could be interpreted as being an enthalpy increase after one turn of the helix is achieved, possibly a hydrogen bond.

The statistical weight of $\sigma^{1/m}$s was chosen instead of s because, although some enthalpy might be gained from one turn of the helix, additional enthalpy could be gained when a forth residue down the sequence was converted into the helical state. At that time, hydrophobic interactions, for example, could be possible between the side chains of residue $i$ and residue $i-3$. This type of side chain interaction is obviously an over simplification because not all amino acid side chains are capable of such hydrophobic interactions. The side chain of L-alanine, for example, is not long enough for such interactions. L-glutamate's side chain is long enough, but it carries a negative charge and would actually have unfavorable near-range interactions. However, if we remember how the experimental data of Figure 3 were obtained, favorable near-range interactions become feasible even with these amino acids.

The "host" used in these experiments was either poly$[N^5-(3'-\text{hydroxypropyl})-L\text{-glutamine}]$ or poly$[N^5-(4'-\text{hydroxybutyl})-L\text{-glutamine}]$. Since the fraction of the "guest", the desired amino acid, was low, the probability of a "guest" being four residues down from a "host" is high. Also, the side chain of the "host" is long enough to be able to interact with a "guest", even L-alanine, four residues down the chain.
Thus, side chain hydrophobic interactions, hydrogen bonding and electrostatic interactions become possible.

As seen from Figure 7, modification 1, $U(5)^{\text{mod} \ 1}$ matrix, did not increase the correlation between $\sigma'$ and $s'$ over the unmodified $U(5)$ matrix, but actually decreased it. The decrease in correlation is not difficult to explain based on the previous results of Figure 4. As more residues at the ends of helical sequences are considered to be "floppy", by expanding $U(m + 1)$, the correlation between $\sigma'$ and $s'$ increases. With modification 1, $U(5)^{\text{mod} \ 1}$, actually less residues at the ends of helical sequences are considered to be "floppy", and hence, the correlation between $\sigma'$ and $s'$ decreases. For this reason modification 2 and 3 were examined.

It is seen in both Figures 7 and 8, that modifications 2 and 3 increase the correlation between $\sigma'$ and $s'$ over the unmodified matrix treatment. In the $U(m + 1)^{\text{mod} \ 3}$ matrix all the statistical weights for residue $i$ in the helical state include an enthalpy term. Modification 3 is felt to be less accurate as a model for treating the $\alpha$-helix-to-coil transition than modification 2. Modification 3 is included only as an "upper limit" on the amount of correlation between $\sigma'$ and $s'$ which may be obtained from a given matrix size.

In comparing the curves of Figures 7 and 8 with the experimental data of Figure 3, all the curves appear to fit the data equally well. As mentioned before none of the curves fit the data past some greater
than zero. Changing the value of $\sigma$ used in $\mathcal{I}_0(m + 1)$ would not appear to influence these conclusions very much. As shown in Figure 5, the $-\ln \sigma'$ vs. $\ln s'$ curves would be shifted either up, decreasing $\sigma'$, or down, increasing $\sigma'$, in the region of $\ln s'$ less than zero. Changing the value of $\sigma$ does not appear to affect the curves in the region of $\ln s'$ greater than zero very much.

Lifson and Roig (63) have shown, in the case of $\mathcal{I}_0(3)$, that the transition point for large $N$, the value of $s$ which gives $\theta = 0.5$, is at $s = 1 + \sigma^{1/2}$, instead of $s = 1$ as in the Zimm-Bragg treatment (62). The difference could be traced to the different ways of assigning the statistical weights in $\mathcal{I}_0$. They plotted the eigenvalues of $\mathcal{I}_0(3)$ $\lambda(3)$, against $s$ for various values of $\sigma^{1/2}$. By so doing, they found that the transition occurred in the vicinity where the value of $\lambda_1$ and $\lambda_2$ deviated from their asymptotic values, in the limit $\sigma^{1/2} \to 0$, to an extend dependent on the value of $\sigma^{1/2}$. Since the solutions to the secular equation of matrix $\mathcal{I}_0(3)$ could not be obtained directly, the $\lambda$ vs. $s$ plots were calculated directly from the secular equation.

In a like manner, it can be shown for matrix $\mathcal{I}_0(4)$, that the transition point occurs at $s = 1 + \sigma^{1/3}$. The secular equation of matrix $\mathcal{I}_0(4)$, $\lambda = \lambda(4)$, is

$$
(\lambda - s) \left( \lambda^3 - \lambda^2 - \sigma^{1/3} \lambda - \sigma^{2/3} \right) - \sigma = 0 \quad (54a)
$$

Rearranging and solving for $s$ gives

$$
s = \lambda - \left[ \sigma / (\lambda^3 - \lambda^2 - \sigma^{1/3} \lambda - \sigma^{2/3}) \right] \quad (54b)
$$
In Figure 10, \( \lambda(4) \) is plotted from this equation, for two values of \( \sigma^{1/3} \), \( \sigma^{1/3} = 0.1 \) and \( \sigma^{1/3} = 0.01 \). The dashed lines indicate the asymptotic behavior of \( \lambda(4) \) in the limit \( \sigma^{1/3} \to 0 \). It can clearly be seen that the two largest eigenvalues, \( \lambda_1 \) and \( \lambda_2 \), deviate from their asymptotic values to an extent dependent on the value of \( \sigma^{1/3} \). Thus, the transition point is expected to be near \( s = 1 + \sigma^{1/3} \).

If this behavior holds for any of the \( \mathcal{U}(m + 1) \) matrices, then, in general, the transition point for matrix \( \mathcal{U}(m + 1) \) is at \( s = 1 + \sigma^{1/m} \). Figure 6 shows that the transition for matrix \( \mathcal{U}(5) \) occurs at \( \ln s = 0.14 \), or \( s = 1.15 \), and that for \( \mathcal{U}(10) \) occurs at \( \ln s = 0.37 \), or \( s = 1.45 \). Since in these curves \( \sigma = 0.0005 \), the expected transition points are calculated to be \( s = 1.15 \) and \( 1.43 \) for the \( \mathcal{U}(5) \) and \( \mathcal{U}(10) \) matrices, respectively. This is excellent agreement with the curves of Figure 6.

The peculiar asymptotic behavior observed in all of the \( - \ln \sigma \) vs. \( \ln s \) plots calculated may be explainable by looking at the transition curves of Figure 9. It can be seen in this figure that the \( \mathcal{U}(4) \) Mod 2 transition curve is always less than the Zimm-Bragg transition curve for a given value of \( \sigma \). Also, the \( \mathcal{U}(4) \) Mod 2 curves level off more quickly than does the Zimm-Bragg curves. The reason for the lower values and flatter curves of the \( \mathcal{U}(4) \) Mod 2 plot at the extremes of the transition is because, as stated earlier, only the residues in the interior of a helical sequence are counted as being helical. The ends of helical sequences are not "observed" as being

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 10. Eigenvalues, $\lambda(4)$, of the $\Psi(4)$ statistical weight matrix plotted against $s$ with $\sigma^{1/3} = 0.1$ (solid), 0.01 (dot-dashed), 0 (dashed).
helical. Therefore, the polymer will never be "observed" to be 100% helical. Also it will be "observed" to approach 0% helicity more rapidly, than a polymer whose ends of helical sequences can be "observed".

As can be seen from the two Zimm-Bragg curves at small values of $s$, $\ln s < 0$, a lower value of $\sigma$ causes the transition curve to attain lower values of $\theta$ more quickly, as observed from right to left, than the higher value of $\sigma$. It should be expected, therefore, that $U_{(4)}^{(4) Mod 2}$ would be fit more easily with modifications in $\sigma'$ than in $s'$. Also, small values of $\sigma'$ would be required to fit this flat region of the $U_{(4)}$ curve. This is exactly what is observed.

At the other end of the transition curve, large values of $s$, $\ln s > 0$, the higher values of $\sigma$ and lower values of $s$ in the Zimm-Bragg treatment causes a lower value of $\theta$. However, since the $U_{(4) Mod 2}$ curve levels off more rapidly than the Zimm-Bragg treatment, lower values of $\sigma'$ are required to fit the $U_{(4) Mod 2}$ curve. A compromise is met and the fitting becomes insensitive to both $\sigma'$ and $s'$. This insensitivity to $\sigma'$ can be seen also from Table 5 as large values of the percent deviation in $\sigma'$.

It appears at this time that no one simple theory for the helix-coil transition when fit to the Zimm-Bragg theory will reproduce the experimental correlation found between $\sigma$ and $s$ over the entire experimental range of $s$, $s$ between 0.6 and 1.35. However, as shown from these results, end effects are important in explaining the correlation in the region of $s$ between 0.6 and 1.2. The results of the
modifications to the \( \Psi(4) \) and \( \Psi(5) \) matrices may indicate the possible importance of near-range interactions in the correlation. Maxfield and Scheraga (134) found for charged side chain amino acids that the probability to be helical, larger \( s \) values, is enhanced greatly when an amino acid whose side chain is oppositely charged is four residues away. Also, if one looks at the experimentally determined \( s \) values for alanine, phenylalanine, leucine and glutamic acid at pH 2.3, their seems to be a correlation between the degree of hydrophobic character of the amino acid side chain and the value of \( s \). In general, the more hydrophobic a side chain is, the larger its \( s \) value will be. This seems to indicate the importance of hydrophobic interaction in the helix-coil transition.

Keeping these observations in mind, the \( \Psi(5) \) Mod 2 matrix treatment appears to be a better model for the helix-coil transition in polypeptides, than does the Zimm-Bragg treatment or the other expanded matrix treatments. However, as stated above, this model is still too simplified to account for the entire correlation found experimentally between \( \sigma \) and \( s \).
CHAPTER 3 POLY (\(\gamma\)-HYDROXY-L-PROLINE)

3.1 Introduction

As mentioned earlier in Chapter 1 glycine constitutes about one-third of the amino acid residues of all collagens whether vertebrate or invertebrate, independent of the species. The reason for this is collagen's unusual three left-handed helices, twisted right-handedly about each other. Glycine occupies the interior positions of this superhelix in which no other amino acid is allowed. Collagen is further characterized by its high pyrrolidine content (proline plus hydroxyproline), which varies from about 10 to 30% of the residues depending on the source.

The thermodynamic stability of the ordered collagen structure as measured by the melting temperature of fibers immersed in a large excess of water or of the individual solubilized molecules is dependent on the total pyrrolidine content. The melting or transformation temperature, \(T_m\), increases with increasing pyrrolidine content (135-138). It has been suggested (135) that the increasing melting temperature could be caused by a decrease in the entropy of fusion that would be expected to accompany the increased concentrations of total proline and hydroxyproline. The melting temperature will be dependent on the properties of both phases and consequently
on the nature of the disordered chain structure as well as the ordered one. The increasing concentration of pyrrolidine rings in the chain can be expected to suppress the configurational freedom of the molecule in the disordered state. A lower entropy of fusion would result if the crystalline state was unaffected, and consequently the melting temperature would increase. The configurational properties of the disordered state of a macromolecule can govern to a large extent its thermodynamic stability, relative to the ordered state (139).

Essentially all the hydroxyproline in tissues of vertebrates is found in collagen, and hydroxyproline accounts for about 10% of the total amino acid residues in most collagens. The collagen polypeptide chains can be well represented as repeating tripeptides of glycine-X-Y. Of the 12% proline content in rat and calf skin collagen, 97% was found to occupy position X (140). Also, of the approximate 11% hydroxyproline content in collagen from the same source, 99% was found to occupy the Y position (140). It is well known that proline residues are hydroxylated after polypeptide chain synthesis by an enzyme, called prolyl hydroxylase, which acts specifically at the Y position (141); hence, the uneven distribution of proline and hydroxyproline residues. To date, however, there is little information as to the functional role of hydroxyproline in collagen.

Hydroxylated proline was found to increase the thermal stability of collagen and collagen like structures (142-148). The melting temperature, $T_m$, of protocollagen, a non-hydroxylated form of collagen
(142), from chick embryonic tendon cells was found to be 15° C lower than the $T_m$ of an hydroxylated form of collagen from the same source (143). Also, the melting temperatures of sequential peptides (Pro-Hyp-Gly)$_x$, where $x = 5$ and 10, were found to be about 35° C higher than the non-hydroxylated sequential peptides (146-148). A third, independent series of observations indicating a role for hydroxy-proline came from thermal transition studies on cyanogen bromide cleavage fragments of collagen. The thermal transition of a cyanogen bromide cleavage fragment of collagen was higher than the thermal transition of a cyanogen bromide peptide which was identical except that it contained 1.8 residues less of hydroxy-proline (158). A suggestion for hydroxyproline's role in stabilizing the triple helix conformation was put forth by Traub (149) and Ramachandran et al. (150). They proposed an intra-chain water bridge involving the OH of hydroxyproline and the carbonyl oxygen of the polypeptide backbone.

In an attempt to understand the role of hydroxyproline in collagen, the study of synthetic homopolymers of both proline and hydroxyproline was undertaken. Poly ($\gamma$-hydroxy-L-proline) was first synthesized by Kurtz et al. (151, 152) through an O-acetyl-N-carboxyhydroxy-L-proline anhydride intermediate. Poly ($\gamma$-hydroxy-L-proline) is very water soluble, but unlike poly(L-proline), is not precipitated from aqueous solution by heating or by trichloroacetic acid (153-156). It is insoluble in glacial acetic acid, cold formic acid and dimethylformamide. It can be precipitated from aqueous solution by
dimethylformamide. The solubility behavior of Poly (γ-hydroxy-L-proline) appears to be determined by the presence of free hydrophilic hydroxyl groups and the absence of amide hydrogens (151). The solubility behavior of poly(L-proline) however, appears to be determined mainly by the absence of amide hydrogens, and hence, is soluble in more non-polar solvents and in aliphatic acids (156).

Poly (L-proline) was observed in the solid state to exist in two different ordered conformations (156), designated form I and form II. Form I contains a cis-peptide bond yielding a right-handed helix with a translation of 1.90 Å per residue and a rotation of 108° or 3.3 residues per turn (2). Form II contains a trans-peptide bond leading to the formation of a left-handed helix (41, 42). This left-handed helix has three residues per turn and has an axial translation of 3.12 Å per residue. Thus, it can be seen that form I is the more compact structure.

Poly (γ-hydroxy-L-proline) can also exist in two modifications in the solid state (44, 157), designated A and B. Unlike the two forms of poly(L-proline), only one of the structures is well defined. The X-ray pattern of poly (γ-hydroxy-L-proline) B exhibits only a few diffuse reflections, and so its structure has not been delineated. It could very well represent a disordered state. A vacuum ultraviolet absorption spectrum of an oriented poly (γ-hydroxy-L-proline) film has been interpreted (159) as evidence that the chain conformation in form B is related to that found in poly(L-proline) form I. Analysis of the more detailed X-ray patterns obtained from poly-(γ-hydroxy-L-proline) A reveals (44) that the individual molecular
chains are similar in conformation to that found in poly(L-proline) form II. Poly(γ-hydroxy-L-proline) A is a left-handed helix with an axial translation of 3.05 Å per residue and approximately three residues per turn. Unlike poly(L-proline) form II, however, there are three polymer chains, as opposed to only one for poly(L-proline), in the unit cell related by a threefold screw axis. These chains are "held together" by -OH --- O interchain hydrogen bonds. Hence, the ordered structure of poly (γ-hydroxy-L-proline) A requires three intertwined helical chains very similar to the general requirements of the ordered collagen structure.

The two forms of poly(L-proline) have also been observed in dilute solution (153, 160-167), and these forms are reversibly interconverted by appropriate changes in solvent (162,163, 167). In aqueous solutions of low ionic strength, poly(L-proline) exists in the trans configuration, form II. The addition of propanol or butanol to a water, acetic acid or formic acid—polymer solution induces a cooperative conformational transition to the more compact right-handed helical conformation containing all cis amide linkages (160), form I. Early evidence for the interconversion of form I to form II was noted by Kurtz et al. (168). They observed that poly-(L-proline) as obtained from its polymerizing medium of pyridine was only slightly water soluble and had a specific rotation, [α]$_{D}^{25}$, of 50°. This form of the polymer was later identified as form I. They also found that this form (form I) slowly mutarotates to a water-soluble highly levorotatory form (form II) when placed in
water-pyridine mixtures, aliphatic acids, benzyl alcohol, or chloroethanol.

Unlike poly(L-proline), poly (γ-hydroxy-L-proline) has not been observed to undergo an ordered to ordered transition. Poly (γ-hydroxy-L-proline) A undergoes a conformational change in neutral salt solutions (160, 174) and in formic acid with limited water content (169), but the resultant conformation does not appear to be ordered. The specific rotation, \([\alpha]_D^{25}\), of the polymer increased from about \(-400^\circ\) in water to \(-168^\circ\) in a 6 M lithium bromide solution (160) but only to \(-236^\circ\) in a 97 mole % formic acid solution (169). Similar changes in optical-rotation was observed for poly(L-proline) in a lithium bromide solution (153). The specific rotation \([\alpha]_D^{25}\) increased from approximately \(-530^\circ\) in water to about \(-250^\circ\) for a saturated salt solution.

The hydrodynamic (153, 154, 160 161, 170) and optical (153, 160, 161, 170-173) properties of poly(L-proline) form II in concentrated aqueous solutions of certain neutral salts, notably calcium chloride and lithium bromide, has been studied extensively. Poly(L-proline) form II, as well as poly (γ-hydroxy-L-proline) A, undergoes a conformational transition, as measured by intrinsic viscosity and circular dichroism, by the isothermal addition of calcium chloride (174). Poly (γ-hydroxy-L-proline) A appears to be more stable to disruption by calcium chloride than does poly(L-proline) form II, however. The intrinsic viscosity and optical rotation of poly(L-proline)
in aqueous salt solutions have values which are quite different from those exhibited by either of the helical forms of the polymer.

Mattice and Mandelkern have shown (170) that poly (L-proline) attains the properties of a statistical random coil at 2.5-4.0 M calcium chloride at 30°C. Because of the similarities between Poly(L-proline) and poly (γ-hydroxy-L-proline) in concentrated salt solutions, poly (γ-hydroxy-L-proline) is anticipated to be disordered under these conditions.

The disordering of the poly(L-proline) chain in concentrated aqueous salt solutions has been interpreted as arising from either (a) an increase in the accessible range of the C\(^\alpha\)–C\(^\gamma\) angle \(\psi\) (54, 160, 172, 173, 175, 176) or (b) the formation of random sequences of cis and trans peptide bonds (153, 177–180). In an attempt to ascertain which of these two possibilities is the more correct, Mattice and Mandelkern (177) estimated the upper limit for the characteristic ratio, \(\langle r^2 \rangle_o / n p^2\), see Chapter 1, of poly(L-proline) in 4.8 M calcium chloride at 30°C to be about 4.6. This ratio is substantially less than the characteristic ratio of approximately 9 determined for homopolypeptides with \(-\text{CH}_2\text{R}\) side chains (75). It is also at most only slightly greater than the characteristic ratio of 2.6±0.5 determined experimentally in water for the sequential copolypeptide (Pro-Gly)\(_x\) (79). It is difficult to explain how such a small \(\langle r^2 \rangle_o / n p^2\) could occur in poly(L-proline) if the peptide bonds are all required to maintain their planar trans conformation. However, if calcium chloride interacts with the peptide group in such a manner as to increase the rotational freedom about the peptide bonds, this could lead to a decrease in the apparent \(\langle r^2 \rangle_o / n p^2\) ratio.
bond (171, 181), or to promote cis-trans isomerization (153, 170, 181), a low $<r^2>_o/n_1^2$ might be attained.

The characteristic ratio of about 14 determined for poly(L-proline) in water at 30°C (177), although higher than the ratio of 9 for $-\text{CH}_2-R$ side chains (75), is also lower than the characteristic ratio predicted from conformational energy maps (45, 46, 49-52) based on rigid, $\psi$ fixed, pyrrolidine rings and trans-peptide bonds. For example, Schimmel and Flory (45) calculated a ratio of 116 for high molecular weight poly(L-proline). However, if the pyrrolidine ring is allowed to be flexible, $\psi$ flexible, than a characteristic ratio for poly(L-proline) with all trans-peptide bonds can be calculated which agrees with the experimentally determined ratio (54). The conformational energy maps used in these calculations yielded two widely separated minima in $\psi$ (53), of not too differing energies. The transition from one minimum to the other is of sufficiently high energy, so that, interconversion from one to the other would be slow on the NMR time scale, and would give rise to two distinct NMR peaks. The second minimum in these maps was calculated to be populated by about 8% at 30°C (54).

Tanaka and Scheraga (182), however, utilizing conformational energy maps which restrict $\psi$ to a narrow region (183), have shown by Monte-Carlo calculations that the random introduction of about 5% cis residues in the poly(L-proline) chain also leads to a characteristic ratio which agrees with experiment. Wu, Komoroski, and Mandelkern (184) have observed, using high resolution proton Fourier transform NMR techniques, that poly(L-proline) in $D_2O$ has about 2-3% of its peptides in the cis configuration. Their resonance peak assignment

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
for the C\textsuperscript{\alpha} cis proton was based on previous work by Torchia and Bovey (178). If this assignment is incorrect the isomer could possibly be assigned to the other low-energy state found by Nishikawa and Ooi (53) for the all-trans poly(L-proline) chain. Similar behavior may be expected for poly (γ-hydroxy-L-proline).

The hydrodynamic behavior of poly (γ-hydroxy-L-proline) as a function of molecular weight, as like poly(L-proline) (177), have not been obtained previously because of the unavailability of high molecular weight samples. Fractionation of a commercial preparation of poly (γ-hydroxy-L-proline) has provided nearly monodisperse samples with molecular weights ranging up to 35,000. In this chapter the hydrodynamic properties of poly (γ-hydroxy-L-proline) are examined in both water and calcium chloride solutions. Conformational energy maps based on rigid pyrrolidine rings are calculated, along with the predicted characteristic ratio from these maps. The implications for the conformational effects arising from the substitution of a γ-hydroxy-L-prolyl residue for an L-prolyl residue in a polypeptide chain are discussed.

3.2 Experimental

**Materials.** Two poly (γ-hydroxy-L-proline) samples were used. The lower molecular weight sample designated HP15, was used as supplied by Miles Laboratories. Gel permeation chromatography of the higher molecular weight sample (Sigma Chemical Co.) using a 5.0 x 100 cm K50/100 preparative Pharmacia chromatographic column packed with 80 cm of Sephadex G-100 produced three fractions, designated F1, F2, and
F3. Solutions were prepared by weight from polymer dried under vacuum using a Dry Ice - 2(2-butoxyethoxy) ethanol trap. Samples were recovered by lyophilization. The calcium chloride samples were recovered by dialyzation against deionized water using a Bio-Rad Bio-Fiber 50 beaker followed by lyophilization.

Calcium chloride was reagent grade. Solutions were prepared by dilution of a stock solution whose concentration was determined by titration with silver nitrate. Deionized water was used throughout.

**Osmometry.** Osmotic pressures were measured at 30°C in water using a Mechrolab 501 high-speed membrane osmometer equipped with an S & S B-20 membrane. The membrane was obtained from ArRo Laboratories, Inc., and was used as supplied. Osmotic pressures were measured by Dr. Wayne L. Mattice. However, their inclusion in the dissertation was felt to be vital to the discussion of the hydrodynamic properties of poly (γ-hydroxy-L-proline).

The number-average molecular weights, $M_N$, of samples F1, F2 and F3 were obtained by application of Equation 55 (186) to the reduced osmotic pressures. The osmotic pressure is $\pi$, $c$ is the concentration, $R$ is the gas constant, $T$ is the temperature, and $A_2$ is the second virial coefficient. Osmotic pressures were not measured for sample HP15 because of the anticipated difficulty with permeation of solute through the membrane for an unfractionated polymer of low weight-average molecular weight.

**Ultracentrifugation.** Weight-average molecular weights, $M_w$, were obtained by sedimentation equilibrium using a Beckman six-channel Yphantis equilibrium cell in a Beckman Model E analytical
ultracentrifuge, equipped with interference optics, utilizing Equation 56 (187,188). Here $c_T$ and $c_B$ are the equilibrium concentrations at the top and the bottom of the cell, respectively, and $M_{\text{App}}$ is the weight-average molecular weight evaluated from the slope, $(d \ln n / dr^2)$, of the logarithm of the fringe number vs. the square of the displacement from the center of rotation, assuming an ideal solution (189), from Equation 56b. $\bar{v}, \rho, \omega, R$ and $T$ have their usual meaning. A partial specific volume, $\bar{v}$, of 0.655 ± 0.010 cm$^3$/g was used for poly $(\gamma$-hydroxy-L-proline) in water (192). $c_T$ was evaluated from Equation 56c (190) where $c_o$ is the initial concentration, and $r_T$ and $r_B$ are the displacements from the center of rotation to the top and bottom of the cell, respectively.

In the derivation of Equation 56a it has been assumed that $A_2$ and $A_3$ are independent of the molecular weight distribution in the ultracentrifuge cell. Approximating $A_3$ by $A_2(2M/3)$ (187) leads to Equation 57, which was used for the determination of $M_w$ for all four samples.

\[
(M_{\text{App}})^{-1/2} = (M_w)^{-1/2} \left[ 1 + A_2 M_w \left( c_T + c_B \right) / 2 \right]
\]
Z-average molecular weights, \( M_z \), were also calculated from the sedimentation equilibrium data by Equation 58 (191) where

\[
(M_z^{\text{App}})^{-1} = (M_z)^{-1} \left[ 1 + A_2 M_z (c_T + c_B) \right]
\]  
(58a)

\( M_z^{\text{App}} \) is given by

\[
\frac{(d \ln c / dr^2)_B c_B - (d \ln c / dr^2)_T c_T}{c_B - c_T} = \frac{M_z^{\text{App}}(1 - \nu \rho) \omega^2}{2 RT}
\]  
(58b)

Here \( (d \ln c / dr^2)_T \) and \( (d \ln c / dr^2)_B \) are the slopes of the logarithm of the fringe number vs. the square of the displacement from the center of rotation at the top and bottom of the cell, respectively.

Sedimentation coefficients, \( s \), were obtained by sedimentation velocity using a Beckman double-sector interference cell in the analytical ultracentrifuge, equipped with Schlieren optics. Extrapolation to the infinite dilution sedimentation coefficient, \( s_0 \), was accomplished using Equation 59 (193). Here \( s \) was evaluated

\[
s^{-1} = s_0^{-1} + k c
\]  
(59a)

from Equation 57b (194) where \( (d \ln r / dt) \) is the slope of the logarithm of the displacement from the center of rotation to the schlieren peak vs. time in seconds.

Viscosities. Flow times were measured at 5°, 30°, and 55°C in water using a Cannon-Ubbelohde 50 semimicro dilution viscometer. Flow times in water ranged from 300 to 150 seconds. The flow times of poly (γ-hydroxy-L-proline) in calcium chloride solutions were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
measured at 30° using Cannon-Ubbelohde 100, 75, and 50 semimicro dilution viscometers. The flow times ranged from 700 to 300 seconds. Intrinsic viscosities, [η], were obtained from the extrapolation of the reduced viscosity, \( \eta_{sp}/c \), to infinite dilution by (195)

\[
\eta_{sp}/c = [\eta] + k[\eta]^2 c
\]

where

\[
\eta_{sp} = \frac{t_s - t_o}{t_o}
\]

\( t_s \) and \( t_o \) are the flow times of the solution and pure solvent, respectively.

3.3 Computations

Structure. The atoms considered in the computation of the conformational energy maps for the internal dipeptide unit in poly (γ-hydroxy-L-proline) are shown in Figure 11. The bond lengths and bond angles shown in the figure are from the solid state structure of poly (γ-hydroxy-L-proline) A determined by Sasishekaran (44). The O-H and C-H bond lengths were 1.00 Å and the C\(^{Y}\)-O-H angle was 97°. The angle between the C-N-C\(^{\alpha}\)-C\(^{\beta}\) and C\(^{\beta}\)-C\(^{Y}\)-C\(^{\delta}\) planes was, in the initial calculations, 17°, with the C\(^{Y}\) atom exo (196). This angle, designated \( \gamma \), was later varied from 8.5° to 45°. An exo atom is out of the plane of the pyrrolidine ring in the direction away from the carbonyl group (197). The rotational angle about the hydroxyl group is designated as \( \chi_5 \). Its zero position is a trans placement for atoms C\(^{\beta}\)-C\(^{Y}\)-O-H. A positive
Figure 11. Internal dipeptide geometry of poly (γ-hydroxy-L-proline) used in the conformational energy from reference 44.
rotation corresponds to a clockwise rotation as viewed down the 
Cγ—O bond. The conformational energy was computed at 5° increments 
of $\psi$ for $\chi_5 = 0, 120,$ and $240^\circ$. $\phi$ was held constant at $120^\circ$. The 
peptide bonds were maintained in the planar trans conformation.

Energy Functions. Generally, the intramolecular potential 
energy included contributions from nonbonded ($E_{\text{nb}}$) and electrostatic 
($E_{\text{el}}$) interactions, as well as from the energy ($E_\theta$) for rotation 
about single bonds. The potential energy for hydrogen-bonding 
($E_{\text{HB}}$) interactions was included in the later calculations. The 
total potential energy, $E_T$, is, thus, given by

$$E_T = E_{\text{nb}} + E_{\text{el}} + E_\theta + E_{\text{HB}} \tag{61}$$

The nonbonded interaction (Lennard-Jones 6-12 potentials) 
between a given pair of atoms is defined by Equation 62, where 
$r_{ij}^0$ (Å) is the distance between the two $i,j$ atoms, and $A_{ij}$ and $B_{ij}$ 
are characteristic constants. Values of $A_{ij}$ and $B_{ij}$ were the

$$E_{\text{nb}} = \sum_{ij} \left( -\frac{A_{ij}}{r_{ij}^0} + \frac{B_{ij}}{r_{ij}^{12}} \right) \tag{62}$$
same as those used by Ooi et al. (71), and are shown in Table 6.
The summation in Equation 62 is over all $i,j$ atom pairs considered.

The electrostatic potential energy was computed for pairwise 
interactions between the partial charges $q_i$ of atom $i$ and $q_j$ 
of atom $j$, using Equation 63, where $r_{ij}$ is defined as before, and 
$D$ represents the dielectric constant. Dielectric constants of 
$$E_{\text{el}} = \sum_{ij} 332 \left( \frac{q_i q_j}{D r_{ij}} \right) \tag{63}$$
4 and 2 were used, which are in the range commonly used for the 
interaction of static charges in polypeptides (199). The summation 
is over all appropriate atoms. The partial charges for the peptide
Table 6. Characteristic constants of nonbonded interaction (71)

<table>
<thead>
<tr>
<th>Atom i</th>
<th>Atom j</th>
<th>A_{ij}^{6} (Kcal Å^6 / mole)</th>
<th>B_{ij}^{12} (Kcal Å^12 / mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
<td>370</td>
<td>286000</td>
</tr>
<tr>
<td>C</td>
<td>H</td>
<td>128</td>
<td>38000</td>
</tr>
<tr>
<td>C</td>
<td>N</td>
<td>366</td>
<td>216000</td>
</tr>
<tr>
<td>C</td>
<td>O</td>
<td>367</td>
<td>205000</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>46.7</td>
<td>4500</td>
</tr>
<tr>
<td>H</td>
<td>N</td>
<td>125</td>
<td>27000</td>
</tr>
<tr>
<td>H</td>
<td>O</td>
<td>124</td>
<td>25000</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>363</td>
<td>161000</td>
</tr>
<tr>
<td>N</td>
<td>O</td>
<td>365</td>
<td>153000</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>367.1</td>
<td>145000</td>
</tr>
</tbody>
</table>

group were the same as those used by Brant et al. (67), and are 0.394 for the carbonyl carbon, -0.394 for the carbonyl oxygen, -0.281 for the nitrogen, and 0.281 for the δ carbon. The partial charges for the hydroxy group, when considered, were those by Poland and Scherage (198), and the charge on the C^¥ atom was chosen to yield a net charge of zero for the C^¥-O-H unit. These charges are -0.462 for the hydroxyl oxygen, 0.302 for the hydroxyl hydrogen, and 0.160 for the C^¥ atom. The factor of 332 in Equation 63 was included for converting the energy to Kcal/mole.

The potential energy for rotation about single bonds can be represented by the three fold rotational potential of Equation 64. θ is the rotational angle which was either φ, ψ, x₁, x₂, x₃, x₄, or x₅. φ, x₁, x₂, x₃, and x₄ are the rotational angles in the pyrrolidine
$E_0$ is the barrier height which had values of 0.2Kcal/mole for $\psi$ (71), 1.5Kcal/mole for $\phi$ (81), 3.0Kcal/mole for $x_1, x_2, x_3,$ and $x_4$ (199), and 1.0Kcal/mole for $x_5$ (198). The summation is over these rotational angles.

The potential energy for hydrogen bonding interactions, when considered, was accounted for by the empirical hydrogen bond function of McGuire, Momany and Scherage (200), given by Equation 65.

$$E_{HB} = \frac{B_k}{r_{o...H}} - A_k / r_{o...H}$$  (65)

Here $A_k$ and $B_k$ are given as 5781.2 and 13340.1, respectively. $r_{o...H}$ (Å) is the interatomic distance between the carbonyl oxygen and the hydroxyl hydrogen. When this hydrogen bonding function was used, the nonbonded interaction between these two atoms was neglected. This hydrogen bonding function has a minimum energy of -5.92Kcal/mole when the o...H distance is 1.66Å.

The method for calculation of intrachain atomic distances as a function of the rotational angles, in this case only $\psi_i$, in the peptide chain was the same as those used by Flory (201), Brant and Flory (81), and Nemethy and Scheraga (202). The procedure is outlined in Appendix C.

Conformational Properties. The averaged transformation matrix, $<T_{1}>$, was obtained from Equation 66 (203) using 5° intervals for $\psi_i$.

$$<T_{1}> = R(\zeta, 0) \frac{R(0^{\alpha}, n - \phi_i)}{R(-n, -\psi_i)}$$  (66a)

The N-C$^\alpha$-C$^\beta$ angle is $\alpha$, $n$ is the angle between the C$^\alpha$-C$^\beta$ bond and the virtual bond, the bond connecting C$^\alpha_i$ with C$^\alpha_{i+1}$, and $\zeta$ is the
angle between the C\(^\alpha\)-N bond and the virtual bond. Figure 12 shows these angles for the poly (γ-hydroxy-L-proline) chain. \(R(\zeta, 0)\), \(R(\theta, \pi - \phi_1)\), and \(R(-\eta_1 - \psi_1)\) are given by

\[
R(\zeta, 0) = \begin{bmatrix}
\cos \zeta & \sin \zeta & 0 \\
-sin \zeta & \cos \zeta & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[R(\theta, \pi - \phi_1) = \begin{bmatrix}
\cos \theta^\alpha & \sin \theta^\alpha & 0 \\
\sin \theta^\alpha \cos \phi_1 & -\cos \theta^\alpha \cos \phi_1 & \sin \phi_1 \\
\sin \theta^\alpha \sin \phi_1 & -\cos \theta^\alpha \sin \phi_1 & -\cos \phi_1
\end{bmatrix}
\]

\[R(-\eta_1 - \psi_1) = \begin{bmatrix}
\cos \eta & -\sin \eta & 0 \\
\sin \eta \cos \psi_1 & \cos \eta \cos \psi_1 & -\sin \psi_1 \\
\sin \eta \sin \psi_1 & \cos \eta \cos \psi_1 & \cos \psi_1
\end{bmatrix}
\]

(66b)  

(66c)  

(66d)  

Since \(\phi\) was held constant in the conformational energy calculations, only matrix \(R(-\eta, -\psi_1)\) need be averaged. This can be accomplished by utilizing Equation 67. The summations are taken over equal

\[
\langle R(-\eta, -\psi_1) \rangle = \frac{\sum_{\psi_1=0}^{2\pi} R(-\eta, -\psi_1) \exp \left[-E_T(\psi_1)/RT\right]}{2\pi} \sum_{\psi_1=0}^{2\pi} \exp \left[-E_T(\psi_1)/RT\right]
\]

(67)  

intervals for \(\psi_1\), in this case over 5° intervals. In the evaluation of \(\langle R(-\eta, -\psi_1) \rangle\), regions with energies above 10Kcal/mole above the minimum energy were ignored. Premultiplication of the quotient in Equation 67 by \(R(\theta^\alpha, \pi - \phi_1)\) and then by \(R(\zeta, 0)\) gives \(\langle T_1 \rangle\). The temperature dependence in \(\langle T_1 \rangle\) was assumed to arise entirely from the
Figure 12. The poly (γ-hydroxy-L-proline) chain with all amide bonds in the trans configuration.
change in the temperature term in the Boltzmann factors.

The unperturbed mean-square end-to-end distance, \( <r^2>_o \), for a chain containing \( n_p \) virtual bonds was computed from Equation 68 (204). Here \( \mathbb{E} \) is the identity matrix of order 3. The subscripts on the brackets denote the 1,1 element of the resultant 3 x 3 matrix.

\[
<r^2>_o = n_p \frac{1}{2} \left[ (\mathbb{E} + <T_1>) (\mathbb{E} - <T_1>)^{-1} - (2<\langle T_1 \rangle/n_p) (\mathbb{E} - <T_1>^p) (\mathbb{E} - <T_1>)^{-2} \right]_{11}
\]

(68)

The configurational entropy per residue, \( S \), was computed from Equation 69 (133) using 10° intervals for \( \psi_i \). The average energy,

\[
S = R \ln \sum_{\psi_i} \sum_{\psi_i} \exp \left[-E_T(\psi_i)/RT \right] + <E>/T 
\]

(69a)

\[
<E> = \frac{\sum_{\psi_i} E_T(\psi_i) \exp \left[-E_T(\psi_i)/RT \right]}{\sum_{\psi_i} \exp \left[-E_T(\psi_i)/RT \right]}. 
\]

(69b)

As before, regions with energies greater than 10Kcal/mole above the minimum energy were ignored.

3.4 Results
Hydrodynamic Properties

Molecular weights. Figure 13 shows the concentration dependence of the number-average molecular weights, \( M_n \), for samples F1, (circles), F2 (squares), and F3 (triangles). As mentioned before, osmotic pressures were not measured for sample HP15 because of the
Figure 13. Concentration dependence in the reduced osmotic pressures for the poly (γ-hydroxy-L-proline) samples F1 (circles), F2 (squares), and F3 (triangles) in water at 30°C.
anticipated difficulty with permeation of solute through the membrane. As might be expected, the points become more scattered the lower the molecular sample.

Typical plots of the sedimentation equilibrium data are shown in Figures 14 and 15 for sample HP15 at an initial concentration of 1.01 mg/ml (the worst), and for sample F2 at an initial concentration of 2.01 mg/ml (the best), respectively. No apparent curvature is observable in Figure 15 indicating a fairly monodisperse sample. However, Figure 14 does display some curvature which is not unexpected since this sample was not fractionated and, hence, should be more heterodisperse. Similar d ln c vs. $r^2$ plots for samples F1, and F3 were more like that of Figure 15. Figure 16 is a plot of $1/M_{App}^{1/2}$ vs. average concentration, c. Extrapolation to zero concentration yielded the weight-average molecular weights, $M_w$.

Crude estimates of the z-average molecular weights, $M_z^*$, were also obtained from the sedimentation equilibrium data utilizing Equation 58. The slopes of the logarithm of the fringe number vs. the square of the displacement from the center of rotation at the top and bottom of the cell, $(d \ln c/ dr^2)_T$ and $(d \ln c/ dr^2)_B$, respectively, were obtained from the first five (top) and last five (bottom) points from the $\ln c$ vs. $r^2$ plots. Table 7 is a summary of the various average molecular weights and their ratios for the four samples. As can be seen from the table all samples were fairly monodisperse. Sample HP15 had the greatest heterodispersity, as mentioned earlier.
Figure 14. The logarithm of the fringe number versus the square of the displacement from the center of rotation for the poly (γ-hydroxy-L-proline) sample HP15 at an initial concentration of 1.01 mg/ml (worst).
Figure 15. The logarithm of the fringe number versus the square of the displacement from the center of rotation for the poly (γ-hydroxy-L-proline) sample F2 at an initial concentration of 2.01 mg/ml (best).
Figure 16. Concentration dependence of the apparent weight-average molecular weights for the poly (γ-hydroxy-L-proline) samples, reading from top to bottom, HP15, F3, F2, and F1 as measured by equilibrium ultracentrifugation in water at 25°C.
Table 7. Molecular weights of the four poly (γ-hydroxy-L-proline) samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$10^{-3} M_N$</th>
<th>$10^{-3} M_w$</th>
<th>$10^{-3} M_z$</th>
<th>$M_w/M_N$</th>
<th>$M_z/M_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP15</td>
<td>———</td>
<td>8.94 ± 0.6</td>
<td>13.7 ± 3.3</td>
<td>———</td>
<td>1.53 ± 0.47</td>
</tr>
<tr>
<td>F3</td>
<td>15.6 ± 1.3</td>
<td>16.5 ± 0.9</td>
<td>17.2 ± 4.0</td>
<td>1.06 ± 0.14</td>
<td>1.04 ± 0.30</td>
</tr>
<tr>
<td>F2</td>
<td>20.4 ± 1.8</td>
<td>20.7 ± 1.2</td>
<td>23.5 ± 3.8</td>
<td>1.01 ± 0.14</td>
<td>1.14 ± 0.25</td>
</tr>
<tr>
<td>F1</td>
<td>35.0 ± 3.0</td>
<td>35.2 ± 3.0</td>
<td>38.9 ± 11.0</td>
<td>1.01 ± 0.18</td>
<td>1.11 ± 0.40</td>
</tr>
</tbody>
</table>

**Intrinsic viscosities.** The intrinsic viscosities in water were obtained at 5°, 30°, and 55°C. The intrinsic viscosities in calcium chloride solutions (1.95, 3.40, and 5.57M) were obtained at 30°C for samples F1, F2, and F3. The lowest molecular sample, HP15, had the same molecular weight as a sample whose calcium chloride intrinsic viscosity behavior was reported previously (174), therefore, this data was not repeated. Figures 17 and 18 show the concentration dependence of the reduced viscosities, $\eta_sp/c$, at 30°C for the four samples in $H_2O$ and for the three highest molecular weight samples in calcium chloride. The samples are F1 (circles), F2 (squares), F3 (triangles), and HP15 (diamonds). Figure 19 is a plot of logarithm of the intrinsic viscosity vs. temperature ($ln[\eta]$ vs. $T$) for the four samples in water. As can be seen from this figure, the temperature coefficients ($d\;ln\;[\eta]/dT$) for the intrinsic viscosities are large and negative. This is also shown in Table 8. The limiting temperature coefficient, obtained as the intercept of $d\;ln\;[\eta]/dT$ vs. $1/M_w$, is -0.005 deg.$^{-1}$. The coefficients are the same sign and magnitudes as found for poly(L-proline) in water (177), Table 8.
Figure 17. Concentration dependence of the reduced viscosities for the poly (γ-hydroxy-L-proline) samples F1 (circles), F2 (squares), and F3 (triangles) in water, and F1, and F2 in 1.95M calcium chloride at 30°C.
Figure 18. Concentration dependence of the reduced viscosities for the poly (γ-hydroxy-L-proline) samples HP15 (diamonds) in water, F3 (triangles) in 1.95M calcium chloride, F1 (circles), F2 (squares), F3 in 3.90M calcium chloride, and F1, F2, and F3 in 5.57M calcium chloride at 30°C. The uncertainties in the calcium chloride solutions are ± 0.02.
Figure 19. Temperature dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) in water. The samples are F1, F2, F3, and HP15, reading from top to bottom.
Equivalent treatment of previously reported (170,177) data for poly-(L-proline) also yields a large negative temperature coefficient, \(-0.008\) deg\(^{-1}\). In this respect poly(L-proline) and poly (\(\gamma\)-hydroxy-L-proline) are reminiscent of various derivatives of cellulose (205).

The dependence of \([\eta]\) on \(M_w\) at 30°C in water for the four poly-(\(\gamma\)-hydroxy-L-proline) samples (open circles), and for six poly(L-proline) samples (filled circles) reported earlier (177) is shown as a plot of logarithm \([\eta]\) vs. logarithm \(M_w\) (Log \([\eta]\) vs. Log \(M_w\)) in Figure 20. As can be seen, the poly (\(\gamma\)-hydroxy-L-proline) points follow the same general curvature as the poly(L-proline).

The nature of the molecular weight dependence of the temperature coefficients is such that the curvature apparent in Figure 20 will diminish as temperature increases. The best straight line through the poly (\(\gamma\)-hydroxy-L-proline) points representing samples F1, F2, and F3 has a slope of 0.68, which is well within the range anticipated for a random coil immersed in a good solvent (78).

<table>
<thead>
<tr>
<th>Sample</th>
<th>(10^{-3} M_w)</th>
<th>([\eta]^a)</th>
<th>(10^{3} d \ln [\eta]/dT^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP15</td>
<td>8.94</td>
<td>0.29</td>
<td>(-2.6 \pm 1)</td>
</tr>
<tr>
<td>F3</td>
<td>16.5</td>
<td>0.63</td>
<td>(-4.7 \pm 1)</td>
</tr>
<tr>
<td>F2</td>
<td>20.7</td>
<td>0.82</td>
<td>(-4.3 \pm 1)</td>
</tr>
<tr>
<td>F1</td>
<td>35.2</td>
<td>1.08</td>
<td>(-4.8 \pm 1)</td>
</tr>
<tr>
<td>poly(L-proline)</td>
<td>3.8 - 53</td>
<td>0.10 - 2.14</td>
<td>(-4 \rightarrow -8)</td>
</tr>
</tbody>
</table>

\(a.\) at 30°C, d/\(g.\)
\(b.\) over the temperature range 5°C - 55°C, deg\(^{-1}\).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 20. Molecular weight dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) (open circles) and poly-(L-proline) (small filled circles) in water at 30°C. Results for poly (L-proline) are from references 170 and 177.
The intrinsic viscosities experience a marked reduction in the presence of calcium chloride. This is shown in Figure 21. The lowest molecular weight sample data being taken from results reported previously (174). For comparative purposes, the dashed lines are the intrinsic viscosity data reported for two samples of poly-(L-proline) (170). The transitions of the poly(L-proline) samples occur at lower calcium chloride concentrations than does the transitions for poly (γ-hydroxy-L-proline), and the intrinsic viscosity of the former polymer shows little change above 3M calcium chloride.

The same data are plotted against the activity of calcium chloride in Figure 22. The activity was calculated from the molal activity coefficient (206) at 25° and the tabulated densities of aqueous calcium chloride (207). The density of 5.57M calcium chloride was estimated by a short extrapolation of the existing data. The change in the intrinsic viscosity of both polymers is most rapid at the lower activities of calcium chloride. However, poly(L-proline) has attained a limiting intrinsic viscosity at activities corresponding to a calcium chloride concentration of about 3M, while the change in the intrinsic viscosity of Poly (γ-hydroxy-L-proline) is still substantial at activities corresponding to a calcium chloride concentration of 4M. It does not appear that a limiting value of the intrinsic viscosity for poly (γ-hydroxy-L-proline) has been attained even in saturated calcium chloride.

Figure 23 shows the Log [n] vs. Log M₆ plots for the four concentrations of calcium chloride studied. The top curve is the
Figure 21. Intrinsic viscosities of the poly (γ-hydroxy-L-proline) samples F1 (circles), F2 (squares), F3 (triangles) and lowest molecular weight (diamonds) at 30°C as a function of calcium chloride concentration. The lowest molecular weight sample data was taken from reference 174. The dashed lines are the intrinsic viscosity data reported for two samples of poly (L-proline) (170).
Figure 22. Intrinsic viscosities of the poly (γ-hydroxy-L-proline) samples at 30°C as a function of calcium chloride activity. The samples are as described in Figure 21. The activity is that of calcium chloride at 25°C.
intrinsic viscosity data in water, and the bottom curve is the data in saturated calcium chloride. The curvature observed in the intrinsic viscosity data in water decreases in 1.95M calcium chloride and diminishes in 3.90M calcium chloride. Similar reductions in curvature of Log $[\eta]$ vs. Log $M_w$ plots were observed with poly-(L-proline) (170), but with lower calcium chloride concentrations.

The values of the slope, $a$, drawn through the four poly ($\gamma$-hydroxy-L-proline) samples, as a function of calcium chloride concentration are presented in Figure 24. The size of the symbols reflect the uncertainty in molecular weight and intrinsic viscosity. The slope, $a$, is seen to vary from 0.97 in pure water to 0.55 in saturated calcium chloride. The results presented in Figure 24 are qualitatively similar to the plots of intrinsic viscosity vs. calcium chloride concentration in Figure 21. In both cases the major change occurs at the lower concentrations of calcium chloride. Similar results (170) were obtained for poly(L-proline) samples ranging in weight-average molecular weights from 4,400 to 16,300. The value of the slope, $a$, was found to vary from 1.4 in pure water to 0.7 at 4.80M calcium chloride. The higher value of $a$ reported for poly(L-proline) in water may reflect the fact that lower molecular weight samples were used. In fact, $a$ for the three lower molecular weight samples of poly ($\gamma$-hydroxy-L-proline) in water has a value of 1.3.

**Sedimentation coefficients.** Figure 25 shows the plots of the logarithm of the displacement of the sedimenting boundary from the center of rotation vs. time (log $r_b$ vs. $t$) for samples $\text{#1}$ (the best),
Figure 23. Molecular weight dependence of the intrinsic viscosities of poly (γ-hydroxy-L-proline) at 30°C, reading from top to bottom, in water, 1.95M, 3.90M, and 5.57M calcium chloride.
Figure 24. Variation of the exponent of the molecular weight dependence of the intrinsic viscosity of poly (γ-hydroxy-L-proline) with calcium chloride concentration at 30°C. The exponent, $a$, is for the best straight lines through the samples in Figure 23.
Figure 25. The logarithm of the displacement of the sedimenting boundary from the center of rotation (cm) versus time (mins.) for the poly (γ-hydroxy-L-proline) samples F1 (best) at an initial concentration of 2.40 mg/ml, and HP15 at an initial concentration of 1.60 mg/ml at 20°C.
and HP15 (the worst). Both plots show approximately the same scattering in the points. Similar scattering was observed for the other samples. The concentration dependence of the sedimentation coefficients for the four samples in water at 20°C is shown in Figure 26. The resultant extrapolation to infinite dilution sedimentation coefficients, $s_0$, are collected in Table 9.

Table 9. Sedimentation and second virial coefficients for poly (γ-hydroxy-L-proline) samples in water

<table>
<thead>
<tr>
<th>Sample</th>
<th>$10^{-3} \bar{M}_w$</th>
<th>$10^{13} s_0$ $^a$</th>
<th>$10^{4} A_2$ $^b,c$</th>
<th>$10^{4} A_2$ $^b,d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP15</td>
<td>8.94</td>
<td>1.08 ± 0.10</td>
<td>--</td>
<td>10</td>
</tr>
<tr>
<td>F3</td>
<td>16.5</td>
<td>1.12 ± 0.03</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>F2</td>
<td>20.7</td>
<td>1.28 ± 0.06</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>F1</td>
<td>35.2</td>
<td>1.48 ± 0.07</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

a. at 20°C, sec$^{-1}$.

b. There is a large uncertainty associated with the second virical coefficients (see text). Units are cm$^3$mole/g$^2$.

c. from Equation 55 and the straight lines in Figure 13.

d. from Equation 57 and the straight lines in Figure 16.

Figure 27 shows the manner in which the sedimentation coefficients depend on molecular weight. Curvature is apparent, as was the case for the equivalent representation of the intrinsic viscosities in water. The best straight line through the error bars for poly (γ-hydroxy-L-proline) samples F1, F2, and F3 has a slope of about 0.38, which suggests that the polymer behaves as a random coil in
Figure 26. Concentration dependence of the sedimentation coefficients for the poly (γ-hydroxy-L-proline) samples HP15 (diamonds), F3 (triangles), F2 (squares), and Fl (circles) at 20°C in water.
a good solvent (78). The slope of the best straight line through the error bars for samples HP15, F3, and F2 has a value of about 0.17. This lower value of the slope reflects a rod-like behavior in a good solvent (78) over a substantial number of residues. An appreciable uncertainty must be assigned to the value of the slopes due to the uncertainties in the sedimentation coefficients and the narrow range of molecular weight covered by the fractions studied.

A poly(L-proline) sample with an estimated molecular weight of 18,000–19,000 has been reported (160) to have an $s_{20,w}$ of $0.83 \times 10^{13}$ sec. This result is about 70% as large as the sedimentation coefficient estimated from Table 9 for a poly(γ-hydroxy-L-proline) sample of the same molecular weight. The buoyancy factor, $1 - \bar{v}$, for poly(L-proline) is only about 70% of that for poly (γ-hydroxy-L-proline), accounting for the difference in the sedimentation coefficients of the two polypeptides.

Second virial coefficients. The data represented in Figures 13 and 16 are sufficiently precise to provide accurate intercepts and molecular weights, approximately 7% error. A considerably higher uncertainty, approximately 50%, exists in the slopes, and hence, there is a large uncertainty in the estimation of the second virial coefficients. The numbers presented in Table 9 correspond to the straight lines in Figures 13 and 16. Second virial coefficients are positive and tend to be large, on the order of $10^{-3}$ cm$^3$ mole/g$^2$. A large second
Figure 27. Molecular weight dependence of the sedimentation coefficients for poly (γ-hydroxy-L-proline) at 20°C in water.
virial coefficient was also obtained for a poly(L-proline) sample in water, with a $M_w$ of 99,000, of $(8.5 \pm 1) \times 10^{-4} \text{ cm}^3 \text{ mole/g}$.

**Error analysis.** The error in the number-average molecular weights, $M_N$, was estimated graphically from Figure 13 from the error in the individual points. The error in $M_w$ was calculated from the scattering in the $\ln c$ vs. $r^2$ plots for each $M_{w}^{App}$. The reciprocal of this error times the lowest error in each set of concentration data for a given sample was then used as a weighting factor for the given $M_{w}^{App}$ point. The error in $M_w$ was then calculated from a weighted least-squares routine, Equations B-10 and B-11, Appendix B. The error in $s_o$ was also calculated in a similar manner. The error in all parameters are reported as twice the unbiased estimate of the standard deviation, or approximately a 95% confidence interval.

**Conformational Calculations**

Representative conformational energy maps with the $C^\gamma$ atom exo by 17°, and $\chi_5 = 240^\circ$, are shown in Figure 28. The curves are, from top to bottom, using a dielectric constant = 4, hydroxyl electrostatic interactions deleted; dielectric constant = 4, hydroxyl electrostatic interactions included, and dielectric constant = 2, hydroxyl electrostatic interactions included. The energy scale for each curve was adjusted to yield a minimum energy of zero. Slightly lower energies at the minimum were obtained with $\chi_5 = 240^\circ$ than with $\chi_5 = 0^\circ$ or $120^\circ$, but the manner in which the conformational energy varies with $\psi_1$ was found to be nearly the same for
Figure 28. Conformational energy maps of the internal dipeptide unit in poly (γ-hydroxy-L-proline) shown in Figure 11 based on a pyrrolidine ring puckering angle of 17° and with $x_5 = 240°$. The energy scale for each curve is adjusted to yield a minimum of zero kcal/mol. The curves are from top to bottom, (a) to (c), as follows: (a) dielectric constant = 4, hydroxyl electrostatic interactions deleted; (b) dielectric constant = 4, hydroxyl electrostatic interactions included; (c) dielectric constant = 2, hydroxyl electrostatic interactions included.
each orientation of the hydroxyl group. The low-energy region occurs at $\psi_1 = 325 \pm 40^\circ$, with the minimum at $\psi_1 = 350^\circ$. These general features are similar to several conformational energy maps obtained at constant $\phi$ for poly(L-proline) (45-52) and poly (γ-hydroxy-L-proline) (208) containing planar \textit{trans} peptide bonds. Another low-energy region, nearly 20 Kcal/mol higher than the minimum at $\psi_1 = 350^\circ$, occurs near $\psi_1 = 130^\circ$.

Inclusion of the electrostatic interactions of the atoms in the hydroxyl group does not alter the conformational energy map if $\chi_5 = 0^\circ$ or $120^\circ$. The hydroxyl proton of residue $i + 1$ participates in a weak electrostatic interaction with the carbonyl oxygen atom of residue $i - 1$ if $\psi_1$ is near $290^\circ$ and $\chi_5 = 240^\circ$, resulting in a slight decrease in the conformational energy. Since the electrostatic energy is more negative when $\psi_1 \approx 290^\circ$ than when $\psi_1 = 350^\circ$, the difference in the conformational energy at these two points is sensitive to the choice of the dielectric constant. The difference in energy is only slightly greater than 1 Kcal/mol if the dielectric constant is two, $\chi_5 = 240^\circ$, and the electrostatic effects due to the hydroxyl group are included.

The predicted characteristic ratios are shown in Table 10. Deletion of the electrostatic interactions due to the hydroxyl group, and the use of a dielectric constant of four, yields a characteristic ratio of about 135 at 30°C. The characteristic ratio is diminished by interactions which decrease the difference in energy between the
conformations with $\psi_1 = 290^\circ$ and $350^\circ$. In all cases $d \ln <r^2_o>/dT$ is strongly negative.

Table 10. Dimensional properties at 30°C with the C$^\gamma$ atom exo by 17°.

<table>
<thead>
<tr>
<th>$\chi_5$ (deg)</th>
<th>Dielectric Constant</th>
<th>Hydroxyl Electrostatic Interaction Included</th>
<th>Characteristic Ratio$^a$</th>
<th>$d \ln &lt;r^2_o&gt;/dT \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>No</td>
<td>134</td>
<td>-6.5</td>
</tr>
<tr>
<td>120</td>
<td>4</td>
<td>No</td>
<td>139</td>
<td>-6.5</td>
</tr>
<tr>
<td>240</td>
<td>4</td>
<td>No</td>
<td>127</td>
<td>-6.5</td>
</tr>
<tr>
<td>240</td>
<td>4</td>
<td>Yes</td>
<td>113</td>
<td>-6.4</td>
</tr>
<tr>
<td>240</td>
<td>2</td>
<td>Yes</td>
<td>49</td>
<td>-3.6</td>
</tr>
</tbody>
</table>

$^a$ At a degree of polymerization of 1025.

The average transformation matrix, $<T_{i1}>$, at 30°C calculated from the conformational energy map for $\chi_5 = 240^\circ$, dielectric constant of two, inclusion of the hydroxy electrostatic interactions, and which yields a characteristic ratio of 49 at a degree of polymerization of 1025 is

$$<T_{i1}> = \begin{bmatrix}
0.51 & 0.67 & 0.47 \\
-0.61 & -0.08 & 0.75 \\
0.60 & -0.69 & 0.37
\end{bmatrix} \quad (70)$$

Mattice and Mandelkern (79) have suggested that an interaction of the hydroxyl group of the pyrrolidine ring of hydroxyproline and the peptide backbone may occur in poly (Hyp-Gly), arguing that the low-temperature poly (Pro) type CD observed for poly (Hyp-Gly), but not for poly (Pro-Gly), might be due to ordered structure resulting in...
from hydrogen bonding of a Hyp OH with the C = O moiety of the pre-
ceeding Hyp residue. Torchia (209) has tentatively suggested an
analogous OH --- O = C hydrogen bond for a poly (γ-hydroxy-L-proline)
chain in water based on NMR studies of the hydroxy-proline pyrrolidine
ring. As noted previously (79), models indicate that a strong OH ---
0 = C hydrogen bond is obtained only if (a) the hydroxy-proline ring
is significantly exo puckered at Cγ (χ2 > 30°) and (b) the residue
rotation angles assumes the values φ1 = 120° and ψ1 = 280°. The
hydroxyl oxygen and carbonyl oxygen distance is ca. + 0.3Å of
the optimum value (210), 2.7-2.8Å, when φ1, ψ1 are restricted at
φ1 = 120 ± 20°, ψ1 = 280 ± 10°. Torchia found that the Cγ atom was
exo in poly (γ-hydroxy-L-proline) and χ2 = 45 ± 10° in agreement with
the requirements for such a hydrogen bond. This hydrogen bonding
arrangement is illustrated in Figure 29.

To test the influence of an intrachain hydrogen bond on the
conformation of a poly (γ-hydroxy-L-proline) chain, conformational
energy maps were calculated, using a 10-12 hydrogen bond potential
function of Poland and Scherage (198), for various values of γ ,
the pyrrolidine ring puckering angle, and χ5°.

Figure 30 shows the conformational energy maps computed using a
dielectric constant of two for four exo puckering positions at the
Cγ atom. The maps shown are those which yielded the lowest positive
energy at ψ1 = 290° relative to the minimum at ψ1 = 350° when χ5°
was incremented by 10° between 240 and 320°. The values of χ5° for
the four maps are 280, 280, 290, and 270°, for γ = 8.5, 17, 25.5, and
Figure 29. Possible hydrogen bond pattern in the poly (γ-hydroxy-L-proline) chain.
Figure 30. Conformational energy maps of the internal dipeptide unit in poly (γ-hydroxy-L-proline) shown in Figure 11 based on pyrrolidine ring puckering angles of 8.5 (arrow, and going in the indicated direction), 17, 25.5, and 45°, and with $\chi_5 = 280, 280, 290$, and 270°, respectively.
45° respectively. The arrows in the figure point to the map for $\gamma = 8.5°$, and proceeding in the indicated direction gives the maps for $\gamma = 17, 25.5, \text{ and } 45°$, respectively. As before, the energy scale for each curve was adjusted to yield a minimum energy of zero. The general features of these maps are similar to those of Figure 28 calculated without the hydrogen bond potential.

Three general features from the maps can be seen going from low degree of puckering, $\gamma = 8.5°$, to high degree of puckering, $\gamma = 45°$.

First, the overall region of the maps slightly expand by about 10° in $\psi$. Second, the minimum at $\psi = 290°$ decreases from about 1 kcal/mol for $\gamma = 8.5°$ to about 0.1 kcal/mol for $\gamma = 45°$. And third, the relative maximum at about $\psi = 315°$ increases from about 2 kcal/mol for $\gamma = 8.5°$ to about 13 kcal/mol, not shown, for $\gamma = 45°$ causing the energy wells at the two minimum to become steeper. The affect of these three features can be seen from Figure 31, as plots of the characteristic ratio at 30°C (solid curve) and the temperature coefficient, (dashed curve), $d \ln \langle r^2 \rangle_0 / dT$, as a function of ring puckering angle and from Table 11. The characteristic ratio is seen to drop from about 55 for $\gamma = 8.5°$ to about 20 for $\gamma = 45°$. This is mainly due to the lowering of the difference in energy, $\Delta E$, between the two minimum at $\psi = 290$ and 350°, Table 11. The slight expansion in the energy maps when going from $\gamma = 8.5$ to $45°$ may also pay a slight roll in lowering the characteristic ratio.

The temperature coefficient, $d \ln \langle r^2 \rangle_0 / dT$, increase from about $-3 \times 10^{-3}$ deg$^{-1}$ for $\gamma = 8.5°$ to about zero for $\gamma = 45°$. The main
Figure 31. Characteristic ratios at 30°C (solid) and temperature coefficients (dashed) for poly (γ-hydroxy-L-proline) based on rigid pyrrolidine rings calculated from the conformational energy maps shown in Figure 30. Both the characteristic ratios and temperature coefficients are calculated at a degree of polymerization of 1025.
reason for the increase is again due to the lowering of the difference in energy between the two minima. The smaller the difference in energy between the two minima, the more equally populated the two conformational states become, and the less sensitive the two populations are to temperature. The steepness of the two minima plays only a minor role in determining the value of \( \frac{d \ln \langle r^2 \rangle_0}{dT} \).

The conformational energy was found to be relatively insensitive to rotation about the hydroxyl group for low puckering angles, as might be expected. However, as the puckering angle is increased to 45° the conformational energy at the second minimum \( \psi_1 = 290° \), was found to be very sensitive to the position of the hydroxyl proton. This can be seen from Table 11. The characteristic ratio varies from 354 to 20 to 385 when \( \chi_5 \) is rotated from 240 to 270 to 300°, respectively. The hydroxyl oxygen-carbonyl oxygen distance was found to vary from 4.05Å for \( \gamma = 8.5° \), and \( \psi_1 = 295° \) to 2.74Å for \( \gamma = 45° \), and \( \psi_1 = 285° \). In agreement with model studies (79), the puckering angle of 45°, or \( \chi_2 = 38° \), and \( \psi_1 = 285° \) yields a hydroxyl oxygen to carbonyl oxygen distance optimal for hydrogen bond formation (210). As seen from the table under H-Bond energy, the best hydrogen bond is formed when \( \chi_5 \) is about 300°. The hydroxyl proton of residue \( i + 1 \) is in line with the carbonyl group of residue \( i - 1 \) when \( \chi_5 \) is near 300°, and \( \psi_1 = 280° \).

The residue entropy, \( S \), is seen to vary from about 2.5 for \( \gamma = 45° \), and \( \chi_5 = 240° \), to 4 for \( \gamma = 8.5° \), and \( \chi_5 = 280° \), when \( \Delta E \) is positive, or when the minimum at \( \psi_1 = 350° \) is lower in energy than.
Table 11. Dimensional properties at 30°C with inclusion of a hydrogen bonding potential \(^{198\text{a}}\) between the hydroxyl proton of residue \(i+1\) and the carbonyl oxygen of residue \(i-1\).

<table>
<thead>
<tr>
<th>(\gamma)</th>
<th>(X_5)</th>
<th>H-Bond (^{b}) Energy Kcal/mol</th>
<th>H-Bond (^{b}) Energy Second Minimum</th>
<th>(\psi)</th>
<th>(O^•...H^b) A</th>
<th>(O^•...O^\gamma) A</th>
<th>(S) e.u.</th>
<th>Characteristic (^d)</th>
<th>(d \ln \langle r^2 \rangle_o / dT \times 10^3) deg(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5</td>
<td>280</td>
<td>-0.05</td>
<td>295</td>
<td>0.94</td>
<td>3.10</td>
<td>4.05</td>
<td>4.18</td>
<td>55</td>
<td>-2.4</td>
</tr>
<tr>
<td>17</td>
<td>280</td>
<td>-0.23</td>
<td>290</td>
<td>0.63</td>
<td>2.65</td>
<td>3.62</td>
<td>3.87</td>
<td>39</td>
<td>-1.6</td>
</tr>
<tr>
<td>25.5</td>
<td>290</td>
<td>-0.52</td>
<td>290</td>
<td>0.49</td>
<td>2.41</td>
<td>3.40</td>
<td>3.63</td>
<td>30</td>
<td>-1.2</td>
</tr>
<tr>
<td>45</td>
<td>270</td>
<td>-3.00</td>
<td>285</td>
<td>0.13</td>
<td>1.94</td>
<td>2.74</td>
<td>3.06</td>
<td>20</td>
<td>0.0</td>
</tr>
<tr>
<td>45</td>
<td>240</td>
<td>-0.57</td>
<td>285</td>
<td>3.75</td>
<td>2.39</td>
<td>2.74</td>
<td>2.49</td>
<td>354</td>
<td>-2.2</td>
</tr>
<tr>
<td>45</td>
<td>300</td>
<td>-5.35</td>
<td>285</td>
<td>-3.21</td>
<td>1.75</td>
<td>2.74</td>
<td>0.22</td>
<td>385</td>
<td>-7.2</td>
</tr>
</tbody>
</table>

a. see text.
b. at second minimum near \(\psi = 290^\circ\).
c. \(\Delta E = (\text{energy at second minimum near } \psi = 290^\circ) - (\text{energy at first minimum near } \psi = 350^\circ)\).
d. at a degree of polymerization of 1025.
the minimum at $\psi_1 = 290^\circ$. When $\Delta E$ becomes negative, or when $\gamma = 45^\circ$, and $\psi_1 = 300$, the residue entropy drops to about 0.2. This is due to the fact that the energy well at $\psi_1 = 290^\circ$ is much narrower than the energy well near $\psi_1 = 350^\circ$. The narrower energy well becomes populated to a greater extent than the broader well when $\Delta E$ becomes negative, and hence the configurational entropy decreases.

The average transformation matrix, $<T_1>$, at 30°C which yields the lowest characteristic ratio in Table 11 was calculated to be

$$<T_1> = \begin{bmatrix} 0.66 & 0.56 & 0.22 \\ -0.41 & 0.10 & 0.79 \\ 0.60 & -0.69 & 0.30 \end{bmatrix} \quad (71)$$

The manner in which this characteristic ratio at 30°C for $\gamma = 45^\circ$ and $\chi_5 = 270^\circ$ depends on the number of residues in the polymer chain, $n_p$, is shown in Figure 32 as a plot of $<r^2>/n_p^{1/2}$ vs. $n_p$ (solid), and $<r^2>/n_p^{1/2}$ vs. $1/n_p$ (dashed). The dashed curve is seen to become linear for $n_p > 50$, and yields a limiting value for the ratio at infinite chain length of about 20.5. Similar plots for the other ratios calculated had the same general curvature, but yielded respectively higher characteristic ratios at infinite chain length.

3.5 Discussion

Molecular weight dependence of the intrinsic viscosity and sedimentation coefficients. The intrinsic viscosity can be written in terms of the root-mean-square end-to-end distance, $<r^2>^{1/2}$, and the molecular weight, $M$, for random coiled polymers as (211)

$$[\eta] = \phi <r^2>^{3/2}/M \quad (72)$$
Figure 32. Characteristic ratio at 30°C for poly (γ-hydroxy-L-proline) based on a pyrrolidine ring puckering angle of 45°, and $\chi_5 = 270^\circ$ as a function of the degree of polymerization, $n_p$, (solid), and $1/n_p$ (dashed).
Here \( \Phi \) is a constant provided the polymer is of sufficiently high molecular weight, usually in excess of about 10,000. Alternatively, Equation 72 may be expressed as

\[
[\eta] = \phi \left( \frac{1}{M_o} \frac{<r^2>_p}{n_{p}^{1/2}} \right)^{3/2} \frac{1}{M^{1/2}} \alpha^3 \tag{73}
\]

where \( <r^2>_p \) is the unperturbed mean-square end-to-end distance, \( M_o \) is the residue molecular weight, \( n_p \) is the number of peptide bonds, \( \alpha \) is the expansion coefficient, and \( l_p \) is the distance between adjacent \( \alpha \) carbon atoms. In the case of random coil polymers, \( <r^2>_o / n_{p}^{1/2} \) approaches a constant value rapidly and is independent of molecular weight. Therefore, if the polymer is in a \( \theta \) solvent, \( \alpha = 1 \), then the viscosity is seen to be proportional to \( M^{1/2} \), or the log [\( \eta \)] vs. log \( M \) plot would have a slope, \( \alpha \), of 0.5. If the polymer is in a "good" solvent, \( \alpha > 1 \), then the intrinsic viscosity will be proportional to \( M^{1/2} \alpha^3 \). The manner in which \( \alpha \) depends on molecular weight can be expressed in general as (212)

\[
\alpha^5 - \alpha^3 = 2C_M \psi_1 M^{1/2} \tag{74}
\]

where \( C_M \) and \( \psi_1 \) are usually considered constants and independent of molecular weight. If the right hand side of Equation 74 was sufficiently large, \( \alpha^3 \) could be neglected compared with \( \alpha^5 \) and we should have \( \alpha \approx M^{1/10} \). Hence, the maximum value of \( \alpha \) could vary from 0.5 to 0.8 if the polymer behaves like a random coil in a good solvent.
The sedimentation coefficient, $s_o$, at infinite dilution is given by (78)

$$s_o = \frac{M(1-\bar{\nu}p)}{N_o f_o}$$

(75)

where $\bar{\nu}$, $p$, and $N_o$ have their usual meaning, and $f_o$ is the frictional coefficient. The frictional coefficient can be written as (213)

$$\frac{f_o}{\eta_o} = \frac{P}{\eta_o} \frac{\phi^{-1/3}}{n} (M [\eta])^{1/3}$$

(76)

where $\eta_o$ is the solvent viscosity and $P$ is a universal parameter which is the analog of $\phi$ of the viscosity treatment. Substituting Equation 76 into Equation 75 yields

$$s_o = \left[ \phi^{1/3} \frac{P^{-1}}{\eta_o} \frac{1-\bar{\nu}p}{N_o} \right] M^{2/3} [\eta]^{-1/3}.$$  (77)

Since everything within the brackets in Equation 77 are usually considered independent of molecular weight, and since we have just shown that for a random coil polymer in a good solvent $[\eta]$ is proportional to $M^{0.5 - 0.8}$, then $s_o$ will be proportional to $M^{0.4 - 0.5}$ under the same conditions.

As stated earlier, the results for the best values for the slopes of the log $[\eta]$ or log $[s_o]$ vs. log $M_w$ plots for samples F1, F2, and F3 are 0.68 and 0.38, respectively. These results compare favorably with 0.5 - 0.8 and 0.4 - 0.5, the ranges, as shown above, anticipated for random coil polymers immersed in a good solvent (78). In contrast, rigid helices would have $d \log [\eta]/d \log M = 1.8$ and $d \log s_o/d \log M = 0.2$ (195), results which are clearly incompatible with the data obtained using the three poly ($\gamma$-hydroxy-L-proline) fractions.
If the behavior predicted by the Kuhn-Houwink-Mark equation given by
\[ [\eta] = K M^a \] 
were obtained over the complete molecular weight range studied, the data in Figure 20 would describe a straight line. This equation frequently fails to predict the behavior of polymers at low molecular weight. The slope often decreases upon going to very low molecular weight for ordinary flexible polymers (214). Less flexible polymers may exhibit curvature in the other direction at low molecular weight. Cellulose obeys the relationship \([\eta] = 0.0177 n_p^{0.92}\) at 20°C in 50% sulfuric acid down to \(n_p = 150\) (215). This corresponds to the \(n_p\) of poly (\(\gamma\)-hydroxy-L-proline) F3. The slope increases at lower molecular weight with a maximum slope of 1.3 at \(n_p=15\) being estimated (215). Recall that this was also the value of \(a\) estimated for the best line through samples HP15, F3, and F2.

The factors which could contribute to a high slope in Equation 78 have been discussed with reference to cellulose trinitrate (216) and to poly(L-proline) (177). These are, see Equation 73, (a) a positive \(d \alpha/dM\), (b) a positive \(d(<r_o^2>/n_p^{12})/dM\) caused by the characteristic ratio not having attained its limiting value, and (c) a positive \(d \phi/dM\) caused by \(\phi\) not having attained its limiting value. The last factor could be due either to a hydrodynamic effect reflecting increasing permeation at the lower molecular weight or to deviations of the polymer configuration from random flight statistics and spherical symmetry which are required for the theory or both (216).

The effect of \(d \alpha/dM\) will be negligible for poly (\(\gamma\)-hydroxy-L-proline) because \(\alpha\) must be unity for the monomer and, like for
poly(L-proline), is only at most 1.1 for $M = 35,200$. The effect of $\frac{d(<r^2_o> / n_{p} l_{p}^2)}{dM}$ has been estimated for poly(L-proline) (177) from the molecular weight dependence of the characteristic ratio calculated from a square-well rotational potential function at $\phi = 102^\circ$ and $\psi = 310 \pm 55^\circ$, which satisfactorily reproduced the experimental characteristic ratio for SCCB, $M_w = 99,000$, in organic solvents at 30°C. Although some curvature in the log $[\eta]$ vs. log $M$ plot calculated using Equation 73 could be accounted for from $d(<r^2_o> / n_{p} l_{p}^2)/ dM$ over the molecular weight range studied, it was readily apparent that this was by no means the complete picture. In order to account for the observed $[\eta]$ of the poly(L-proline) sample with $M_w = 4,400$, $\phi$ would have to be about half the value used for the poly(L-proline) sample with $M_w = 99,000$. Similar conclusions can be reached for poly ($\gamma$-hydroxy-L-proline) based on the $d(<r^2_o> / n_{p} l_{p}^2)/ dM$ calculated from the conformational energy map which yielded the lowest characteristic ratio in Table 11. Although this map yields higher $[\eta]$ than the experimental data over the molecular weight range studied, $M_w = 8,940 - 35,200$, there was no observable curvature in the calculated log $[\eta]$ vs. log $M$ over this molecular weight range. This is in agreement with the calculated curve for poly(L-proline) in this molecular weight range (177). Also if the intrinsic viscosity calculated for sample F2 is adjusted to give the experimental value, then $\phi$ would have to be about 70% of the value used for F2 to reproduce the $[\eta]$ of sample HP15. This percentage reduction in $\phi$ is also required for the calculated $[\eta]$ of a poly(L-proline) sample,
M_\iota = 9,000, to agree with the experimental [\eta] of that molecular weight. The reduction in \phi was suggested in both cellulose trinitrate (216) and in Poly(L-proline) (177) to be caused by deviations from Gaussian statistics. Although a hydrodynamic effect (214) reflecting increased permeation at low molecular weight may also be operative.

The nature of the curvature in log \eta vs. log M_\omega at low molecular weight, Figure 27, is consistent with the curvature in log [\eta] vs. log M.

**Temperature coefficient.** As stated in the results, the limiting temperature coefficient of [\eta] for poly (\gamma-hydroxy-L-proline) is large and negative. This was also the case found for poly(L-proline) (177) and for cellulose derivatives (205). For both poly(L-proline) and the cellulose derivatives the large negative d ln [\eta]/dT could not be accounted for by changes in \alpha with temperature. The expansion coefficient for poly(L-proline) was found to only vary from 1.11 at 5°C to 1.09 at 30°C. Instead, the decrease in [\eta] with temperature must reflect an unusually rapid decrease in \frac{\langle r^2 \rangle_o}{n_p l_p^2} with temperature, a conclusion also previously deduced in the study of cellulose tributyrate (217). The rapid decrease in \frac{\langle r^2 \rangle_c}{n_p l_p^2} with temperature was ascribed to considerable modifications to the potentials restricting rotations about the other linkages in the cellulose derivatives. In poly(L-proline) the large negative d ln [\eta]/dT could be ascribed to arise from either (a) an increase in the allowable range in \psi, an increase in the population about \psi = 130° (54), or (b) \text{cis-trans} isomerization about the peptide bond.
In either case a lower $<r^2>_o / n_p \eta_p^2$ would result.

The temperature coefficients, $d \ln(<r^2>_o / n_p \eta_p^2) / dT$, calculated from the conformational maps based on rigid pyrrolidine rings described in the results were found to be sensitive to the difference in energy between the two minima at $\psi = 290^\circ$ and $350^\circ$, Tables 10 and 11, and Figure 31. If the temperature dependence of $\alpha$ is negligible, then $d \ln [\eta] / dT = 3/2 d \ln(<r^2>_o / n_p \eta_p^2) / dT$ for large molecular weight polymers. From the calculations it can be seen that the lower the characteristic ratio, then the smaller is $d \ln(<r^2>_o / n_p \eta_p^2) / dT$ in the absolute sense. If the characteristic ratio for poly (γ-hydroxy-L-proline) is large, then these maps can account for the observed temperature dependence of $[\eta]$. But as stated earlier the intrinsic viscosity calculated from Equation 73 using the lowest characteristic ratio in Table 11 was higher than the experimentally determined $[\eta]$. This seems to indicate that the conformational energy maps calculated for poly (γ-hydroxy-L-proline) with all trans peptide bonds based on rigid pyrrolidine rings cannot account for the observed hydrodynamic properties of this polymer.

**Combination of sedimentation coefficient and intrinsic viscosity.** A further test of the validity of the conclusions obtained from $d \log [\eta] / d \log M$ and $d \log s_o / d \log M$ is provided by combination of the sedimentation coefficient, intrinsic viscosity, and molecular weight according to Equation 79 (218, 219).

$$\beta = \frac{N_o s_o [\eta]^{1/3} \eta_o}{M^{2/3} (1 - \bar{v}_p)}$$

(79)
\( \eta_0 \) is the solvent viscosity in poise, and \( N_0, s_0, [\eta], M, \bar{v}, \) and \( \rho \) have their usual meaning. The \( \beta \)'s for the four poly (\( \gamma \)-hydroxy-L-proline) samples are presented in Figure 33. The error bars in the figure reflect the uncertainties in \( s_0, M, \) and \( \bar{v}. \) The \( \beta \)'s tend to decrease with increasing molecular weight, and the best straight line drawn through the four points has an intercept of 2.46. Clearly a line could just as easily be drawn through the error bars which intercepts the graph at 2.15 or 2.9. The results for the three fractions, F1, F2, and F3, are consistent with a molecular weight independent \( \beta \) of \((2.5 - 2.8) \times 10^6\), dashed lines. Values of \( \beta \)'s for six different polymer-solvent systems (polystyrene-methyl ethyl ketone, polystyrene-toluene, cellulose acetate-acetone, polysarcosine-water, polyisobutylene-cyclohexane, and poly(methylmethacrylate) - acetone) \((78)\) are in the range \((2.3 - 2.7) \times 10^6\). Consequently the results obtained with the poly (\( \gamma \)-hydroxy-L-proline) fractions are in the range obtained with random coil polymers. This conclusion is in harmony with that obtained from \( d \log [\eta]/d \log M \) and \( d \log s_0/d \log M. \)

It is of interest to compare the experimental \( \beta \) with those which should be obtained if the form A chain geometry were rigorously maintained in solution. The length of the helix would then be \( 3.12\AA \) per residue \((44)\). The diameter of the helix was chosen to be \( 10\AA \), a value obtained from a CPK model of the form A helix. The predicted \( \beta \)'s are obtained from the calculated axial ratios and the tables reported.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 33. Molecular weight dependence of $\beta$ (see text) for poly (γ-hydroxy-L-proline) at 20°C in water. The curved line is the $\beta$'s predicted for a rod-like molecule with an acid translation of 3.12 Å per residue (44) and a diameter of 10Å. The predicted $\beta$'s are obtained from the calculated axial ratios and the tables reported in reference 219 for a prolate ellipsoid.
in reference 219, and are shown as the curved line labeled ROD in Figure 33.

The β calculated for sample HP15, lowest molecular weight sample, is $2.72 \times 10^6$, which is consistent with the experimental result of $(2.95 \pm 0.37) \times 10^6$. The β's calculated for the solid state geometry increases with increasing molecular weight. This effect is not reproduced by the experimental β's. For the highest molecular weight sample, F1, the calculated value for form A is $3.21 \times 10^6$. This is much larger than the result obtained experimentally with F1.

The hydrodynamic properties of the poly (γ-hydroxy-L-proline) fractions, F1, F2, and F3, can be summarized as follows:

$\frac{d \log [\eta]}{d \log M}, \frac{d \log s}{d \log M}, d \beta/dM,$ and $\beta$ all exhibit the behavior observed with random coil polymers in good solvents, and they conflict with the values anticipated if the form A chain configuration were rigorously maintained in solution.

The frictional and diffusion coefficients for the various poly (γ-hydroxy-L-proline) samples can be calculated from the sedimentation coefficients and molecular weights utilizing Equations 75 and 80, respectively.

$$D_o = \frac{s_o \, RT}{M \,(1 - \gamma_p)}$$

The values for these coefficients along with their uncertainties for the four poly (γ-hydroxy-L-proline) samples studied are reported in Table 12. No new conclusions can be stated about these coefficients because they are calculated rather than measured quantities. They are presented here only as an aid to the reader for comparing these results with similar results in the literature.
Unperturbed dimensions. As mentioned in the general introduction, Chapter 1, the characteristic ratio for poly (γ-hydroxy-L-proline) can be calculated from the data assembled in Tables 8 and 9 using Equations 3-5 as has been done previously using data obtained with other random coil polypeptides in good solvents (75, 80, 82-84, 177). No consideration need be given to polydispersity for poly (γ-hydroxy-L-proline) samples F1, F2, and F3 because, as stated earlier in the results, their molecular weight distributions are narrow. The ratio of the z-average to the weight-average molecular weight for sample HP15 indicates a slight heterodispersity, but this sample too was treated as though it were monodisperse.

The expansion coefficients, \( \alpha \), obtained from the second virial coefficients of the osmotic pressure, OP, and sedimentation equilibrium, SE, data from Equation 5 are shown in Table 12. Fortunately, the large uncertainty in the second virial coefficient does not produce a similar large uncertainty in \( \alpha \). The expansion coefficients are slightly greater than unity. This was also the case for a poly(L-proline) sample of \( M_w = 99,000 \) (177). Characteristic ratios are shown in Table 12 and Figure 35. The dark circles in Figure 35 are the ratios calculated using the \( A_2 \) and \( M_w \) of the sedimentation equilibrium data. The triangles in the figure are the ratios calculated from the \( A_2 \) and \( M_N \) of the osmotic pressure data. The error bars reflect the estimated errors in \( A_2 \) and the molecular weight. The intercept of the best straight line drawn through the points yields an \( \langle r^2 \rangle_\infty / n \langle l_p^2 \rangle_\infty \) of 15.8 with an uncertainty of about 10%. For comparison, the result obtained (177) with poly(L-proline) in
Figure 34. Molecular weight dependence of the characteristic ratio of poly (γ-hydroxy-L-proline) in water at 30°C. Circles denote results obtained using $M_w$ and $A_2$ obtained from sedimentation equilibrium. Triangles denote results obtained using $M_N$ and the $A_2$ obtained using osmometry.
Table 12. Parameters calculated from the experimental results for poly (γ-hydroxy-L-proline).

<table>
<thead>
<tr>
<th>Sample</th>
<th>$10^{-3}M_w$</th>
<th>$10^{-6}c^a$</th>
<th>$10^7D^{a,b}_o$</th>
<th>$10^8f^{a,b}_o$</th>
<th>$\alpha^{d,e}$</th>
<th>$\langle r^2_o/n \rangle_{p^{1/2}}^{n_p}$</th>
<th>$\langle r^2_o \rangle_{p^{1/2}}^{1/3,12n_p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP15</td>
<td>8.94</td>
<td>2.93 ± 0.37</td>
<td>8.6 ± 1.1</td>
<td>4.7 ± 0.6</td>
<td>1.09</td>
<td>----</td>
<td>8.6 --</td>
</tr>
<tr>
<td>F3</td>
<td>16.5</td>
<td>2.62 ± 0.24</td>
<td>4.8 ± 0.4</td>
<td>8.4 ± 0.8</td>
<td>1.10</td>
<td>1.09</td>
<td>11.6  11.8</td>
</tr>
<tr>
<td>F2</td>
<td>20.7</td>
<td>2.82 ± 0.32</td>
<td>4.3 ± 0.5</td>
<td>9.3 ± 1.1</td>
<td>1.04</td>
<td>1.02</td>
<td>12.5  14.7</td>
</tr>
<tr>
<td>F1</td>
<td>35.2</td>
<td>2.51 ± 0.33</td>
<td>3.0 ± 0.4</td>
<td>13.6 ± 1.8</td>
<td>1.10</td>
<td>1.10</td>
<td>12.9  13.9</td>
</tr>
</tbody>
</table>

a. at 20°C.
b. diffusion coefficient.
c. frictional coefficient.
d. using the $A_2$ and $M_w$ obtained from sedimentation equilibrium.
e. using the $A_2$ and $M_N$ obtained from osmotic pressure.
water at 30°C is 13.7 ± 0.9. While the characteristic ratio of poly (γ-hydroxy-L-proline) might be slightly larger than that of poly(L-proline) the difference does not exceed the experimental uncertainty in the measurements. This result provides confirmation for the prediction by Schimmel and Flory (220) that poly(L-proline) and poly (γ-hydroxy-L-proline) would have equivalent unperturbed dimensions. As stated in the introduction concerning poly(L-proline), the characteristic ratio for poly(L-proline), and hence, poly (γ-hydroxy-L-proline), exceeds those ratios of about 9 obtained for homopolypeptides bearing -CH₂R side chains in L configuration (75, 80), for various copoly­peptides (83, 84), and for proteins (82).

The ratio of $<r^2>^{1/2}_o$, obtained from the characteristic ratios shown in Table 12, to the end-to-end distance for a poly (γ-hydroxy-L-proline) A helix is a means of representing the deviation of the poly (γ-hydroxy-L-proline) chain conformation in solution from that adopted by form A in the solid state (44). The values of this ratio are collected in Table 12. All samples studies have a root-mean-square end-to-end distance which is less than half the value characteristic of the form A helix. This ratio decreases to one-fourth for F1, the sample with a molecular weight of 35,000. Similar results were obtained for poly(L-proline) (177). In fact, the $<r^2>^{1/2}_o$ of the poly(L-proline) sample with $M_w = 99,000$ was only one-fifth that of the end-to-end distance of a poly(L-proline) form II chain. With both polymers a high degree of coiling is indicative.
The conformational energy maps calculated above for rigid pyrrolidine rings clearly do not predict characteristic ratios, Tables 10 and 11, which are in the range of the observed ratio. Of the ratios calculated, the characteristic ratio of 20 calculated from the energy map based on a pyrrolidine ring puckering angle of 45° comes closest to the observed value of about 15. However, as mentioned above, \( \frac{d \ln \left( \frac{\langle r^2 \rangle_{o}}{1_n \text{p}^2} \right)}{dT} \) is zero for this map, which is incompatible with experimental results. Similar conformational energy maps for poly(L-proline) and poly (\( \gamma \)-hydroxy-L-proline) (177, 220, 221) based on rigid pyrrolidine rings also gave larger characteristic ratios than the observed ratios for either polymer. It therefore appears that poly (\( \gamma \)-hydroxy-L-proline) as well as poly(L-proline) possesses a source of flexibility which is not adequately represented in the computations leading to the conformational energy maps based on rigid pyrrolidine rings.

Proton NMR measurements in aqueous solution demonstrate that the pyrrolidine ring in poly(L-proline) interconverts rapidly between two puckered conformations (197), while a single conformation is favored in poly (\( \gamma \)-hydroxy-L-proline) (209). In both cases the pyrrolidine ring was puckered at the \( \text{C}^\gamma \) atom. Pulse Fourier transform \(^{13}\text{C}\) NMR shows rapid ring motion in both homopolymers, with the pyrrolidine ring in poly(L-proline) being the more mobile (222). These observations have a direct bearing on the unperturbed dimensions because the
severity of the steric interaction between hydrogens on adjacent pyrrolidine rings, and hence the potential function for $\psi$, depend on the conformation adopted by the pyrrolidine rings (54-56). As stated in the introduction flexibility in the pyrrolidine ring yields conformational energy maps for poly(L-proline) which predict characteristic ratios (54) compatible with those determined experimentally for this polymer. Also these maps yield temperature coefficients which are large and negative. Similar conformational energy maps based on flexible pyrrolidine rings for poly ($\gamma$-hydroxy-L-proline) also predict (56) characteristic ratios and $d \ln(\langle r^2 \rangle_o /d n \langle L^2 \rangle_p) /dT$ which agree with the experimentally determined properties reported here.

An alternative source of the added flexibility demanded by the experimentally determined characteristic ratios of both poly(L-proline) and poly ($\gamma$-hydroxy-L-proline) is rotation about the peptide bond. The effect of a small number of cis peptide bonds in poly(L-proline) has already been discussed in the introduction. Experimental work based on proton Fourier transform NMR (184) indicated that 2-3% of the peptide bonds adopt the cis conformation. Monte Carlo calculations (183) showed that characteristic ratios in the range found experimentally could be computed based on a small incorporation, about 5%, of cis peptide bonds in a poly(L-proline) chain.

It is tempting to assert that the source of the added flexibility should be the same for both homopolypeptides because of the similarity in their structures, characteristic ratios, and
temperature coefficients. There is no convincing evidence for the existence of cis peptide bonds in poly (γ-hydroxy-L-proline), either in the solid state or in solution. It is possible that a few cis peptide bonds are present in poly (γ-hydroxy-L-proline) when dissolved in water and that they have simply escaped detection. Indeed, the region in the proton NMR spectrum, resonance at 4.3 ppm, normally attributed to the Cα cis proton is obscured by the Cγ proton resonance in aqueous poly (γ-hydroxy-L-proline) (223).

Effect of calcium chloride. Previous work (174) has shown that calcium chloride reduces the intrinsic viscosity and eliminates the positive 225 nm circular dichroism band in a poly (γ-hydroxy-L-proline) sample with Mw = 9,000. Similar effects, although at somewhat lower salt concentrations, occur with poly(L-proline) (174). The results reported here for poly (γ-hydroxy-L-proline), as with poly(L-proline) (170, 174), show that log [η] becomes more sensitive to calcium chloride as the molecular weight increases. As a consequence d log [η]/d log M for both polypeptides decreases in size as the calcium chloride concentration increases with poly(L-proline) being more sensitive to disruption by calcium chloride than poly (γ-hydroxy-L-proline). Proton (178) and 13C NMR (180) studies along with infra-red studies (179) on poly(L-proline) demonstrate that the induced collapse in the polymer dimensions caused by concentrated solutions of potassium iodide, sodium thiocyanate, calcium chloride, and lithium bromide arises from a salt-induced isomerization of
randomly located peptide bonds to the cis conformation. The similarity in the effects of calcium chloride on the two polypeptides suggests that a poly (γ-hydroxy-L-proline) chain should also exist with an appreciable number of its peptide bonds in the cis conformation. However, experimental support for the presence of cis peptide bonds in poly (γ-hydroxy-L-proline) is lacking. The $^{13}$C NMR spectrum of poly (γ-hydroxy-L-proline) in 6 M lithium bromide has been reported to show the existence of only one dominant isomer (224).

**Implications for the stability of the collagen helix.** As mentioned in the introduction, recent work on synthetic sequential copolypeptides (146) and modified collagens (143, 158, 225) has shown that the melting temperature, and hence the stability of the collagen helix, increases upon the substitution of γ-hydroxy-L-proline for L-proline. Also, the stereochemistry of the attachment of the hydroxyl group to the C$^γ$ atom also affects the stability of the helix (226). Two peptides containing allohydroxy-proline, (aHyp-Pro-Gly)$_{10}$ and (Pro-aHyp-Gly)$_{10}$, were synthesized. Examination of the peptides by optical rotation and circular dichroism showed that neither of the peptides formed triple-helical structures in aqueous solutions. Since the peptides had less tendency than (Pro-Hyp-Gly)$_{10}$ to become helical, the results demonstrated that the trans-γ-hydroxyl group of hydroxyproline makes a specific contribution to the stability of the triple helix. Also, since the peptides had less tendency than (Pro-Pro-Gly)$_{10}$ to become helical, then the cis-γ-hydroxyl group on allohydroxyproline decreases the stability of the triple helix.
One explanation offered for the hydroxyproline effect is the formation of an intrachain hydrogen bonded bridge from the hydroxyl group to a carbonyl oxygen atom via a water bridge (149, 150). Or, if the pyrrolidine ring in hydroxyproline is sufficiently puckered, then a hydrogen bond between the hydroxyl group and a carbonyl group in the same chain can be formed (79, 207). If strong hydrogen bonds of this type were present in collagen in an aqueous environment, they should also be found in aqueous poly (γ-hydroxy-L-proline). These hydrogen bonds would then serve to stiffen the poly (γ-hydroxy-L-proline) chain relative to poly(L-proline). As demonstrated by the conformational energy calculations above, a strong hydroxyl group and carbonyl group interaction can occur, supporting previous results, provided that the pyrrolidine ring at the Cγ atom is puckered greater than 25° and ψ is near 290°. However, the conformational properties predicted using these conformational energy maps are too large to account for the observed conformational properties of poly (γ-hydroxy-L-proline). Also, the importance of such hydrogen bonds in aqueous solutions must be questioned. The interaction of the hydroxyl group and carbonyl group with water molecules or ions may be at least as strong as the proposed hydrogen bonds. The observations of virtually identical characteristic ratios for poly (γ-hydroxy-L-proline) and poly(L-proline) also argues against the importance of such intra-chain hydrogen bonds.

An alternative suggestion for the hydroxyproline effect on the stability of the collagen helix has come from conformational energy...
calculations based on flexible pyrrolidine rings for poly (γ-hydroxy-L-proline) (56). The configurational entropy per residue was found to be slightly lower for poly (γ-hydroxy-L-proline) than for poly-(L-proline) based on similar flexible ring maps (54). These configurational entropy calculations are consistent with the observation that the pyrrolidine ring of poly(L-proline) is more mobile than the ring of poly (γ-hydroxy-L-proline) (222). A reduction in configurational entropy for the statistical coil when γ-hydroxy-L-proline is substituted for L-proline would increase the stability of the ordered structures formed by that polypeptide.

The hydroxyproline effect on the stability of the collagen helix appears therefore, to be a consequence of alterations in the flexibility of the pyrrolidine ring (56) or possibly has an inter-chain origin (226) or both.
REFERENCES

10. A. Wada, see Ref. 9, p. 369
13. W. B. Gatzer, see Ref. 9, p. 177
14. T. Miyazawa, see Ref. 9, p. 69

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
22. H. Benoit, L. Freund, G. Spach, see Ref. 9, p. 105
23. S. Beychok, see Ref. 9, p. 293
24. J. T. Yang, see Ref. 9, p. 239
26. G. D. Fasman, see Ref. 9, p. 499-604
27. E. M. Bradbury, B. G. Carpenter, and H. Goldman, Biopolymers, 6, 837 (1968).
34. T. M. Birshtein and O. B. Ptitsyn, Biopolymers, 5, 785 (1967).
59. G. D. Fasmon, see Ref. 9, 764 pp.


73. P. K. Ponnuswamy and V. Sasisekharan, Biopolymers, 10, 565 (1971).


91. G. W. Brady and R. Salovey, Biopolymers, 5, 331 (1967).


105. see Ref. 61, Appendix C, p. 316.


115. The LAPS hierarchy programs has been deposited as Document No. NAPS-01460 with the ASIS National Auxiliary Publication Service, c/o CCM Information Corp., 909 3rd Ave., New York, N. Y. 10022. A copy may be obtained by citing the document number and by remitting $2.00 for microfiche or $7.40 for photo copies. Advance payment is required. Make checks or money orders payable to: CCMIC-NAPS.


121. see Ref. 61, p. 171

122. see Ref. 61, p. 42


127. see Ref. 124, p. 339

128. see Ref. 124, p. 310


133. see Ref. 58, P. 274


159. J. L. Bensing and E. S. Pysh, Biopolymers, 10, 2645 (1971).
179. C. A. Swenson, Biopolymers, 10, 2591 (1971).

186. see Ref. 78, p. 531


201. see Ref. 58, Appendix B, p. 385


203. see Ref. 58, p. 275.

204. see Ref. 58, p. 25.


211. see Ref. 78, p. 611

212. see Ref. 78, p. 600

213. see Ref. 78, p. 627


223. L. Mandelkern, personnal communication.


APPENDIX A

Statistical weight Matrices $U(m + 1)$

4 x 4 Matrix (3 neighbors)

$m = 3$

$J^* = (0001)$

$J = \text{col}(0101)$

5 x 5 Matrix (4 neighbors)

$m = 4$

$J^* = (00001)$

$J = \text{col}(01001)$
\(6 \times 6\) Matrix (5 neighbors)

\[
\Psi(6) =
\begin{array}{cccccc}
\begin{array}{cccccc}
& i+1 & h & c & h & h & h & h & \text{hUc} \\
& i & h & h & h & h & c \\
\end{array} \\
\begin{array}{cccccc}
i-1 & h & \text{hUc} & h & c & h & h & \text{hUc} \\
i-2 & h & \text{hUc} & h & c & h & \text{hUc} \\
i-3 & h & \text{hUc} & c & h & \text{hUc} \\
i-4 & h & \text{hUc} & c & h & \text{hUc} \\
i-3 & h & \text{hUc} & c & h & \text{hUc} \\
n & s & 1/5 & o & 0 & 0 & 0 & 0 \\
\end{array}
\end{array}
\]

\(m = 5\)

\(J^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\)

\(J = \text{col} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}\)

\(7 \times 7\) Matrix (6 neighbors)

\[
\Psi(7) =
\begin{array}{cccccccc}
\begin{array}{cccccccc}
& i+1 & h & c & h & h & h & h & \text{hUc} \\
& i & h & h & h & h & h & c \\
\end{array} \\
\begin{array}{cccccccc}
i-1 & h & \text{hUc} & h & c & h & h & \text{hUc} \\
i-2 & h & \text{hUc} & h & c & h & \text{hUc} \\
i-3 & h & \text{hUc} & c & h & \text{hUc} \\
i-4 & h & \text{hUc} & c & h & \text{hUc} \\
\end{array}
\end{array}
\]

\(m = 6\)

\(J^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\)

\(J = \text{col} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\)
\begin{array}{cccccccc}
& & & & & & & \\
\text{8 x 8 Matrix (7 neighbors)} & & & & & & & \\
& i+1 & h & c & h & h & h & h & hUe \\
& i & h & h & h & h & h & h & c \\
& i-1 & h & hUe & h & h & c & hUe & c \\
& i-2 & h & hUe & h & h & c & hUe & c \\
& i-3 & h & hUe & h & c & hUe & c & hUe \\
& i-4 & h & hUe & h & c & hUe & c & hUe \\
& 1-5 & h & hUe & c & hUe & c & hUe & c & hUe \\
\end{array}

\begin{array}{cccccccc}
U(8) & i - 6 & i - 5 & i - 4 & i - 3 & i - 2 & i - 1 & i \\
\text{h} & h & h & h & h & h & h & s \\
\text{Ue} & hUe & hUe & hUe & hUe & h & c & o \\
\text{c} & h & h & h & h & h & h & \sigma^{1/7} \\
\text{Ue} & c & h & h & h & h & h & \sigma^{1/7} \\
\text{c} & hUe & c & h & h & h & h & \sigma^{1/7} \\
\text{Ue} & c & hUe & c & h & h & h & \sigma^{1/7} \\
\text{c} & hUe & c & hUe & c & h & h & \sigma^{1/7} \\
\text{Ue} & c & hUe & c & hUe & c & hUe & \sigma^{1/7} \\
\end{array}

m = 7

\begin{align*}
J^* &= (00000001) \\
J &= \text{col } (01000001)
\end{align*}
$9 \times 9$ Matrix (8 neighbors)

$$
\Psi(9) = \begin{array}{cccccccccc}
\text{i+1} & h & c & h & h & h & h & h & h & hUc \\
i & h & h & h & h & h & h & h & c \\
i-1 & h & hUc & h & h & h & h & c & hUc \\
i-2 & h & hUc & h & h & h & c & hUc & c \\
i-3 & h & hUc & h & h & h & c & hUc & c & hUc \\
i-4 & h & hUc & h & h & c & hUc & c & hUc & c \\
i-5 & h & hUc & h & c & hUc & c & hUc & c & hUc \\
\end{array}
$$

$$
\begin{array}{cccccccc}
h & h & h & h & h & h & s & 1/8 & o & o & o & o & o & o & o & o & 1 \\
hUc & hUc & hUc & hUc & hUc & hUc & c & o & o & o & o & o & o & o & o & 1 \\
c & h & h & h & h & h & o & 1/8 & 1/8 & o & o & o & o & o & o & o & 1 \\
hUc & c & h & h & h & h & o & 1/8 & 1/8 & o & o & o & o & o & o & o & 1 \\
c & hUc & c & h & h & h & h & o & 1/8 & 1/8 & o & o & o & o & o & o & 1 \\
hUc & c & hUc & c & h & h & h & o & 1/8 & 1/8 & o & o & o & o & o & 1 & 1 \\
c & hUc & c & hUc & c & h & h & h & o & 1/8 & 1/8 & o & o & o & o & 1 & 1 \\
\end{array}
$$

$m = 8$

$J^* = (000000001)$

$J = \text{col } (010000001)$
10 x 10 Matrix (9 neighbors)

<table>
<thead>
<tr>
<th>( i )</th>
<th>( h )</th>
<th>( c )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h )</th>
<th>( h \Ue)</th>
<th>( h \Uc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i+1 )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( i )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( i-1 )</td>
<td>( h \Uc)</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( h )</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( i-2 )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h )</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
<tr>
<td>( i-3 )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
<tr>
<td>( i-4 )</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( i-5 )</td>
<td>( h \Uc)</td>
<td>( h )</td>
<td>( h )</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( i-6 )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
</tr>
</tbody>
</table>

\( U(10) = \)

<table>
<thead>
<tr>
<th>( \sim )</th>
<th>( i-8 )</th>
<th>( i-7 )</th>
<th>( i-6 )</th>
<th>( i-5 )</th>
<th>( i-4 )</th>
<th>( i-3 )</th>
<th>( i-2 )</th>
<th>( i-1 )</th>
<th>( i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h )</td>
<td>( s )</td>
<td>( \sigma^{1/9} )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
</tr>
<tr>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
<td>( h \Uc)</td>
</tr>
<tr>
<td>( c )</td>
<td>( \sigma^{1/9} )</td>
<td>( \sigma^{1/9} )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
</tr>
<tr>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
<tr>
<td>( c )</td>
<td>( \sigma^{1/9} )</td>
<td>( \sigma^{1/9} )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
</tr>
<tr>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
<tr>
<td>( c )</td>
<td>( \sigma^{1/9} )</td>
<td>( \sigma^{1/9} )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
<td>( o )</td>
</tr>
<tr>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
<tr>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
<td>( h \Uc)</td>
<td>( c )</td>
</tr>
</tbody>
</table>
APPENDIX B

Least-Squares Fitting Technique

This procedure takes a model function $F$ and fits it to a set of $n$ pairs of values $(x_i^*, y_i^*)$, $i = 1, \ldots, n$.

B.1 General Fitting Method

The least-squares criterion seeks to choose values $p_1^*, \ldots, p_m^*$ for the parameters of the model function $F$ which minimize the weighted sum of squares

$$S = \sum_{i=1}^{n} w_i (F_i - y_i^*)^2 = \sum_{i=1}^{n} w_i \delta_i^2$$

B-1

The weighting factor $w_i$ is included to allow for non-equalivnet reliability in the observed $(x_i^*, y_i^*)$, $i=1, \ldots, n$ pairs. The condition that $S$ be a minimum requires

$$\left( \frac{\partial S}{\partial \tilde{p}_j} \right)_{\tilde{p} \neq \tilde{p}_j} = 0$$

B-2

for all $\tilde{p}_j$'s.

Generally the most useful way of finding $\{\tilde{p}\}$, the set of minimum parameter values satisfying Equation B-2, is the Gauss-Newton method (123). A general review of this procedure is given in Leach (124). In this procedure the model function $F$ is expanded as a Taylor's series and second- and higher-order terms are neglected. Thus, $F$ is written as

$$F_i = F_i^0 + \sum_{j=1}^{m} \left( \frac{\partial F_i}{\partial p_j} \right)^0 \Delta p_j$$

B-3
where \( F_i = F_i(P_1, ..., P_m) \), \( F_i^0 = F_i(P_1^0, ..., P_m^0) \), \( \{P^0\} \) is the set of initial guesses of the parameter values,

\[
(\partial F_i/\partial P_j)^0 = (\partial F_i/\partial P_j)(P_1^0, ..., P_m^0)
\]

and

\[
\Delta P_j = P_j - P_j^0.
\]

The minimized deviation \( \delta_i \) will be given by

\[
\delta_i = \delta_i^0 + \sum_{j=1}^{m} (\partial F_i/\partial P_j)^0 \Delta P_j
\]

The condition that minimizes \( S \) requires that

\[
(\partial S/\partial P_k) = \sum_{i=1}^{n} w_i \delta_i (\partial \delta_i/\partial P_k) = 0
\]

for \( k = 1, ..., m \).

Substituting Equation B-4 and approximating \( (\partial \delta_i/\partial P_k) = (\partial F_i/\partial P_k) \)

by \( (\partial F_i/\partial P_k)^0 \) into Equation B-5 gives

\[
\sum_{j=1}^{m} \Delta P_j \sum_{i=1}^{n} w_i (\partial F_i/\partial P_j)^0 (\partial F_i/\partial P_k)^0 = 0
\]

for \( k = 1, ..., m \) or

\[
C_{11} \Delta P_1 + C_{12} \Delta P_2 + \cdots + C_{2m} \Delta P_m = d_1
\]

\[
C_{21} \Delta P_1 + C_{22} \Delta P_2 + \cdots + C_{2m} \Delta P_m = d_2
\]

\[
\vdots
\]

\[
C_{m1} \Delta P_1 + C_{m2} \Delta P_2 + \cdots + C_{mm} \Delta P_m = d_m
\]

where

\[
C_{jk} = C_{kj} = \sum_{i=1}^{n} w_i (\partial F_i/\partial P_j)^0 (\partial F_i/\partial P_k)^0
\]

\[
C_{ij} = 0 \quad \text{for} \quad i > j
\]

\[
C_{ij} = 0 \quad \text{for} \quad i < j
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
and
\[ d_k = - \sum_{i=1}^{n} W_i \delta_i (\partial F_i / \partial P_k)^o \]  

Solving this set of \( m \) simultaneous linear equations yields a new set of minimum parameter values \( \{P^1\} \), \( P_j^1 = P_j^o + \Delta P_j^o \). These new \( P_j^1 \)'s are then used in the above equations in place of the \( P_j^o \)'s, and the cycle is repeated. This iterative process is continued until the computed parameter adjustments become negligible as measured by the weighted sum of squares \( S \), Equation B-1.

In some instances, e.g., when the initial parameter values are badly chosen, the successive \( S \)'s diverge. Or the convergence process is so slow that excessive computer time is required. In order to overcome these difficulties, damping and scaling factors have been introduced into the computing procedure.

Damping was achieved by multiplying the diagonal terms, \( \{C_{jj}\} \) in Equation B-6, by \((1 + D^2)\). In the program that follows damping was carried out only when \( S^r + 1 > S^r \). Here \( S^r \) and \( S^r + 1 \) are the sum of squares of the \( r \)th and the \( r + 1 \)th refinement, respectively. The value of \( D \) was chosen to be \( 2^v \), where \( v \) is an integer and is the \( v \)th successive iteration after \( r + 1 \) which causes \( S^r + 1 < S^r \).

Scaling was achieved by multiplying the calculated parameter adjustments, \( \Delta P_j^o \)'s, by a constant factor \( h \). The value of \( h = k^v \) was determined by finding a value of \( v \) such that
\[ S^r + 1 (k^v) < S^r + 1 (k^v - 1) \]  
and
\[ S^r + 1 (k^v) < S^r + 1 (k^v + 1) \]  
The value of \( k \) depends on the function to be fit, but in general has values ranging from 1.2 to 2.
B.2 Errors in Parameter Estimates

When the sum of squares, $S$, has been minimized, an unbiased estimate of the standard deviation is given (125) by

$$\hat{\sigma} = \left( \frac{S_{\text{min}}}{R} \right)^{1/2}$$

where

$$R = \frac{(n - m)}{n} \sum_{i=1}^{n} W_i$$

$S_{\text{min}}$ is the minimum value of $S$, and $m$ is the number of independent parameters in the fitting function $F$. Unbiased estimates of the standard deviation of the parameter values may be calculated from the relation

$$\hat{\sigma}_j = \hat{\sigma} \sqrt{(c^{-1})_{jj}}$$

where $(c^{-1})_{jj}$ is the $j$th, $j$th element of the inverse of the coefficient matrix, $(c_{jk})$, whose elements are defined by Equation B-7.

B.3 Computer Program

The main fitting program is written in subroutine form and has the name ALSTSQ. The coefficients $c_{jk}$ and $d_k$ of Equations B-7 and B-8 are calculated in subroutine COEF. Function DELF is used to calculate the $(\frac{\partial F_i}{\partial P_j})^0$'s in COEF. The model function to be fit is written in function subprogram form and is labeled F. After the coefficient matrix, $(c_{jk})$, has been set up the $P_j$'s of Equation B-6 are calculated by the Gauss-Seidel iterative method (126) for solving simultaneous equations in subroutine SIMEQ. The standard deviation given by Equation B-10 is calculated in function SIGMA. And the inverse of the coefficient matrix, $(c_{jk})$, is calculated in
APPENDIX C

Calculation of Intrachain Distances (81, 201, 202).

For the purpose of calculating the intrachain distances as functions of the rotational angles in the chain, let a right-handed Cartesian coordinate system be defined for each bond of the chain skeleton. As illustrated in Figure 1, the x-axis is in the direction of a given bond and the y-axis is in the plane defined by this bond and the preceding one. The positive direction of the y-axis is chosen by requiring it to make an acute angle with the preceding bond. All of the z-axes are perpendicular to the plane of the diagram with their directions alternating up and down from one coordinate system to the next. $\theta_i$ is the supplement to the fixed bond angle between two consecutive bonds. $\phi_i$ is the rotational angle about bond $i$.

Consider a vector $v_{i+1}$ with components $x_{i+1}$, $y_{i+1}$, $z_{i+1}$ in the coordinate system of bond $i+1$. Its components in the coordinate system of bond $i$, $x_i$, $y_i$, and $z_i$, are given by

\begin{align}
  x_i &= x_{i+1} \cos \theta_i + y_{i+1} \sin \theta_i + z_{i+1} \cos \phi_i \\
  y_i &= x_{i+1} \sin \theta_i \cos \phi_i - y_{i+1} \cos \theta_i \cos \phi_i + z_{i+1} \sin \phi_i \\
  z_i &= x_{i+1} \sin \theta_i \sin \phi_i - y_{i+1} \cos \theta_i \sin \phi_i - z_{i+1} \cos \phi_i
\end{align}
Figure 1. Cartesian coordinate system for consecutive bonds of a chain. The z-axes, not shown, are perpendicular to the plane, with the directions alternating up and down from one coordinate to the next.
The same result can be expressed in matrix notation as follows:

\[ v^i_{i+1} = T^i_1 v^i_{i+1} \]  

(C-4a)

where \( T^i_1 \) is the orthogonal matrix

\[
T^i_1 = \begin{bmatrix}
\cos \theta^i_1 & \sin \theta^i_1 & 0 \\
-\sin \theta^i_1 \cos \phi^i_1 & -\cos \theta^i_1 \cos \phi^i_1 & \sin \phi^i_1 \\
\sin \theta^i_1 \sin \phi^i_1 & -\cos \theta^i_1 \sin \phi^i_1 & -\cos \phi^i_1
\end{bmatrix}
\]

and \( v^i_{i+1} \) and \( v^i_{i+1} \) are expressed as column vectors:

\[ v^i_{i+1} = \text{col} \left( x^i_{i+1}, y^i_{i+1}, z^i_{i+1} \right) \]  

(C-4c)

\[ v^i_{i+1} = \text{col} \left( x^i_{i+1}, y^i_{i+1}, z^i_{i+1} \right) . \]  

(C-4d)

In a like manner, \( v^i_{i+1} \) may now be transformed to the coordinate system of bond \( i - 1 \) as

\[ v^i_{i+1} = T^{-1}_{i-1} v^i_{i+1} = T^{-1}_{i-1} T^i_1 v^i_{i+1}. \]  

(C-5)

\( T^{-1}_{i-1} \) is the transformation matrix for bond \( i - 1 \), and is given in a like manner by Equation (C-4b) with replace of \( \theta^i_1 \) and \( \phi^i_1 \) by \( \theta^i_{i-1} \) and \( \phi^i_{i-1} \), respectively.

Since it is the distance between any atom in the polymer chain that we are interested in, we must add the bond vector \( v_i \) to Equation (C-4a). In this way, not only do we transform the coordinate system of bond \( i + 1 \) to that of bond \( i \), but we also transform the origin of \( v^i_{i+1} \) to the origin of \( v^i_i \). Thus, the transformations of Equations C-4a and C-5 become

\[ v^i_{i+1} = T^i_1 v^i_{i+1} + v^i_i \]  

(C-6)
\[ \vec{r}_{i+1}^2 = T_i - 1 \vec{v}_{i+1} + \vec{v}_i - 1 = T_i - 1 T_i \vec{v}_i + 1 \]

\[ + T_i - 1 \vec{v}_i + \vec{v}_i - 1 \]

The distance, \( r_{i-1}, i+1 \), from atom \( i-1 \) and \( i+1 \), for example, may now be calculated by

\[ r_{i-1}, i+1 = (\vec{v}_{i-1})^T \vec{v}_{i+1} \]

where \((\vec{v}_{i-1})^T\) is a row vector and is the transpose of \(\vec{v}_{i+1}\).

The transformation matrix \(T_i\) may be formulated in more general terms by execution of two rotations: (a) a rotation of \(x_{i+1}, y_{i+1}, z_{i+1}\) about the \(z_{i+1}\) axis through the angle \(\tau\) that makes \(x_{i+1}\) coincident with (or parallel to) \(x_i\), and (b) another rotation of \(x_{i+1}, y_{i+1}, z_{i+1}\) through an angle \(\rho\) about the axis \(x_i\) that makes the two coordinate systems parallel. The angles \(\tau\) and \(\rho\) are measured in a right-handed sense. \(T_i\) can now be rewritten in terms of \(\tau\) and \(\rho\) as \(R_{i}\) given by

\[
R_i = \begin{bmatrix}
\cos \tau_i & \sin \tau_i & 0 \\
-sin \tau_i \cos \rho_i & \cos \tau_i \cos \rho_i & \sin \rho_i \\
\sin \tau_i \sin \rho_i & -\cos \tau_i \sin \rho_i & \cos \rho_i
\end{bmatrix}
\]

Using this method of vector transformations, the distance of any pair of atoms in the poly (\(\gamma\)-hydroxy-L-proline) chain can be calculated.
APPENDIX D

Glossary of Principal Symbols and Conventions

1. Mathematical Conventions

\( \langle \rangle \) The statistical mechanical average of the quantity enclosed, taken over all configurations of the chain.

\( \langle \rangle_0 \) The corresponding average for the chain unperturbed by interactions of long range, or by external constraints.

\( | | \) The absolute value of the quantity enclosed.

2. English Letter Symbols

\( \text{Å} \) Angstrom unit, \( 10^{-8} \) cm

\( A_2, A_3 \) Second and third virial coefficients

\( A_{ij}, B_{ij} \) Characteristic constants of the nonbonded interactions, Table 6, p. 77

\( c, c_0 \) Concentration (weight/volume) or coiled state in a polypeptide, p. 17; initial concentration.

\( d \) Distance (Å) traversed down a helical axis per amino acid residue.

\( D, D_0 \) Dielectric constant; diffusion coefficient, Eq. 80, p. 125.

\( E, E(\Phi_i, \psi_i), E_T, E_{nb}, E_{el}, E_0, E_{HB} \) Energy (per mole), energy associated with the configuration (\( \Phi_i, \psi_i \)); total conformational potential energy, Eq. 61, p. 76; nonbonded potential energy, Eq. 62, p. 76; electrostatic potential energy, Eq. 63, p. 76; torsional potential energy for rotation about Eq. 64, p. 78; potential energy for hydrogen bonding interactions, Eq. 65, p. 78.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
$f_A$, $f_o$  
Average fraction of A units in a copolymer of A and B units, Eq. 31, p. 26; frictional coefficient, Eq. 75, p. 118.

$F_A$  
Theoretical fraction of A units in a copolymer of A and B units, p. 28.

$G_{(h,c)}$  
Free energy (per mole) associated with the configuration $(h,c)$; free energy change per residue in going from a coil to helical state.

$\Delta G_{res}$  
Residue enthalpy change (per mole) in going from a coil to helical state.

$J, J^*$  
Column vector col (010...01) of order $m + 1$; row vector (00...01) of order $m + 1$.

$l_p$  
Virtual bond length, the distance from successive $\alpha$ carbon atoms.

$L(N)$  
Average number of amino acid residues per unbroken helical section, Eq. 13, p. 19.

$m, m'$  
Number of near-neighbor residues to residue $i$ considered in matrix $U_\nu (m+1)$, p. 33; LAPS hierarchy approximation, p. 30.

$M, M_\nu, M_{N}, M_z$  
Molecular weight (g/mol); number-average molecular weight; weight-average molecular weight; z-average molecular weight.

$n, n_p$  
Number of residues per turn in a helix, p. 4; number of residue in the polymer chain.

$N, N_o$  
Number of residues in the polymer chain; Avogadro's number.
\[ p \quad \text{Pitch of a helix (n x d)} \]

\[ q_i \quad \text{Partial charge on atom } i, \text{ p. 76} \]

\[ Q(N) \quad \text{Partition function for a chain molecule of length } N, \text{ Eqs. 1, 2, p. 10, 11.} \]

\[ r, r_b, r_T, r_{ij} \quad \text{Displacement (cm) from the center of rotation in ultracentrifugation; displacement to Schlieren boundary; displacement to bottom of centrifuge cell; displacement to top of centrifuge cell; interatomic distance (Å) from atom } i \text{ to atom } j. \]

\[ r, R \quad \text{Defined by Eq. 43; Eq. 44, p. 41, or gas constant.} \]

\[ <r^2>, <r^2>_o/n^1 \quad \text{Mean-square unperturbed end-to-end distance, characteristic ratio, p. 12.} \]

\[ s, s_o \quad \text{Zimm-Bragg helix-coil transition parameter, p. 14, or sedimentation coefficient, Eq. 59b; sedimentation coefficient at infinite dilution, Eq. 59a, p. 73.} \]

\[ s^{'}, s_{\text{min}} \quad \text{Fitted helix-coil transition parameter, p. 38; asymptotic limit in } s^{'}, \text{ p. 40.} \]

\[ S, \Delta S_{\text{res}} \quad \text{Configurational entropy per residue, Eq. 69a, p. 81; residue entropy change in going from a coiled to a helical state, p. 51.} \]

\[ t, t_o, t_s \quad \text{Time (sec); flow time of solvent; flow time of solution.} \]

\[ T, <T_i> \quad \text{Absolute temperature (°K); averaged transformation matrix for transforming the coordinate system of bond } i + 1 \text{ (virtual bond } i + 1) \text{ to that of bond } i \text{ (or virtual bond } i), \text{ Eq. 66a, p. 78.} \]
\[ u \] Lifson-Roig helix-coil transition parameter for coiled state, p. 18.

\[ U \] Matrix of statistical weights; expanded statistical weight matrix of order \( m + 1 \); modified expanded statistical weight matrix of modification \( i \).

\[ U^{(m+1)} \] Modified expanded statistical weight matrix of modification \( i \).

\[ v \] Lifson-Roig helix-coil transition parameter for residue \( i \) in the helical state and residue \( i - 1 \) in the coiled state, p. 18.

\[ V(N) \] Average number of unbroken helical sections, Eq. 12, p. 19.

\[ w \] Lifson-Roig helix-coil transition parameter for residue \( i \) in the helical state and residue \( i - 1 \) in the helical state, p. 18.

3. Greek Letter Symbols

\[ \alpha, \alpha' \] Expansion coefficient; ratio of the a priori probabilities of occurrence of A and B units in a random copolymer of A and B units, Eq. 29, p. 25.

\[ [\alpha]_D \] Specific rotation at the sodium D line at 25°C

\[ \beta \] Defined by Eq. 79, p. 122.

\[ \gamma \] Pyrrolidine ring puckering angle, p. 74.

\[ \xi, \eta, \epsilon \] Angle between the \( C^\alpha-N \) bond and the virtual bond; angle between the \( C^\sim-C^\alpha \) bond and the virtual bond; angle \( N-C^\alpha-C^\beta \), see Fig. 12, p. 80.

\[ \eta_o, \eta_{sp} \] Solvent viscosity; specific viscosity, Eq. 60b, p. 74; reduced viscosity; intrinsic viscosity, Eq. 60a, p. 74; intrinsic viscosity in a theta solvent.

\[ [n]_o \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\( \theta(N) \)  
Average fraction helical content, Eq. 11, p. 19.

\( \lambda, \lambda_{(m+1)} \)  
Eigenvalues of the statistical weight matrix \( U \); eigenvalues of the matrix \( U_{(m+1)} \).

\( \nu, \tilde{\nu} \)  
Coefficient of variation, Eq. 49, p. 50; partial specific volume (cm\(^3\)/g).

\( \varepsilon_i \)  
Free energy of a given residue \( i \) when the polymer is in a specific conformation.

\( \pi \)  
Osmotic pressure, Eq. 55, p. 71.

\( \phi \)  
Solvent density (g/cm\(^3\)).

\( \sigma, \sigma^\prime \)  
Zimm-Bragg helix-coil transition parameter for initiation of a helical sequence, fitted \( \sigma^\prime \), p. 38; asymptotic limit in \( \sigma^\prime \), p. 40; maximum value attained at large \( s^\prime \), p. 40.

\( \sigma_{\text{max}}, \sigma_{\text{max}}^\prime \)  
Rotational angle about \( N-C^\varepsilon \), Fig. 1, p. 3, or universal constant in viscosity theory, p. 12.

\( \chi_1, \chi_2 \)  
Rotational angles in the pyrrolidine ring of hydroxyproline or proline.

\( \chi_3, \chi_4 \)  
Rotational angle about the hydroxyl group in hydroxyproline.

\( \chi_5 \)  
Rotational angle about \( C^\varepsilon-C^\prime \), Fig. 1, p. 3.

\( \psi \)  
Rotational angle about the peptide bond, Fig. 1, p. 3, or angular velocity (rad/sec).
4. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>aHyp</td>
<td>All-Y-hydroxy-L-proline</td>
</tr>
<tr>
<td>Ala</td>
<td>L-Alanine</td>
</tr>
<tr>
<td>Glu.</td>
<td>L-Glutamic Acid</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>Hyp</td>
<td>Y-Hydroxy-L-Proline</td>
</tr>
<tr>
<td>Leu</td>
<td>L-Leucine</td>
</tr>
<tr>
<td>PHBG</td>
<td>Poly ( N^5 - (4^\prime\text{-hydroxybutyl}) - L\text{-glutamine} )</td>
</tr>
<tr>
<td>Phe</td>
<td>L-Phenylalanine</td>
</tr>
<tr>
<td>PHPG</td>
<td>Poly ( N^5 - (3^\prime\text{-hydroxypropyl}) - L\text{-glutamine} )</td>
</tr>
<tr>
<td>Pro</td>
<td>L-Proline</td>
</tr>
<tr>
<td>Ser</td>
<td>L-Serine</td>
</tr>
<tr>
<td>Val</td>
<td>L-Valine</td>
</tr>
</tbody>
</table>
APPENDIX E

Least Squares Fitting Program

SUBROUTINE ALSTSG(L,F)
IMPLICIT REAL A-H,0-2,Z,1
DIMENSION RA(45),XA(45),YE(45),C(6,6),D(6,6),SIG(2),XO(6),
TJ(N,N),XJ(N),ASIG(3)
DIMENSION SIG(6)
COMMON /TEMP1/XJ,N,DX,M2,N
COMMON /TEMP2/YE,RA,W
COMMON /TEMP3/C,D
COMMON /TEMP4/SCALE
EXTERNAL F
C YE = EXPERIMENTAL VALUES (20 MAX), FINAL YE(1)=STANDARD DEVIATION.
C YX(2)=CORRELATION COEFFICIENT BETWEEN XO AND YO
C W = WEIGHTS
C M2 = NUMBER OF EXPERIMENTAL POINTS
C XO,YO = INITIAL GUESS OF X,Y (PARAMETERS TO BE FITTED), ALSO FINAL
C FITTED PARAMETERS
C DELX,DELY = INITIAL SLOPE ADJUSTMENT PARAMETERS, ALSO FINAL
C SIGNIFICANCE OF FITTED PARAMETERS
C N = DEGREE OF POLYMERIZATION
C NOTE: XO,YO,DELX,DELY MUST BE VARIABLES IN MAIN
C L 1. IF L=0, NO PRINT, L=1, PPINT ONLY FITTED PARAMETERS
C L=2, PRINT ALL - EXPERIMENTAL AND REST
C C BUILDING COEFFICIENT MATRIX C AND CONSTANT VECTOR D
C L=1
L=1
J=1
DAMJ=1,
DJ=5 I=1,6
SIG(I)=0.0
GO TO 120
110 CONTINUE
L=1
J=2
120 CONTINUE
J=1
J=2
H=1.0
GO TO 10
CALL GUEF(CAMF)
C SAVING FOR NEW X AND Y
CALL 'SINGO(C*D,N,TI)
DO 100 I=1,N
K=I+1
AX(I+1)=X(I+1)
XU(I+1)=AX(I)+TT(K,1)*H
10 CONTINUE
AX(I+1)=AX(I)+TT(K,1)*H
110 CONTINUE
H=SCALE
J=J+1
DI I, I=1,N
F (I+1)=AX(I)+TT(K,1)*H
120 CONTINUE
AX(I+1)=SIGMA(XU,M2,F)
IF(JJ)+1.0,10,11
IF(ASIG(2)+GT,ASIG(1))GO TO 13
IF(ASIG(2)+LE,ASIG(J))GO TO 15
ASIG(1)=ASIG(3)
167

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
IF(J2.GT.99) GO TO 14
J2=J2+1
J1=1
GO TO 11
13 CONTINUE
H=H/(SCALE*SCALE)
DO 14 M=1,N
K=1+I
X(J1)=AX(I)*TT(K,1)*H
14 CONTINUE
SIG(J)=$SIG(J1)
GO TO 17
15 CONTINUE
SIG(J)=$SIG(2)
16 CONTINUE
IF(J.EQ.1)GO TO 110
SUN=(SIG(1)-SIG(2))/SIG(1)
ERROR=GAUSERROR
C TOLERANCE = 1.10**-06
IF(UPPER.LE.1.0D-07)GO TO 140
IF(SIG(2).LT.SIG(1))GO TO 130
L6=1
DAMP=DAMP*2.0
IF(L6.GT.20) GO TO 500
L6=L6+4
GO TO 120
130 CONTINUE
IF(L6.GT.99) GO TO 140
L6=L6+1
SIG(1)=SIG(2)
SIG(06)=SIG(1)
L6=L6+1
IF(L6.GT.6) L6=1
DAMP=1.
GO TO 110
140 CONTINUE
WRITE(6,6010)L5,SCALE
6010 FORMAT(1X,"NUMBER OF ITERATIONS = ",IS,5X," SCALE = ",F10.4)
WRITE(6,6031)(SIGI,1=1,6)
6031 FORMAT(1X,6,5X,6,I4)
SIG=SIGMA(XU,M2,N,F)
DAMP=1.0D0
CALL COEFF(DAMP,F)
CALL FATINV(C,N,N0)
DO 20 I=1,N
DX(I)=STAND*(USDOT(C(I,1))
20 CONTINUE
L6=L6+1
GO TO (110,160,170),L1
160 CONTINUE
WRITE(6,6001)STAND
6001 FORMAT(1X,"STANDARD DEVIATION = ",D15.8+)
DO 40 I=1,N
40 WRITE(6,6056)(XU(I),DX(I))
6056 FORMAT(1X,8X,"X(",I4,",*) = ",D15.8,)+/-",D15.8)
OILHELI PRECISESS FUNCTION F(X/R)

IMPLICIT REAL*A-H,O-Z,S

DIMENSION AL1(2),AL2(2),PF(2),PP(2),PG(2)

DIMENSION X(6)

SIG=X(1)
S=X(2)
N=R
AN=H
SIG=SINSIG
IF(SIG*GE.1.E-6)SIG=0.99
S=DABS(S)
AL1=(1.-S)*(1.-S)+4.*SIG*S
DET=1.0*H*(DET)
ALAMD1=(S+1.-DET)/2.
ALAMD2=(S+1.-DET)/2.
A1=AN*LOG10(ALAMD1)
L=A1
B1=(A1-L)
AL1(1)=10.**(B1)
AL1(2)=L
A1=AN*LOG10(ALAMD2)
L=A1
B1=(A1-L)
AL2(1)=10.**(B1)
AL2(2)=L
A1=(ALAMD1-1.)
A2=(1.-ALAMD2)
PF(1)=AL2*ALAMD1*AL1(1)
PF(2)=AL1(2)
CALL XPDG(IF(1),PF(2))
PF(1)=AL1*ALAMD2*AL2(1)
PF(2)=AL2(2)
CALL XPDG(IF(1),PF(2))
CALL CALD(PF,PP)
DDET=4.*SIG-2.*2.*S
ROE=DET/(2.*DET)
DALAM1=(1.+40.)/2.
DALAM2=(1.-40.)/2.
DL1DL2=(UU/I(2.*0.)/DET)
I1=(AN+1.)*A1*DLAM1-DALAM2*DLAM1
PI(1)=ALP(1)+R1
PI(2)=ALP(1+)
CALL XPH(PI(1),PP(2))
I1=DALAM1+DLAM2*(AN+1.)*A1*DALAM2
PI(1)=ALP(1)+R1
PI(2)=AL2(2)
CALL XPG(PI(1),PG(2))
CALL FAUD(PI,PG)
SN=SAN
PG(1)=(5N*PP(1))/PF(1)
PF(2)=PP(2)/PF(2)
AV=PG(1)*10.*PP(2))
AV=AV-5N*DL1DL2
F=AV
RETURN
END

SUBROUTINE XPG(PI,PG)
IMPLICIT REAL*8(A-H,O-Z,S)
IF(UM1S(A1),LT,1+0.3A1=0.D00)
I1=1.
IF(A1)<10.3A1=1.
J=10.
AT=DL1DL10(A1)
ATX=AT
I1=AT*ALXP
A1=AL1*(10.*PP(1))
A2=ALXP*AL1
U1=U
J=J.
CONTINUE.
K=0.
A2=2J
50 RETURN
END

SUBROUTINE FAUD (PI,B)
IMPLICIT REAL*8(A-H,O-Z,S)
IMPLICIT REAL*8(A-H,O-Z,S)
A = A + B
C=A(2)-B(2)
CS=A(2)-B(2)
IF(C*CS<30.)CS=-30.
II(5.A)=30.2A
10 CONTINUE
A(1)=B(1)*A(1)*(10.*CS)
A(2)=B(2)
G0 TO 40
20 CONTINUE
A(1)=A(1)*B(1)*(10.*CS)
G0 TO 40.
10 CONTINUE
A(1)=A(1)+B(1)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
DOUBLE PRECISION FUNCTION F1(X,R)
IMPLICIT REAL*8 (A-H,O-Z,S)
DIMENSION X(6)
F1=X(1)*R+X(2)
RETURN
END

DOUBLE PRECISION FUNCTION F2(X,R)
IMPLICIT REAL*8 (A-H,O-Z,S)
DIMENSION X(6)
F2=X(1)*R*R*X(2)+X(3)
RETURN
END

DOUBLE PRECISION FUNCTION F3(X,R)
IMPLICIT REAL*8 (A-H,O-Z,S)
DIMENSION X(6)
ALPHA=5.X(1)*R*R*X(2)
F3=X(1)*DEXP(ALPHA)
RETURN
END

DOUBLE PRECISION FUNCTION DELF (X0,AX,R,N,F)
IMPLICIT REAL*8 (A-H,O-Z,S)
C X0 = INITIAL VALUES FOR X
C AX = SCALLED VARIANCE IN X
C N = NUMBER OF PARAMETERS TO BE FITTED
DIMENSION X0(6),AX(6),X1(6),X2(6)
DIMENSION AM(2)
DO 5 K=1,N
X1(K)=X0(K)
X2(K)=0(K)
5 CONTINUE
IF(L=1.0-0.9)
J=1
DO 20 K=1,N
1=1
DX=DANS(AX(K))
IF(AX(1)=1.0+30)GO TO 10
GO TO 110
20 CONTINUE
1=2
110 CONTINUE
IF(AX(1)=0.9)DX=0.9
DEL X=X*X0(K)
X1(K)=1.0+DX
X2(K)=1.0-DX
X1(K)=X1(K)*X0(K)
X2(K)=X2(K)*X0(K)
DF=F(X1,R)-F(X2,R)
DX=2.*DF/LX
AM(I)=AM(I)+DX
DX=LX/2.0
IF(I=4+1)GO TO 120
ERR=ERR+DANS(AM(I)-AM(2))
IF (ERROR.LT.TOL) GO TO 130
IF (J.GT.20) GO TO 130
J=J+1
AM(1)=AM(2)
GO TO 120
10 CONTINUE
20 CONTINUE
130 CONTINUE
DEL=AM(2)
RETURN
END

SUBROUTINE COEF(DA,F)
IMPLICIT REAL*(A-H,O-Z,T-Z)
DIMENSION YE(45),R(45),W(45),C(6,6),D(6),XO(6),DX(6)
DIMENSION SUM(6,6),SUMD(6),DFDX(6),AX(6)
COMMON /TEMP1/XO,DX,M2,N
COMMON /TEMP2/YE,R,W
COMMON /TEMP3/C,U
EXTERNAL F
FOR=(1,C04DA)
IF (DA+LL.1.O0D0) FOR=1.00D0
DO 45 I=1,6
DO 50 J=1,6
SUM(I,J)=C.00D0
50 CONTINUE
AX(I)=0.0000
SUMH(I)=0.0000
DFDX(I)=C.00D0
45 CONTINUE
DU 1 DO I=1,M2
RP=H(I)
DU 60 K=1,N
AX(K)=DX(K)
DFDX(K)=DELF(XO,AX,RP,N,F)
AX(K)=C.00D0
60 CONTINUE
SUM(K,L)=SUM(K,L)+W(I)*DFDX(K)*DFDX(L)
70 CONTINUE
SUMD(K)=SUMD(K)+W(I)*DFDX(K)
80 CONTINUE
100 CONTINUE
DO 120 L=K,N
C(K,L)=SUMD(K)
120 CONTINUE
D(K)=-SUMD(K)
130 RETURN
END

DOUBLE PRECISION FUNCTION SIGMA(XO,M2,M,F)
IMPLICIT REAL*(A-H,O-Z,T-Z)
DIMENSION YE(45),RA(45),W(45),XO(6)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
COMMON / TEMP2/YE,RA,W
SUM=0.
SUMW=0.
DO 10 J=1,N2
H=HA(I)
DEL0=1/(X0(I)P)-YE(I)
SUMW=SUMW+W(I)
SUM5=SUM5+DEL0*DEL0*W(I)
100 CONTINUE
AH=P2-M
AM=M2
R=AH/AM
VAR=SUM5/(SUMW*R)
SIGMA=DSQRT(VAR)
RETURN
END

SUBROUTINE M E Q J ( A ,E.NV,NE)
IMPLICIT REAL*8 (A-H,0-Z,S)
DIMENSION A(6,6),E(6,6)
C A=A
DO 200 J=1,NE
DO 100 I=1,NV
BE(I,J)=A(I,J)
100 CONTINUE
200 CONTINUE
RETURN
END

SUBROUTINE M UL T I (A,NA,B,NNB,NEC,C)
IMPLICIT REAL*8 (A-H,0-Z,S)
DIMENSION A(6,6),B(6,6),C(6,6)
C C=A*B
IF(NA,NE,NNB)*0 TO 400
DO J=1,NA
DO 200 I=1,NEB
CSUM=0.
DO 100 K=1,NNB
CSUM=CSUM+A(I,K)*B(K,J)
100 CONTINUE
C(I,J)=CSUM
200 CONTINUE
300 CONTINUE
RETURN
400 CONTINUE
WRITE(6,600)
600 FORMAT(1H,5X,*MATRIX A AND B ARE NOT CONFORMABLE*)
RETURN
END

SUBROUTINE SIMO(C,D,T,TT)
IMPLICIT REAL*8 (A-H,0-Z,S)
C C = COEFFICIENT MATRIX OF ORDER N (SQUARE)
C D = CONSTANT VECTOR
C TT = SOLUTION MATRIX (SOLUTIONS IN FIRST COLUMN)
DIMENSION C(6,6),D(6),CI(6,6),TN(6,6),T(6,6),TT(6,6)
DO 10 I=1,6
DO 10 J=1,6

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CI(I,J)=0.

10 CONTINUE
C

C GENERATE IDENTITY MATRIX CI
IP=N+1
DO 100 I=1,N
100 CI(I,I)=1.
C
CALL MUX(CI,1,IP,IP)
C
ASSMLING MATRIX T
DO 200 I=1,N
CALL MUX(CI,TN,IP,IP)
K=I+1
DO 300 J=2,N
300 TN(K,J)=-CI(I,J-1)/CI(I,I)
TN(K,K)=0.
TN(K,1)=DI(I)/CI(I,1)
CALL MUX(TN,IP,IP,NP,IP,IP)
CALL MEG(TT,TN,IP,IP)
CONTINUE

C

CONTINUATIONS ARE ITERATED TO 2**NN
NN=8
DO 400 KM=1,NN
CALL MUX(T,TN,IP,IP,NP,IP)
C
USING A TOLERANCE OF 1.0E-05
SUM=0.
DO 500 K2=2,NP
IF(DAYS(T(K2,1),LT,1.0E-30)GO TO 500
SUM=SUM+DAYS((TT(K2,1)-T(K2,1))/T(K2,1))
500 CONTINUE
IF(SUM.LE.1.0E-10)GO TO 600
CALL MUX(T,TN,IP,IP)
600 CONTINUE
RETURN
END

SUBROUTINE MATINV(A,B,N,L)
IMPLICIT REAL*8(A-H,O-Z)
C
IF L=5 RETURNS INVERSE OF A IN A, IF L=1 SOLUTION OF AX=B IN B
DIMENSION A(6,6),B(6),IP(6),IN(6,2)
DU IF | =1.E
17 1P(I)=C
0=1.
DU 12 I=1,N
AMAX=0.
DU 3 J=1,N
IF(IP(J).GT.C)GO TO 3
IF(IP(J).LT.0)GO TO 4
DI I=1,N
IF(IP(K).GT.1)GO TO 2
IF(IP(K).LT.1)GO TO 4
IF(DABS(A(J,K)).LE.AMAX)GO TO 2
I=J
IC=K
AMAX=DABS(A(J,K))
2 CONTINUE
3 CONTINUE
IP(IC)=IP(IC)+1
IF(AMAX.GT.1.0E-60)GO TO 6
4 WRITE(6,5)
5 FORMAT(*16H SINGULAR MATRIX)
RETURN
6 IF(IR.EQ.IC)GO TO 8
D=0
DO 7 K=1,N
AMAX=A(K,IR)
A(IR,K)=A(IC,K)
7 A(IC,K)=AMAX
IF(L.UG.0)GO TO 8
AMAX=B(IR)
B(IR)=IC
IC=AMAX
8 IN(I,1)=IN
IN(1,2)=IC
AMAX=A(IC,IC)
D=0*AMAX
A(IC,IC)=1.
U(I) Y K=1,N
9 A(IC,K)=A(IC,K)/AMAX
IF(L.UG.0)GO TO 10
B(IC)=B(IC)/AMAX
10 DU 12 J=1,N
IF(J.UG.IC)GO TO 12
AMAX=A(J,IC)
A(J,IC)=0.
D(J) K=1,N
11 A(J,K)=A(J,K)-A(IC,K)*AMAX
IF(L.UG.0)GO TO 12
B(J)=B(J)-B(IC)*AMAX
12 CONTINUE
IF(L.UG.1)RETURN
DU 14 I=1,N
J=IN(I,1)
IF(IN(J,1).NE.IN(J,2))GO TO 14
I3=IN(J,1)
IC=IN(J,2)
DO 13 K=1,N
AMAX=A(K,IR)
A(K,IN)=A(K,IC)
13 A(K,IC)=AMAX
14 CONTINUE
RETURN
END
VITA

Name: Donald Stephen Clark

Born: November 29, 1949, at New Orleans, Louisiana.

Marital Status: Single

Formal Education:

B.S., 1972, Louisiana State University. Major: Biochemistry; Minor: Chemistry.


Publications:


EXAMINATION AND THESIS REPORT

Candidate: Donald Stephen Clark

Major Field: Biochemistry

Title of Thesis: Conformational Properties of Polypeptides: Helix-Coil Transition, and the Conformational and Hydrodynamic Properties of Poly (gamma-Hydroxy-L-Proline)

Approved:

[Signatures]

Wayne J. Matthee
Major Professor and Chairman

James J. Faynham
Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination:

October 1, 1976