1987

Polypeptides: Conformational Transition and Complex Formation With Catechins and Procyanidins.

Luanne Faith Tilstra
Louisiana State University and Agricultural & Mechanical College

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

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/4427

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.
Polypeptides: Conformational transition and complex formation with catechins and procyanidins

Tilstra, Luanne Faith, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1987
POLYPEPTIDES:  
CONFORMATIONAL TRANSITION AND  
COMPLEX FORMATION WITH  
CATECHINS AND PROCYANIDINS

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 Chemistry

by

Luanne Faith Tilstra  
B.A., Central University of Iowa, 1983  
August 1987
ACKNOWLEDGMENTS

My gratitude to Professor Wayne Mattice for patiently guiding my floundering attempts at research. Thanks to Dr. Phyllis E. Stoll for helping to make this text presentable.

Special thanks to Phillip C. Smith for encouragement, confidence, and love.

I would also like to acknowledge the Alumni Federation of Louisiana State University whence came the Fellowship that made this work financially possible.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ....................................... ii
LIST OF TABLES .......................................... vi
LIST OF FIGURES ......................................... viii
ABSTRACT ............................................... xiii

PART I
CONFORMATIONAL TRANSITIONS OF POLYPEPTIDES:
THE MEAN SQUARE END TO END
DISTANCE FOR THE β-SHEET TO RANDOM COIL TRANSITION

Chapter

ONE. HISTORICAL BACKGROUND ......................... 2
   Two Conformations of Polypeptides ............... 2
      Solid State Conformation .................. 4
      Conformation in Solution ................ 8
      Kinetics of Transition ................... 10
      Conformations in Proteins ............... 11

TWO. THE CONFORMATION PARTITION FUNCTION ......... 14
   Statistical Mechanics ....................... 14
      Conformation Partition Function ....... 14
      Additional Equations ................... 16
      Helix-Coil Transition .................... 18
      Possible Conformations ................. 18
      Statistical Weight Assignments ....... 20
      Determining the Conformation
         Partition Function ................... 22
      Experimental Values .................... 25
   Sheet-Coil Transition ....................... 26
      Possible Conformations ................. 26
      Statistical Weight Assignments ....... 29
      Determining the Conformation
         Partition Function ................... 30

THREE. THE MEAN SQUARE END TO END DISTANCE ....... 38
   Defining Matrices ........................... 38
      Transformation Matrix .................. 39
      The Generator Matrix .................. 41
      The F Matrix ............................ 43
      Sheet-Coil Transition .................... 44
      Defining the Transformation Matrix ..... 44
      Constructing the Generator Matrix ..... 47
      Building the F Matrix ................... 48
<table>
<thead>
<tr>
<th>Four. Computation</th>
<th>........................................</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Conformation Partition Function</td>
<td>........................................</td>
<td>56</td>
</tr>
<tr>
<td>The Mean Square End to End Distance</td>
<td>........................................</td>
<td>60</td>
</tr>
<tr>
<td>Modifications</td>
<td>........................................</td>
<td>60</td>
</tr>
<tr>
<td>Testing the Program</td>
<td>........................................</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Five. Results of Calculations</th>
<th>........................................</th>
<th>69</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Value of I</td>
<td>........................................</td>
<td>69</td>
</tr>
<tr>
<td>Values of Statistical Weights</td>
<td>........................................</td>
<td>71</td>
</tr>
<tr>
<td>Calculation of Other Parameters</td>
<td>........................................</td>
<td>74</td>
</tr>
<tr>
<td>Group I: $\delta, r = 1.0, 0.1$</td>
<td>........................................</td>
<td>74</td>
</tr>
<tr>
<td>Group II: $\delta, r = 0.3, 0.6, 0.9$</td>
<td>........................................</td>
<td>91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part II</th>
<th>Complex Formation of Polypeptides and Proanthocyanidins</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Six. Background and Literature Review</th>
<th>........................................</th>
<th>107</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catechins and Procyanidins</td>
<td>........................................</td>
<td>107</td>
</tr>
<tr>
<td>Structure</td>
<td>........................................</td>
<td>107</td>
</tr>
<tr>
<td>Circular Dichroism</td>
<td>........................................</td>
<td>112</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>........................................</td>
<td>114</td>
</tr>
<tr>
<td>Interaction of Catechins and Procyanidins</td>
<td>........................................</td>
<td>115</td>
</tr>
<tr>
<td>with Polymers</td>
<td>........................................</td>
<td>118</td>
</tr>
<tr>
<td>Polymers</td>
<td>........................................</td>
<td>118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seven. Theory of Experimental Methods</th>
<th>........................................</th>
<th>124</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultraviolet-Visible Absorption</td>
<td>........................................</td>
<td>124</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>........................................</td>
<td>125</td>
</tr>
<tr>
<td>Circular Dichroism</td>
<td>........................................</td>
<td>133</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eight. Materials</th>
<th>........................................</th>
<th>144</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Nine. Methods</th>
<th>........................................</th>
<th>148</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Ten. Results</th>
<th>........................................</th>
<th>154</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorbance</td>
<td>........................................</td>
<td>154</td>
</tr>
<tr>
<td>Catechins and Procyanidin B-1</td>
<td>........................................</td>
<td>160</td>
</tr>
<tr>
<td>Circular Dichroism</td>
<td>........................................</td>
<td>160</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>........................................</td>
<td>166</td>
</tr>
<tr>
<td>Synthetic Polymers</td>
<td>........................................</td>
<td>169</td>
</tr>
<tr>
<td>Expected Behavior of Polypeptides</td>
<td>........................................</td>
<td>172</td>
</tr>
<tr>
<td>Poly [N-hydroxybutyl glutamine]</td>
<td>........................................</td>
<td>174</td>
</tr>
<tr>
<td>Circular Dichroism</td>
<td>........................................</td>
<td>174</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>........................................</td>
<td>176</td>
</tr>
<tr>
<td>Poly [N-hydroxypropyl glutamine]</td>
<td>........................................</td>
<td>181</td>
</tr>
<tr>
<td>Circular Dichroism</td>
<td>........................................</td>
<td>181</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>........................................</td>
<td>181</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1. Atomic Coordinates of the Residues in the Repeating Unit of the Antiparallel-Chain Pleated Sheet Structure in the $\beta$ form of Poly (L-Alanine).</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2. Helix Coil Transition Statistical Weights.</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>3. Allowed Conformations for a Polypeptide Undergoing Sheet Coil Transition.</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>4. Sheet-Coil Transition Statistical Weights.</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>5. Values of $Z$ when $r = \delta = t = 1.0$.</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>6. The Conformation Partition Function Calculated by Hand and by Computer: $r = 0.1$, $\delta = 0.5$, $t = 2.0$.</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>7. Mean Square End to End Distance for a Straight Line Transformation.</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>8. Sheet-Coil Transition. $\delta = 1.0$, $r = 1.0$, $I = 12$, $n = 300$.</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>9. Sheet-Coil Transition. $\delta = 1.0$, $r = 0.1$, $I = 6$, $n = 300$.</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>10. Sheet-Coil Transition. $\delta = 0.1$, $r = 1.0$, $I = 20$, $n = 300$.</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>11. Sheet-Coil Transition. $\delta = 0.3$, $r = 0.3$, $I = 20$, $n = 300$.</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>12. Sheet-Coil Transition. $\delta = 0.3$, $r = 0.6$, $I = 25$, $n = 300$.</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>13. Sheet-Coil Transition. $\delta = 0.3$, $r = 0.9$, $I = 45$, $n = 300$.</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>14. Sheet-Coil Transition. $\delta = 0.6$, $r = 0.3$, $I = 12$, $n = 300$.</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>15. Sheet-Coil Transition. $\delta = 0.9$, $r = 0.3$, $I = 10$, $n = 300$.</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>16. Maximum Polymer Concentrations.</td>
<td>149</td>
<td></td>
</tr>
</tbody>
</table>
17. Absorption Coefficients of Polymers, (+)-Catechin, and (-)-Epicatechin. . . . . . . 159

18. Circular Dichroism of Catechins and Procyanidin B-1. . . . . . . . . . . . . . . . . . . 164
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Polypeptide chain structure.</td>
<td>3</td>
</tr>
<tr>
<td>2. Stereochemistry of amino acids.</td>
<td>5</td>
</tr>
<tr>
<td>4. Some potential conformations of a heptapeptide and their corresponding statistical weights.</td>
<td>19</td>
</tr>
<tr>
<td>5. Determining the conformation partition function, Z, for the helix coil transition by enumeration for n = 4.</td>
<td>23</td>
</tr>
<tr>
<td>6. Defining potential conformations for a polypeptide undergoing the sheet coil transition.</td>
<td>27</td>
</tr>
<tr>
<td>7. Determining Z for the sheet coil transition when n = 4, I = 2.</td>
<td>36</td>
</tr>
<tr>
<td>8. Definition of the coordinate system for a virtual bond (from Flory 1969).</td>
<td>40</td>
</tr>
<tr>
<td>9. A chain in which every residue is in a β-sheet.</td>
<td>45</td>
</tr>
<tr>
<td>10. The generator matrices used in calculations.</td>
<td>49</td>
</tr>
<tr>
<td>11. A chain with a defined conformation used to define which generator matrix should go with which statistical weight in the program.</td>
<td>51</td>
</tr>
<tr>
<td>12. The F matrix for I = 3. Also shown are the appropriate initial row and final column.</td>
<td>53</td>
</tr>
<tr>
<td>13. Possible conformations of short chains.</td>
<td>58</td>
</tr>
<tr>
<td>14. End effects. (a) Expected allowable conformations. (b) Conformations that become allowable when the length of the Nth bond is set equal to zero. (c) Conformation constructed by the F matrix in place of corresponding conformations illustrated in (a) due to the 'begin' effect.</td>
<td>62</td>
</tr>
<tr>
<td>15. The ratio of values expected (column (a) in Figure 14) to the values calculated by the program (column (c) in Figure 14).</td>
<td>67</td>
</tr>
</tbody>
</table>
16. The effect of increasing $I$. Shown is data for
   $n = 300, \tau = 0.3, \delta = 0.9, I = 5$ (□);
   $I = 10$ (+); $I = 15$ (◇). .................... 72

17. A comparison of the mean square end to end
distance as a function of $n$ for a random coil
(□) and the average sheet (+) when
   $\delta = \tau = 1.0$. ........................ 76

18. Effect of $\tau$ on the square root of the mean square
    end to end distance when $n = 300$:
    $\delta = \tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (+); and
    $\delta = 0.1, \tau = 1.0$ (◇). (a) $f_1 = 0$;
    (b) $f_1 = 0$. ............................. 81

19. A plot of the square root of the mean square end
to end distance as a function of the average
number of residues per sheet for $n = 300$:
   $\delta = 1.0, \tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (+);
   and $\delta = 0.1, \tau = 1.0$ (◇). ............ 86

20. The effect of $\tau$ on the fraction of residues in a
    sheet with $n = 300$. ...................... 88

21. The relationship between $\langle r^2 \rangle$ and the fraction
    of residues in a sheet for $n = 300$: $\delta = 1.0$,
    $\tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (+); $\delta = 0.1$,
    $\tau = 1.0$ (◇). ....................... 90

22. The effect of $\tau$ on the square root of the mean
    square end to end distance for $\delta = \tau = 0.3$ (□);
    $\delta = 0.6, \tau = 0.3$ (+); $\delta = 0.9, \tau = 0.3$ (◇). 97

23. The effect of $\tau$ on the square root of the mean
    square end to end distance for $\delta = \tau = 0.3$ (□);
    $\delta = 0.3, \tau = 0.6$ (+); $\delta = 0.3, \tau = 0.9$ (◇).
    (a) $f_1 = 0.78$; $f_1 = 0$. ............... 98

24. The effect of $\tau$ on the fraction of residues in a
    sheet for $\delta = \tau = 0.3$ (□); $\delta = 0.6, \tau = 0.3$
    (+); $\delta = 0.9, \tau = 0.3$ (◇). (a) $f_1 = 0.78$;
    (b) $f_1 = 0$. .......................... 99

25. The effect of $\tau$ on the fraction of residues in a
    sheet for $\delta = \tau = 0.3$ (□); $\delta = 0.3, \tau = 0.6$
    (+); $\delta = 0.3, \tau = 0.9$ (◇). .......... 100

26. The relationship between $\langle r^2 \rangle$ and fraction of
    residues in a sheet for $n = 300$: $\delta = 0.3$,
    $\tau = 0.3$ (□); $\delta = 0.6, \tau = 0.3$ (+); $\delta = 0.9$,
    $\tau = 0.3$ (◇). ........................ 103

ix
The relationship between $\langle r^2 \rangle$ and fraction of residues in a sheet for $n = 300$: $
abla = 0.3$, $\tau = 0.3$ (□); $\delta = 0.3$, $\nu = 0.6$ (+); $\theta = 0.3$, $\tau = 0.9$ (Δ) .................................. 104

The skeletal structure of flavonoids including the numbering system ............................... 108

The structure of the flavan-3-ol monomers .............................. 110

The four 4β-8 dimers of (+)-catechin and (-)-epicatechin ....................... 113

The structure of synthetic polymers used in this study .............................. 120

The structures of poly (L-glutamates) used in this study .............................. 121

Structure of homopolypeptides used in this study .............................. 123

Block diagram of the Hewlett-Packard 8451A Diode Array Spectrophotometer .............................. 126

Jablonski Diagram .............................. 127

Block Diagram of the SLM 8000C Spectrofluorometer .............................. 131

Comparison of the experimental and theoretical circular dichroism curves for "infinite" right-handed α helical poly (L-alanine) in trifluoroacetic acid, 98.5:1.5 v/v (Woody 1968) .................................. 136

Comparison of experimental and theoretical circular dichroism curves for high molecular weight poly (L-lysine) and an antiparallel β pleated sheet of two strands, each ten residues long (Pysh 1970) .................................. 138

Circular dichroism spectra of poly (L-lysine) in the α helical (1), β (2), and random (3) conformations (from Greenfield and Fasman 1969) .................................. 139

A diagram of the JASCO 500-A .................................. 141

Absorbance spectra of (a) (+)-catechin, (b) (-)-epicatechin, and (c) procyanidin B-1. 155

Absorbance spectra of the polymers used in this study. (a) poly vinyl pyrrolidone; (b) polyacrylamide; (c) poly ethylene glycol;
(d) hydroxypropyl cellulose; (e) poly [N-hydroxybutyl glutamine], molecular weight = 60,000; (f) poly [N-hydroxybutyl glutamine], molecular weight = 400,000; (g) copoly [N-hydroxyethyl:hydroxybutyl glutamine]; (h) poly [N-hydroxypropyl glutamine]; (i) copoly [(N-hydroxybutyl glutamine):arginine]; (j) poly (L-glutamic acid); (k) poly (D-glutamic acid); (l) poly (S-carboxymethyl cysteine); (m) poly (L-arginine); (n) poly (L-lysine).

43. Ellipticity of (a) (+)-catechin, (b) (-)-epicatechin, and (c) procyanidin B-1. 161

44. Corrected fluorescent emission of (+)-catechin and procyanidin B-1. 161

45. Relative fluorescence of (+)-catechin (□) and procyanidin B-1 (+) in solutions of polymer as a function of polymer concentration. (a) poly vinyl pyrrolidone, 0 to 8 mg/ml. Filled circles are results of Bergmann (1986) (b) polyethylene glycol, 0 to 8 mg/ml (c) polyacrylamide, 0 to 2 mg/ml (d) hydroxypropyl cellulose, 0 to 2 mg/ml. 170

46. The temperature-induced transition of PHBG as measured by circular dichroism. The inset shows the [θ] at 222 nm as a function of temperature (□) and also with (-)-epicatechin concentrations of 0.01 mg/ml (+), 0.05 mg/ml (◇), and 0.1 mg/ml (△). Error bars on the middle line apply to all three spectra. 175

47. The effect of PHBG on the fluorescence of (+)-catechin and procyanidin B-1. 177

48. A Stern-Volmer plot for butanolamine as a quencher of (+)-catechin (□) and procyanidin B-1 (+) fluorescence. 179

49. The temperature-induced transition of PHPG as measured by circular dichroism. The inset shows [θ] at 222 nm as a function of temperature for PHPG (□), PHPG + (+)-catechin (+), and PHPG + procyanidin B-1 (◇). 182

50. The effect of polymer on the fluorescence of (+)-catechin and procyanidin B-1. PHBG [MW = 60000] (□); PH(E:B)G (+); PHPG (◇); P(HBG:Arg) (△). 183
51. The temperature-induced transition of PH(E:B)G as measured by circular dichroism. Plots are for 5°C (□), 25°C (+), and 40°C (◇). 184

52. The temperature-induced transition of PH(E:B)G (□) and the effect of (a) (+)-catechin and (b) procyanidin B-1 at proanthocyanidin concentrations of 0.05 mg/ml (+) and 0.1 mg/ml (◇). 186

53. The temperature-induced transition of P(HBG:Arg). The cd spectra are shown for 30°C (□) and 60°C (+). The inset is a plot of [θ] vs temperature for P(HBG:Arg) and the 60,000 PHBG. 187

54. The effect of (+)-catechin (a) and procyanidin B-1 (b) on the temperature-induced transition of P(HBG:Arg). Shown for the absence of (□) and three concentrations of (+)-catechin and procyanidin B-1; 0.01 mg/ml (+), 0.05 mg/ml (◇), and 0.1 mg/ml (△). 188

55. The effect of (a) poly (L-arginine) (pH = 8.2) and (b) poly (L-lysine) (pH = 7.8) on the fluorescence of (+)-catechin (□) and procyanidin B-1 (+). 190

56. The pH-induced transition of poly (L-glutamic acid). (a) Ellipticity at 3 pH as a function of wavelength. (b) Ellipticity at 220 nm as a function of pH and the effect of (□) 0.01 mg/ml (+)-catechin, (+) 0.05 mg/ml (+)-catechin, and (◇) 0.1 mg/ml (+)-catechin. 192

57. The effect of (a) poly (L-glutamic acid) and (b) poly (D-glutamic acid) on the fluorescence of (+)-catechin (□) and procyanidin B-1 (+) as poly (Glutamic acid) goes through a pH-induced transition. 193

58. Circular dichroism spectra of poly (S-carboxymethyl cysteine) (DP = 330) at pH = 4.3 (□) and pH 5.9 (+). Also shown are results obtained by Maeda (1982) (dashed line). 195

59. pH-induced transition of poly (S-carboxymethyl cysteine) as measured by circular dichroism. (a) DP = 330 (□), with (+)-catechin added (◇). Dotted line indicates results of Maeda (1984). (b) DP = 560 at 200 nm (□) and at 220 nm (+); with (+)-catechin added after equilibration (◇). 196
Theoretical explanations of the α-helix to coil transition which employ rotational isomeric state calculations are reviewed. Recent advances in this field include the development of similar fundamentals for the intramolecular antiparallel β-sheet to coil transition. This dissertation adds to the current body of knowledge by outlining a formularization to calculate the mean square end-to-end distance, $<r^2>_0$, of a homopolypeptide for the transition from random coil to intramolecular antiparallel β-sheet conformation. The formularization is translated into a computer program in the C language and values of $<r^2>_0$ are calculated for a number of different statistical weights. These are analyzed parallel to other statistical parameters such as the fraction of residues in the β sheet conformation, the average number of residues per sheet, the average number of strands per sheet, and the average number of residues per strand. Included are plots of $\sqrt{<r^2>}$ against the fraction of residues in a sheet, two experimentally obtainable parameters.

Catechins and procyanidins, flavonoids synthesized by plants, are the monomers and dimers of the group of larger molecules known as tannins. Tannins are known to interact with proteins, rendering certain
digestive enzymes inactive and causing the precipitation of otherwise soluble proteins. Experimental research revolves around the effect of catechins and procyanidins on the conformational transitions mentioned above. The effect of catechins and procyanidins on the transition is studied by following the transition of a number of polypeptides in the presence and absence of catechins and procyanidins. The polypeptides vary by the degree of hydrophobic nature, and charge density in their side chains. The systems are also observed by fluorescence to determine the degree of interaction of catechins and procyanidins with the various side chains. Results of the experimental work support earlier proposals that the interaction of tannins and proteins is uniquely favored for the amino acid residue, proline. There may also be a slight effect on the $\beta$-sheet coil transition of poly (S-carboxymethyl cysteine). This is proposed to be due to a stabilization of the initial strand of the sheet by hydrogen bonding to the backbone.
PART I:
CONFORMATIONAL TRANSITIONS OF POLYPEPTIDES:

THE MEAN SQUARE END-TO-END DISTANCE
FOR THE $\beta$-SHEET TO RANDOM COIL TRANSITION
Polypeptides are polymers of amino acids. The name arises from the peptide bond $N - C$ which connects consecutive amino acid residues. The bond angles and lengths of a polypeptide chain, defined in Figure 1, are virtually constant. Variations in the conformation of the polymer occur primarily by rotation about the $N - C$ bond, $\phi$, and about the $C^{\alpha} - C$ bond, $\psi$. According to international convention, $\phi$ and $\psi$ are equal to $0^\circ$ when $C^{\alpha} - R$ is trans to $N - H$ and $C - O$ respectively. Because of the resonance of the double bond between the $C - O$ and the $C - N$ positions, atoms enclosed within the dashed-line boxes lie on a plane.

A polypeptide is often considered to have a relatively simple composition that can be synthesized in the laboratory. Although proteins are technically polypeptides, the term protein usually connotes a polymer with a more complicated composition than that of a polypeptide, and a polymer that was first synthesized in nature rather than in a laboratory. These are not officially sanctioned definitions but are in accord with common usage.
Figure 1. Polypeptide chain structure. Two peptide bonds are shown, each enclosed in a dashed line rectangle. The six atoms enclosed in each rectangle are co-planar. Data adapted from Ramachandran (1974).
Two Conformations of Polypeptides

In the early 1950's Pauling and Corey published their discussions of proposed structures of polypeptides. Two of these are the helix (Pauling and Corey 1950; Pauling and Corey 1951a; Pauling and Corey 1951b) and sheet (Pauling and Corey 1951c; Pauling and Corey 1953). The α-helix conformation is a regular repeating pattern with 3.7 residues per turn. Amino acids have a chiral carbon and can exist in two configurations (Figure 2). Pauling and Corey suggest that the L-amino acids should form a right-hand helix while D-amino acids should form a left-hand helix. The two corresponding helices are also illustrated in Figure 2. It is the L-configuration of amino acids that occurs in proteins, hence the right-handed helix is dominant in proteins.

In the sheet structure, nearly extended polypeptide chains lie parallel to each other to form sheets. With respect to the sequence -NH-CHR-CO-, the sense of the strands alternates in the antiparallel sheet (Figure 3) and is the same for all strands in the parallel sheet.

Solid State Conformation

When x-ray diffraction was first being used to characterize the structure of fibrous proteins, certain proteins appeared to have a repeating pattern. Of particular note are the patterns obtained from silk (Meyer and Mark 1928) and from stretched mammalian β-keratin
Figure 2. Stereochemistry of amino acids. The two configurations form helices which rotate in the opposite direction relative to translation. Helix sense is sketched below the corresponding configuration.
Figure 3. Anti-parallel sheet conformation of polypeptides (Pauling and Corey 1951c).
Silks are often classified on the basis of the predominant form of regular secondary structure as determined by x-ray diffraction. The group known as \( \beta \) silks has a pseudo-repeat of about 0.7 nm parallel to the fiber axis (Brill 1923). Later studies yield comparable values. Kratky and Kuriyama (1931) report 0.695 ± 0.025 nm. Values reported by Bamford et al. (1953) and Marsh et al. (1955) are 0.694 and 0.697 ± 0.003 nm respectively. Kratky and Kuriyama (1931) also identify reciprocal lattice vectors perpendicular to the fiber axis. While these results clearly indicate a repeating structure, the nature of that structure was not initially clear.

Shortly after the appearance of the papers by Pauling and Corey, it was established that the overall distribution of intensity in the diffraction patterns of certain synthetic polypeptides was consistent with the presence of an \( \alpha \)-helix. One well-studied homopolypeptide is poly (L-alanine). Oriented specimens of poly (L-alanine) in which the majority of the polymer has an \( \alpha \)-helical conformation are prepared by stretching spun fibers (Elliot 1967). Although early reports suggested a slight favoring for a left-hand helix for poly (L-alanine) (Brown and Trotter 1956), studies of Elliot and Malcolm (1958) helped clarify this potential disagreement with Pauling and Corey. If the chain direction with respect to the sequence \(-NH\-CHR\-CO-\) is random within a fiber, the intensities of reflections can
be explained by a crystal model of right-hand helices. The
diffraction pattern of stretched poly (L-alanine) was
remeasured (Arnott and Wonacott 1961a; Arnott and Wonacott
1961b; Arnott and Dover 1967; Arnott 1968) to obtain
quantitative intensity data on 61 reflections. A limited
least squares refinement allowed calculation of atomic
coordinates.

Specimens of poly (L-alanine) obtained by steam-
stretching fibers spun from dichloroacetic acid solution
yield a different diffraction pattern than those obtained
by the previous preparation. A limited least squares
refinement of the antiparallel-chain pleated-sheet model
was carried out on the basis of 29 x-ray intensity data
(Arnott and Dover 1967). The atomic coordinates of the
four residues in the repeating unit of the antiparallel-
chain pleated-sheet, based on the values derived by Arnott
and Dover (1967) are given in Table 1. The fiber structure
of the $\beta$ silks is also based on an antiparallel-chain
pleated sheet. The obvious agreement of this proposed
structure with x-ray data establishes the $\beta$-sheet as a
common structure in proteins.

Conformation in Solution

In the preceding section, polypeptides discussed
are in the solid state; i.e., crystalline or fibrous
protein. The conformation of a polypeptide in solution may
be slightly different. Certainly it is not rigidly fixed.
Polymers in dilute solution are able to displace solvent
Table 1.—Atomic Coordinates of the Residues in the Repeating Unit of the Antiparallel-Chain Pleated-Sheet Structure in the $\beta$ form of Poly (L-alanine)

<table>
<thead>
<tr>
<th>Atom</th>
<th>$u(\AA)$</th>
<th>$v(\AA)$</th>
<th>$w(\AA)$</th>
<th>Hydrogen bonds</th>
<th>$u(\AA)$</th>
<th>$v(\AA)$</th>
<th>$w(\AA)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>2.695</td>
<td>-0.837</td>
<td>0.130</td>
<td>6.765</td>
<td>0.837</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3.665</td>
<td>-0.857</td>
<td>-0.100</td>
<td>5.795</td>
<td>0.857</td>
<td>-0.100</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2.225</td>
<td>0.383</td>
<td>0.790</td>
<td>7.235</td>
<td>-0.383</td>
<td>0.790</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1.155</td>
<td>0.453</td>
<td>0.700</td>
<td>8.305</td>
<td>-0.453</td>
<td>0.700</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2.605</td>
<td>0.353</td>
<td>2.280</td>
<td>6.855</td>
<td>-0.353</td>
<td>2.280</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2.855</td>
<td>1.608</td>
<td>0.130</td>
<td>6.605</td>
<td>-1.608</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>4.065</td>
<td>1.638</td>
<td>-0.130</td>
<td>5.395</td>
<td>-1.638</td>
<td>-0.130</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2.035</td>
<td>2.608</td>
<td>-0.130</td>
<td>7.425</td>
<td>4.282</td>
<td>-0.130</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1.065</td>
<td>2.588</td>
<td>0.100</td>
<td>8.395</td>
<td>4.302</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2.505</td>
<td>3.828</td>
<td>-0.790</td>
<td>6.955</td>
<td>3.062</td>
<td>-0.790</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3.575</td>
<td>3.898</td>
<td>-0.700</td>
<td>5.885</td>
<td>2.992</td>
<td>-0.700</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2.125</td>
<td>3.798</td>
<td>-2.280</td>
<td>7.335</td>
<td>3.092</td>
<td>-2.280</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.875</td>
<td>5.053</td>
<td>-0.130</td>
<td>7.585</td>
<td>1.837</td>
<td>-0.130</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.665</td>
<td>5.083</td>
<td>0.130</td>
<td>8.795</td>
<td>1.807</td>
<td>0.130</td>
<td></td>
</tr>
</tbody>
</table>

molecules and change conformations. Even a very stable state which has a large probability of occurrence may have some fluctuation about its dihedral angles. At high temperatures one may expect a more random conformation due to an increase in thermal energy that makes accessible additional states obtained by rotation about $\phi$ and $\psi$. The nature of the solvent may also affect the conformation assumed by the polypeptide. If the solvent interacts favorably with the side chains--e.g., both are hydrophobic--the polymer can be expected to assume a conformation that exposes side chains to the solvent. Contrariwise, an unfavorable solvent-side chain interaction would favor a conformation that keeps side chains in close contact with each other; e.g., the $\alpha$-helix. The pH of a solution may affect the conformation of ionic polymers. If the pH is such that all side chains are similarly charged, their side chains will repel each other and the favored conformation will be a random coil. If, however, some side chains are neutralized, a conformation in which the side chains are close to each other may form; e.g., a helix in the case of poly (glutamic acid) (Lotan, et al. 1965) and a $\beta$-sheet in the case of poly (S-carboxymethyl cysteine) (Maeda, et al. 1982). Complete neutralization usually results in the aggregation and eventual precipitation of the polymer with ionic side chains.

**Kinetics of Transition**

The kinetics of formation of an $\alpha$-helix from a
random coil in dilute solution are very fast. The transition is on a microsecond time-scale (Lader and Mandelkern 1979). The time required for most experimental procedures which measure conformation transitions is much greater than the time required for the random coil to $\alpha$-helix transition.

It has been demonstrated that formation of a sheet from random coil takes much longer than for the transition to an $\alpha$-helix (Maeda, et al. 1982). A solution of poly (S-carboxymethyl cysteine) in water adjusted to a pH that favors sheet formation must be allowed to equilibrate 24 hours before the transition is complete (as shown by circular dichroism at 222 nm). Because of the extra time required for this transition, measurements are more difficult. While the transition is occurring, the pH drifts because the transition is accompanied by a change in pK for the carboxyl group. It often happens that the pH drifts out of the sheet-forming range. Readjustment and the consequent 24 hour wait must follow. Measurements are much more tedious for the sheet-coil transition, and acquiring a value at a given desired pH is difficult.

Conformations in Proteins

The helix and sheet conformations are not unique to homopolypeptides. Recent studies have shown clear evidence of helices in globular proteins; e.g., hemoglobin (Perutz 1951). Examples of $\beta$-sheet conformation in proteins are silk, and $\beta$-keratin. There are also examples
of sheet in other proteins. One of these is the tail fiber of Adenovirus (Green, et al. 1983).

The relationship between the structure and function of a protein is well established. Proteins which have lost their native conformation no longer perform the function for which they were originally intended. Many proteins have specific sites at which the native conformation is critical. The 'shape' of the rest of the macromolecule is important to the protein's function only in so much as it affects the conformation at the active site. While there may be some flexibility in the allowable structure of a protein, the relationship between structure and function is certain. For this reason, a variety of experimental and theoretical structural studies of the structure have been undertaken.

The shape of proteins depends upon the rotation about backbone bonds for each amino acid residue. Within a polypeptide chain, bond angles are essentially fixed at the angles shown in Figure 1. Overall chain conformation is determined by the value of the dihedral angles $\phi$ and $\Psi$. Globular proteins have a clearly defined conformation. It may not be a regular repeating pattern of $\phi$ and $\Psi$, but it is defined. In a crystal of several molecules of hemoglobin, all the molecules have almost the same conformation. A randomly coiled polypeptide, on the other hand, has no defined conformation.

The conformation of a protein is very likely to
be directly related to the identity and order of the amino acid residues that make up its chain. There are several types of potential interaction between the side chains which may affect this relationship. Three of them include nearest-neighbor interaction, interaction with a residue a fixed distance down the chain, and interaction with a residue any distance down the chain. The identity of the residue will affect how and with which other residues it interacts.

Experimental studies of homopolypeptides allow the study to focus on the effect of solution conditions on the random coil to $\alpha$-helix or the random coil to $\beta$-sheet transition for a single residue. In this way, it is possible to isolate the probability of a given amino acid being in or initiating a given conformation. The study is therefore limited to nearest-neighbor interactions.
CHAPTER TWO
THE CONFORMATION PARTITION FUNCTION

For the results from experiments on simple polypeptides to be useful to the understanding of the stability of the conformation of proteins, there must be an established theoretical bridge. The most thorough review of the theory presented in this dissertation has been given by Flory (1974). The goal of this analysis is a tenable definition of the conformation partition function, the sum of the statistical weights of all the possible conformations of a polymer.

Statistical Mechanics

Conformation Partition Function

The conformation partition function for the rotational states of a polymer chain can be defined as

$$ z = \int e^{-E(\Omega)/kT} \, d(\Omega) $$

(1)

where $\Omega$ indicates that all the dihedral angle variables for all the residues are to be included in the integration, $E(\Omega)$ represents the energy of these many states, $k$ is the Boltzmann constant, and $T$ is the temperature. A special case of Equation (1) is

$$ z = \left[ \int e^{-E(\Phi, \Psi)/kT} \, d(\Phi, \Psi) \right]^n = s^n $$

(2)
where $Z$ is the conformation partition function for one residue and the $n$ identical residues are assumed to be independent so that $Z = a^n$. This is a good approximation in the case of the random coil of poly-(L-alanine) (Brant and Flory 1965).

If the residues are interdependent Equation (1) must be evaluated as a nested series of integrals. This can be approximated by a nested sum.

$$Z \approx \sum_{w_1} \sum_{w_2} \cdots \sum_{w_n} e^{-E(\Omega)/kT}$$

where each sum is for a given residue. Here $w_n$ represents the dihedral angles for one residue. For polypeptides there are two dihedral angles so each summation is actually a double sum. The sums are nested and inseparable if the statistical weight for the conformation of a residue is dependent on the conformation of the other residues. This nested sum is frequently handled by matrix methods.

Serial matrix multiplication, where each matrix represents one residue, retains specific information about interactions between neighboring residues. If $U$ is a matrix of statistical weights applicable to the possible conformations in the transition; i.e., $\alpha$-helix or random coil in the helix-coil transition and strand/bend/loop or random coil in the sheet-coil transition, the conformation partition function for the homopolypeptide is given by

$$Z = (\text{row}) U^n - 1 (\text{column})$$

Values in the initial row and final column will be defined in a later section.
Additional Equations

The fraction of residues in a given conformation, $b$, is a function of the probability of that conformation.

$$f_b = \frac{1}{n} \int n_b P(\Omega) d(\Omega)$$

$$f_b = \frac{n}{n_b} \frac{P(\Omega) d(\Omega)}{\Omega}$$

If $\int P(\Omega) d(\Omega) = 1$.

Here $n$ is the total number of residues in a chain and $n_b$ is the number of residues in the $b$ conformation. $P(\Omega)$ is the probability for a given set of dihedral angles. This can be written as

$$f_b = \frac{1}{n} \int n(b) s\omega(\Omega) d(\Omega) = \frac{1}{n} \int n(b) s\omega(\Omega) d(\Omega).$$

In this equation $s\omega(\Omega)$ refers to the statistical weight for a given set of dihedral angles. Once again, the integral can be estimated by a series of nested sums.

$$f_b = \frac{1}{n} \sum_{\Omega} n_b s\omega(\Omega)$$

This equation is also more manageable as a serial matrix multiplication.

One useful feature of equation (6) is the presence of statistical weights in the numerator and in the denominator. This means that if all the statistical weights are multiplied by some constant, there is no effect on the value of $f_b$. It is this feature that allows the arbitrary assignment of a numerical value for the statistical weight of a random coil, usually $z = 1$. In this case $z$ for the random coil is equal to one. The absolute value of $z$, however, loses significance. For the
If this is replaced with
\[ Z = (1)^n = 1 \]  \hspace{1cm} (9)
then \( Z \) has been multiplied by \((\pi)^{-n}\). For this reason the \( Z \) obtained from the assignment of \( \pi = 1 \) for the random coil residues is different from the true \( Z \) by \((\pi)^{-n}\).

One more equation bears mention. The mean square end-to-end distance is defined as
\[ <r^2> = \int r^2(\eta)P(\eta)d(\eta) \]  \hspace{1cm} (10)
if \( \int P(\eta)d(\eta) = 1. \)

As for the previous discussion
\[ <r^2> = \int r^2(\eta)sw(\eta)d(\eta) \]  \hspace{1cm} (11)
\[ = z^{-1}\Sigma \cdots \Sigma r^2(\eta)sw(\eta). \]

In Equation (7) pertinent information about the number of residues in the desired conformation, \( n_b \), is contained in the matrices used to determine \( Z \). A more complicated situation is seen in Equation (10). Determining the mean square end-to-end distance requires information about the chain geometry (length of bonds, bond angles, and dihedral angles). It is still true that numerical values for the statistical weight for the random coil can be arbitrarily assigned because of the presence of the same statistical weights in the numerator and denominator. However, an additional serial matrix multiplication must be developed to tie the appropriate
geometries with each statistical weight.

**Helix-Coil Transition**

Matrices required for the serial matrix multiplication which describes the probability for transition between the random coil and α-helix conformation were developed by 1960 (Zimm and Bragg 1959). Use is made of the partition function from statistical mechanics to develop a statistical approach to polypeptide conformation. Statistical weights for every potential conformation in the helix coil transition are defined. These are then assigned to every potential conformation of a polypeptide, the statistical weights of every potential conformation are summed to yield the conformation partition function.

**Possible Conformations**

In the helix coil transition a residue may have one of two states. It may be either an α-helix with the defined dihedral angles or a random coil, which is in this case a residue with any dihedral angles other than those required for the α-helix. Consider the example in Figure 4. Given a chain seven residues long, allow each residue to have one of two states. Two possibilities for the conformation of a chain seven residues long are sketched. Rather than drawing dots and lines for all possibilities, a notation is introduced in which the letter 'c' represents a residue in the random coil conformation, and 'h' represents a residue in the helix conformation. The two conformations drawn are listed along with some additional possibilities.
Figure 4. Some potential conformations of a heptapeptide and their corresponding statistical weights.
Statistical Weight Assignments

The next step in this analysis is to assign statistical weights to the possible conformations for each residue. The statistical weights of a given conformation for each residue in a chain are multiplied together to give the statistical weight for that conformation for the entire chain. If a residue is a random coil, it is assigned a statistical weight of "1". This value can be assigned arbitrarily if the final calculated value of $Z$ is used only in equations which are of the form of Equations (6) and (11) where the statistical weights are relative values. If the conformation of the residue is that of an $\alpha$-helix, it is assigned a statistical weight of $s$, the precise value of $s$ depending on the identity of the residue. The nature of the $\alpha$-helix is such that there is a decrease in enthalpy, resulting in a decrease in free energy, once the first hydrogen bond is formed between residues $i$ and $i + 4$. Forcing the polypeptide to hold one specific conformation causes a decrease in entropy. Before the first hydrogen bond has formed, each residue that is added to the helix causes an increase in free energy because of the decrease in entropy. For this reason the theory of helix coil transition penalizes the initiation of a helix by assessing an additional statistical weight of $\sigma$, which penalizes the overall statistical weight of the helix for the effects of starting a helix.

Consider the following chains.
The statistical weight of this chain is $\sigma^4$. Another example is
\begin{equation}
\text{c h h h h c}
\end{equation}
which would have a statistical weight of $\sigma^5$. Finally the chain
\begin{equation}
\text{c h h c h h h}
\end{equation}
has a statistical weight of $\sigma^2\sigma^5$. These three possible conformations for a heptapeptide illustrate the relative effect of different values for $s$ and $\sigma$. The ratio of the statistical weights of the first two conformations is $\sigma^5/\sigma^4 = s$. If one considers a polypeptide chain going through the transition from structure (12a) to structure (12b), then $s$ can be considered to be an equilibrium constant. If $s < 1$, then the conformation labelled (b) is less probable than the conformation labelled (a), the equilibrium lies toward conformation (a). If $s > 1$ the opposite is true.

Consider the transition between conformations (12b) and (12c). In this instance the equilibrium constant is simply $\sigma$. The cost of going from (12b) to (12c) is the value of $\sigma$. If $\sigma < < 1$ then the transition to (12c) is less probable.

The ratio of the statistical weights of conformation (12c) to conformation (12a) is $\sigma^2\sigma^5/\sigma^4 = \sigma\sigma$.

If $\sigma\sigma < 1$, conformation (12a) is more probable. If $\sigma\sigma > 1$ conformation (12c) is more probable. It is possible for
conformation (12a) to be more probable when \( s > 1 \). This will be true when \( \sigma \) is so small that its product with \( s \) is less than one. Another way of saying this is that when the cost of initiation is high, multiple helices are less probable than a single helix with the same number of residues. However, even with a very small \( \sigma \) (large penalty for initiation) it is possible to choose \( n \) large enough that \( \sigma s^n > 1 \) if \( s > 1 \). In this case the preferred conformation should depend on \( n \). When the contribution of \( s \) to the statistical weight is large enough, the helix conformation will be present. The point to note is that both \( \sigma \) and \( s \) make a unique contribution to the statistical weight of a conformation.

**Determining the Conformation Partition Function**

If all possible conformations of a given chain length are enumerated and then added together, the result would be the conformation partition function, \( Z \). For the case of a chain four residues long, the value of \( Z \) is determined in Figure 5.

For very long chains using discrete enumeration to determine the conformation partition function becomes very tedious. An alternative approach is described. If one were to make a table to define all potential statistical weights for a residue which can be in a helix or a random coil, one can see that this table would be complete with four entries (Table 2). Looking at the first row, one can see that if residue \( i - 1 \) is in a random
<table>
<thead>
<tr>
<th>Letter Code</th>
<th>Statistical Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>c c c c</td>
<td>1</td>
</tr>
<tr>
<td>c c c h</td>
<td>$\sigma s$</td>
</tr>
<tr>
<td>c c h c</td>
<td>$\sigma s$</td>
</tr>
<tr>
<td>c h c c</td>
<td>$\sigma s$</td>
</tr>
<tr>
<td>h c c c</td>
<td>$\sigma s$</td>
</tr>
<tr>
<td>c c h h</td>
<td>$\sigma s^2$</td>
</tr>
<tr>
<td>c h h c</td>
<td>$\sigma s^2$</td>
</tr>
<tr>
<td>h h c c</td>
<td>$\sigma s^2$</td>
</tr>
<tr>
<td>h c h c</td>
<td>$\sigma^2 s^2$</td>
</tr>
<tr>
<td>c h c h</td>
<td>$\sigma^2 s^2$</td>
</tr>
<tr>
<td>h c c h</td>
<td>$\sigma^2 s^2$</td>
</tr>
<tr>
<td>c h h c</td>
<td>$\sigma^3 s^3$</td>
</tr>
<tr>
<td>h h h c</td>
<td>$\sigma^3 s^3$</td>
</tr>
<tr>
<td>h h c h</td>
<td>$\sigma^3 s^3$</td>
</tr>
<tr>
<td>h c h h</td>
<td>$\sigma^3 s^3$</td>
</tr>
<tr>
<td>h h h h</td>
<td>$\sigma^4 s^4$</td>
</tr>
</tbody>
</table>

$$Z = 1 + 4\sigma s + 3\sigma s^2 + 3\sigma^2 s^2 + 2\sigma s^3 + 2\sigma^2 s^3 + \sigma s^4$$

Figure 5. Determining the conformation partition function, $Z$, for the helix coil transition by enumeration for $n = 4$. 
Table 2.—Helix-coil Transition Statistical Weights

<table>
<thead>
<tr>
<th>conformation of residue i - 1</th>
<th>conformation of residue i</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>s</td>
</tr>
<tr>
<td></td>
<td>ss</td>
</tr>
</tbody>
</table>

If residue i has a coil conformation, and residue i is also a random coil, then the contribution of residue i to the statistical weight of the conformation of the chain is 1. If, however, residue i has an α-helix conformation, the contribution of residue i to the statistical weight of the conformation of the chain is ss. Looking now at the second row, if residue i - 1 has a helical conformation and residue i is a random coil, the contribution of residue i to the statistical weight of the conformation of the chain is still 1. If residue i is also a helix, its contribution to the statistical weight of the conformation of the chain is s.

Such a table can be used to determine the conformation partition function. The table is representative of the statistical weight of residue i with respect to its nearest preceding neighbor, residue i - 1. The table is made into a 2 x 2 matrix and one matrix is used to represent each residue in the chain. When these matrices are preceded by an appropriate row and followed by the appropriate column, and the whole is multiplied, a
Scalar value for the partition function is obtained. If, by definition, the nonexistent residue before the beginning of the chain is said to be a random coil (which is the same as saying a nonexistent residue cannot contribute to the stability of the helix), then the initial row is

\[ \begin{bmatrix} 1 & 0 \end{bmatrix} \]

The final column translates the vector into a scalar. The equation for \( Z \) becomes

\[
Z = \begin{bmatrix} 1 & \sigma s \end{bmatrix} \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix}^n - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{13}
\]

or

\[
Z = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{14}
\]

where \( n \) is the number of residues. Serial matrix multiplication for \( n = 4 \) leads to

\[
Z = 1 + 4\sigma s + 3\sigma s^2 + 2\sigma s^3 + \sigma s^4 + 3\sigma^2 s^2 + 2\sigma^2 s^3. \quad \tag{15}
\]

**Experimental Values**

Experiments done with the synthetic homopolypeptides have led to numerical values for \( \sigma \) and \( s \) (Sueki, et al. 1984). Conformational transitions are studied for certain modified homopolypeptides such as poly (hydroxyalkyl glutamines). In some cases a homopolypeptide will not form an \( \alpha \)-helix under any conditions. The effect of such amino acid residues on the well-characterized conformational transition of poly [N-hydroxybutyl glutamine] is used to determine the value of \( \sigma \) and \( s \) for
that amino acid residue. Values reported for $\sigma$ for amino acid residues in aqueous solution lie in the range $1 \times 10^{-5}$ to $2 \times 10^{-2}$, while values for $s$ vary from 0.59 to 1.35 (Sueki, et al. 1984).

**Sheet-Coil Transition**

When the same approach is used to define $Z$ for the $\beta$-sheet random coil transition a new problem is present. The $\alpha$-helix is essentially a one-dimensional conformation. After initiation, the helix either continues or it doesn't. Hydrogen bonds always form between residue i and residue $i+4$. An $\alpha$-helix of $n$ residues has only one possible shape. The sheet, however, has a two-dimensional character. A sheet of $n$ residues may assume many shapes when $n$ is large. The initial strand can be short or as long as the chain itself. A residue continuing the sheet may continue along the strand or start a bend to initiate a new strand. There are more options available to any one residue in a $\beta$-sheet. It is possible for residue i to join in hydrogen bond formation with essentially any other residue in the chain.

**Possible Conformation**

An approach similar to that taken to determine the conformation partition function for the $\alpha$-helix has been taken for the $\beta$-sheet (Mattice and Scheraga 1984). Look at Figure 6. Here is illustrated a sheet which contains at least one residue in every conformation that will be considered in this dissertation. Below is written
Figure 6. Defining potential conformations for a polypeptide undergoing the sheet coil transition.
the same conformation in a letter code. The lower case \( b_i \) represents a residue that is part of a sheet but does not have a corresponding residue in the previous strand of the sheet. Its subscript indicates its position from the beginning of the strand. The uppercase \( B_{ij} \) represents residues in the sheet conformation which have a corresponding residue in the previous strand. The notation includes two subscripts, \( i \) and \( j \). The first, \( i \), represents the position of the residue from the beginning of the strand. The second, \( j \), represents the number of residues in the preceding strand. Clearly \( i \) must be less than or equal to \( j \). The third new possibility is the residue that is in a loop between two strands. Its notation is \( l_i \), and the subscript refers to the distance from the beginning of the loop. Random coil is still defined as \( o \). A residue labelled \( 1^j \) is distinguishable from \( o \). This loop is under the restriction that it conclude at a position which starts a new strand to resume the building of the sheet from which it took off. The end of a sequence of \( 1 \)'s is restrained in a manner not seen in a sequence of \( o \)'s.

In the case of the sheet, the end effects last the entire length of one strand and the degree of free energy gain depends on the length of the first strand. All subsequent strands form hydrogen bonds with the previous strand when there is a corresponding residue in the previous strand and may contribute to a decrease in the free energy. For example, consider the conformation
illustrated in Figure 6. Let each dot represent a residue and each connecting line the virtual bond that connects α carbons. For the conformation illustrated here the labelling would be

\[ \text{cccb}_{1}\text{b}_{2}\text{B}_{12}\text{B}_{22}\text{b}_{3}\text{B}_{13}\text{B}_{23}\text{B}_{33}\text{l}_{1}\text{l}_{2}\text{l}_{3}\text{l}_{4}\text{l}_{5}\text{B}_{13}\text{B}_{23}\text{cc} \]

(16)

Statistical Weight Assignments

The statistical weight of each residue in a β-sheet includes \( t \) in the same way that \( s \) is included in the statistical weight for every residue in an α-helix. As in the theory for α-helix formation, there is penalty ascribed for starting the regular conformation. For the helix the penalty is defined as \( \sigma \). In the case of the β-sheet each residue in the initial row is penalized by a factor of \( t \). Initiation of sheet formation requires laying down the entire first strand. For this reason the initiation parameter for a sheet that is comparable to \( \sigma \) for the helix is \( T_{n}^{*} \) where \( n^{*} \) is the number of residues in a strand with no corresponding residues in a preceding strand.

There are two possible residue conformations that have no corresponding part in the helix. One is due to the two-dimensional nature of the sheet. The residues at a bend are different than those in the middle of a strand. The statistical weight penalty for this is \( \delta \). At a bend it is possible for the chain to leave the sheet and form an interstrand loop. If the loop consumes \( n_{l} \) residues, the interstand loop contributes a statistical weight of \( f_{l}^{n_{l}} \).
The value of \( f_1^n \) is not identical with the probability for the cyclization of a random flight chain but it provides a good approximation when \( f_1 = 0.78 \) (Mattice and Scheraga 1985). If \( f_1 = 0 \), loops are forbidden and all strands in a sheet must be connected by tight bends. For a residue in a random coil, the statistical weight is one. The contribution of the conformation in Figure 6 to the partition function for a chain eleven residues long that can be in either the sheet or random conformation is 

\[ 5^3 \cdot 10 f_1^5. \]

**Determining the Conformation Partition Function**

As in the theory for \( \alpha \)-helix formation it is possible and advantageous to make a table and build matrices for the determination of the conformation partition function. It is wise to include an additional variable in the theory. The regularity of the \( \alpha \)-helix requires that there always be a given number of residues per turn. Poly (L-alanine), for example, has 3.7 residues per turn. However, the potential architecture of a sheet is more varied. A chain of six residues, for example, can form any of the sheets shown in Table 3 when \( f_1 = 0 \). Because of this versatility, an additional parameter is needed in the theory for the transition between random coil and \( \beta \)-sheets. This parameter limits the number of residues per strand. By definition it can range from a value of 2 to the number of residues in the chain. Therefore, for \( n = 8 \), the parameter \( I \) can be 2, 3, 4, 5, 6, 7, or 8. The
Table 3.—Allowed Conformations for a Polypeptide Undergoing Sheet-Coil Transition.

<table>
<thead>
<tr>
<th>I</th>
<th>conformation</th>
<th>degeneracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ccccccc</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 ccccc$</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 c b_1 b_2 c$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 B_12 B_22 ccc$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 B_12 B_22 B_12 B_22$</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>$b_1 b_2 b_3 ccc$</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 b_3 B_13 B_23 B_33$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 b_3 B_13 B_23 c$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 b_3 c b_1 b_2$</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>$b_1 b_2 b_3 b_4 ccc$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$b_1 b_2 b_3 b_4 B_14 B_24$</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>$b_1 b_2 b_3 b_4 b_5 c$</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>$b_1 b_2 b_3 b_4 b_5 b_6$</td>
<td>1</td>
</tr>
</tbody>
</table>
theory also allows for strands shorter than I within one sheet. In theory, the value of I should be equal to the value of n since there is no reason to assume that a sheet of one continuous strand cannot exist. In fact, as I approaches n, a better approximation to the true value of Z is calculated. If an algorithm is available for computation of a property of the chain as a function of I, the interest is in the limit where I goes to n.

In order to develop the algorithm for calculating the means square end-to-end distance, it is easiest to define the matrices for I = 2, and then increase the value of I. A table and matrix can be constructed for a chain of any length with I = 2 (Table 4). This table defines all the potential statistical weights of residue i as they relate to residue i - 1. In the first row one can see that if residue i - 1 is a random coil, and residue i is also a random coil the statistical weight is 1. If residue i is the first residue in the first strand of a β-sheet, its statistical weight is rt. If residue i - 1 is a random coil, it is impossible for residue i to be the second residue in the first strand of a β-sheet, or to be in the second strand of a β-sheet. For this reason these statistical weights are set to zero. The blank entries in Table 4 are zeros. Similarly, if residue i - 1 is the first residue in the first strand of a β-sheet, the only conformation residue i can have is that of the second residue in the first strand. (Strands of one residue are
Table 4.—Sheet-Coil Transition Statistical Weights

<table>
<thead>
<tr>
<th>conformation of residue i - 1</th>
<th>conformation of residue i</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>b₁</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
</tr>
<tr>
<td>b₁</td>
<td>(\tau t)</td>
</tr>
<tr>
<td>b₂</td>
<td>(\delta t)</td>
</tr>
<tr>
<td>B₁₂</td>
<td>(t)</td>
</tr>
<tr>
<td>B₂₂</td>
<td>(t)</td>
</tr>
</tbody>
</table>

not permitted by definition.)

The conformation partition function is a scalar value, not a matrix. In this formulation the first matrix is the first row of the square matrix and the last includes selected values of the first column of the square matrix. A more appropriate means of dealing with this row and column is to have an initial row and final column in which the values are all '0' and '1'. The initial element of the initial row is a one followed by a range of zeros. Whether the value of each element in the final column is zero or one is justified as follows. The hypothetical residue that follows the last residue of the chain cannot be the first residue in a new strand, that is it can not be a \(b₁\) or a \(B₁x\). The final column contains zeros in the rows that correspond to these conformations.

The value of \(Z\) can be defined as

\[ Z = \text{row}(1,0,0,0,0) \cup^n \text{col}(1,0,1,0,1) \quad (17) \]
where
\[
U = \begin{bmatrix}
1 & r_t & r_t \\
1 & r_t & \delta t \\
1 & \delta t & t \\
1 & \delta t & t \\
\end{bmatrix}
\]

If \( I = 3 \) the table and the matrix expand. Now the conformation partition function can be defined as follows.
\[
Z = \text{row}(1, 0) \ U^n \ \text{col}(1, 0, 1, 1, 0, 0, 1, 1, 1) \quad (18)
\]

where
\[
U = \begin{bmatrix}
1 & r_t & r_t & \delta t & t \\
1 & r_t & \delta t & t \\
1 & \delta t & t \\
1 & \delta t & t \\
\end{bmatrix}
\]

Conformation labels for the rows and columns are \( c, b_1, b_2, b_3, B_{12}, B_{13}, B_{22}, B_{23}, \) and \( B_{33} \) (in order from top to bottom and left to right).

This formulation can be expanded to allow for randomly coiled loops between strands of a single sheet. These have been named interstrand loops (Mattice 1985). One means of including loops requires expanding the \( U \) matrix beyond its current size. A simpler method is to insert elements into the current \( U \) matrix. The positions of these elements are defined by
\[
i = j = I + 2, I + 3, \ldots, 2I. \quad (19)
\]
The effect of this insertion is to allow the possibility
for a loop of any length to take off from and return to the position of the first residue of a new strand.

This rather involved method of determining $Z$ is more convincing when the same value is determined by another means which may be more understandable. Given a chain four residues long, one can draw sticks and dots or use the single letter notation to determine all possible conformations. Because this chain is only four residues long, no interstrand loops can form. A sheet with two strands of the shortest allowable length consumes all the residues. For this reason, interstrand loops are excluded from this example. The statistical weight of each of these can be assigned according to the definitions in the preceding paragraphs (Figure 7). These are then added to give the value of $Z$ in terms of $r, δ, f,$ and $t$. If the appropriate matrixes are multiplied together, the same result is attained.

The dimensions of $U$ can be defined by the value of $J$ where $J = I(I + 3)/2$. As the value of $I$ increases the size of the $U$ matrix increases. It is easy to see that such a huge matrix quickly becomes unmanageable by computers of reasonable size. Since many of the elements of $U$ are zeros, it is not practical to store the entire matrix. Rather than building the matrix, the location of nonzero elements is defined in terms of $I$, stored, and called upon for matrix multiplication. The positions of nonzero elements are defined in Appendix IV. Use of them will be
\[ Z = 1 + (3\tau^2 t^2) + \tau^2 \delta t^4 \]

*Figure 7. Determining Z for the sheet coil transition when n = 4, I = 2.*
described more fully in Chapter Three.

This discussion presents a means of calculating the conformation partition function for the antiparallel intramolecular $\beta$-sheet to coil transition for very large values of $n$ and $I$ where $n$ is the number of residues. Having a means of calculating $Z$ opens possibilities for determining the value of other parameters which require a knowledge of the conformation partition function. One example of this is the fraction of residues in a given conformation in the chain. Another example is the mean square end-to-end distance. Formulations for each of these are described in the next chapter.
CHAPTER THREE
THE MEAN SQUARE END-TO-END DISTANCE

A calculation of the mean square end-to-end distance averaged over all possible conformations in the \( \beta \)-sheet to coil transition can be accomplished by making use of the value of \( Z \) determined in the manner described in the preceding chapter. However, computing the mean square end-to-end distance requires defining and manipulating another matrix. The nature of this matrix multiplication is, as in the previous discussion, based on the concept of one matrix for one residue. The order of the matrices is directly related to the order of the residues in the chain. Although calculations in the present application are done on homopolypeptides— that is, the statistical weights are the same for all residues in a chain— this formalism allows the possibility of calculations on polypeptides in which the statistical weights for each residue are unique.

Defining Matrices

The eventual goal of this section is to develop and describe the matrix needed to calculate the mean square end-to-end distance, the \( F \) matrix. This matrix is closely related to the direct product of the statistical weight matrix, \( U \) and a generator matrix, \( G \), for each statistical weight. It differs from the simple direct product only in
that the specific forms of $G$ depend on $u_{ij}$, where $u_{ij}$ is an element of $\mathcal{U}$. The $G$ matrix as defined by Flory (1974) contains the transformation matrix, $T$. The $F$ matrix contains information about both the geometrical conformation of virtual bonds and the statistical weight of each conformation.

**Transformation Matrix**

If one wishes to describe a vector for the end-to-end distance of a chain, one must choose a frame of reference. The frame of reference chosen here is the coordinate system of the first bond (Flory 1969). Clearly, this is not consistent throughout the chain, but the coordinate system of bond $i + 1$ can be transformed into the coordinate system of bond $i$. The matrix which describes the transformation required to accomplish this is

$$
T = \begin{bmatrix}
\cos \theta & \sin \theta & 0 \\
\sin \theta \cos \phi & -\cos \theta \cos \phi & \sin \phi \\
\sin \theta \sin \phi & -\cos \theta \sin \phi & -\cos \phi
\end{bmatrix}
$$

$T$ can be interpreted by referring to Figure 8. Theta is the supplement of the angle between the bonds. Phi, the dihedral angle about bond $i$, is defined as zero for the trans conformation, and increases with right handed rotation. The $x$ axis of the Cartesian coordinate system for bond $i$ is along the bond and is positive in the chosen direction of propagation. The $y$ axis is in the plane of $x_i$ and $x_{i-1}$ perpendicular to $x_i$. Its positive direction is that which yields a positive projection on $x_{i-1}$. The $z$
Figure 8. Definition of the coordinate system for a virtual bond. The dihedral angle, $\phi$, is zero for the trans conformation and is positive for rotation that is clockwise when looking in the direction of the progression of the chain. Theta is the supplement to the bond angle. The z axis is that which completes a right-hand coordinate system (from Flory 1969).
axis is that which completes a right hand coordinate system.

**The Generator Matrix**

In the case of polypeptides, the distance transversed by one residue is best thought of in terms of the virtual bond. The virtual bond can be described as the line connecting the alpha carbon of residue $i$ to the alpha carbon of residue $i + 1$. Let this bond be defined as

$$ l_i = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} $$

in the $i$ coordinate system, and

$$ = T_{i-1}l_i $$

in the $i - 1$ coordinate system.

The end-to-end vector is the sum of all $l_i$ if they are in the same coordinate system. Since they are not when each one is written as shown in Equation (21), the coordinate system of each virtual bond is transformed into the coordinate system of the previous bond.

$$ r = l_1 + T_{1}l_2 + T_{1}T_{2}l_3 + \ldots + T_{1}T_{2} \ldots T_{N-1}l_N. $$

$N$ is the number of virtual bonds.

The value for the square of the end-to-end distance of a polymer in a specific conformation is

$$ r^2 = r \cdot r = l_1 \cdot l_2 + l_2 \cdot l_2 + \ldots + l_N \cdot l_N + 2(l_1 \cdot l_2) + \ldots. $$

In this equation, the dot product is a means of accounting for the fact that the various virtual bonds ($l_i$) are
pointing in a variety of directions. It is a means of transforming them all into the same coordinate system. In other words,

$$l_i \cdot l_j = l_i T_i T_{i+1} \cdots T_{j-1} l_j$$  \hspace{1cm} (25)$$

Then $\langle r^2 \rangle$, where the brackets indicate the average over many conformations, can also be written as

$$\langle r^2 \rangle = \sum_{i \leq N} 1_i' + 2 \sum_{i < j \leq N} l_i T_i T_{i+1} \cdots T_{j-1} l_j$$  \hspace{1cm} (26)$$

If the conformation at each residue is independent of the conformation at any other residue

$$\langle r^2 \rangle = \sum_{i \leq N} 1_i + 2 \sum_{i < j \leq N} l_i T_i T_{i+1} \cdots T_{j-1} l_j$$  \hspace{1cm} (27)$$

Matrices can be developed that will accomplish this summation when they are multiplied serially. The equation becomes

$$\langle r^2 \rangle = \begin{bmatrix} 1 & 2 l_i T_{i} & 1_i' \\ \\ 1 & 2 l_2 T_{2} & 1_2' \\ \\ 0 & 0 & 1 \end{bmatrix} \cdots$$  \hspace{1cm} (28)$$

$$\begin{bmatrix} 1 & 2 l_{N-1} T_{N-1} & 1_{N-1}' \\ \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1_{N-1} \\ 1_{N-1} \end{bmatrix}$$

Multiplying these matrices yields the same result as writing out the summation in Equation (27).

These five by five matrices are the generator or $G$ matrices. Equation (27) can be rewritten as

$$r^2 = G_1 G_2 \cdots G_{n-1} G_n.$$  \hspace{1cm} (29)$$

where $G_1$ indicates the first row of $G_1$ and $G_n$ indicates the final column of $G_n$. Equation (29) is for a polymer with a specific conformation, i.e., each $T$ in Equation (28) is for a specified conformation of the indicated bond.
If the residue conformations in the polymer are interdependent, the statistical weight of each conformation is then multiplied by the mean square end-to-end distance for each specific polypeptide conformation and all of these multiples are added, the sum being divided by the conformation partition function. In equation form,

\[ <r^2> = \frac{u(a)r^2(a) + u(b)r^2(b) + u(c)r^2(c) + \cdots}{u(a) + u(b) + u(c) + \cdots} \quad (30) \]

where the parenthetical variables for u and r^2 denote unique conformations and r^2 is defined in Equation (29).

The G's differ from each other by the different transformation matrices required for different conformations and the corresponding virtual bond lengths.

**The F Matrix**

The summation of statistical weights for the determination of the value of Z can be accomplished with a matrix multiplication method already described. This approach is also useful for the summation of statistical weights multiplied by their corresponding generator matrices, G. Consider the example of the statistical weight matrix for an α-helix for interdependent bonds where the statistical weight of the conformation at bond i depends on the conformation at bond i - 1. Then

\[ Z = \begin{bmatrix} 1 & 0 \\ 1 & s \end{bmatrix} \begin{bmatrix} 1 & s^N \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (31) \]

To calculate \( <r^2> \) the generator matrix for a
given conformation is multiplied by the statistical weight for that conformation.

\[
<r^2> = z^{-1} \begin{bmatrix} 1 & 0 \\ G_c & sG_h \end{bmatrix} \begin{bmatrix} g_n \\ g_n \end{bmatrix}^N \tag{32}
\]

where \( G_c \) is the generator matrix for a random coil and \( G_h \) is the generator matrix for a helix. This equation also defines the \( F \) matrix.

\[
F = \begin{bmatrix} G_c & sG_h \\ G_c & sG_h \end{bmatrix} \tag{33}
\]

Sheet-Coil Transition

In order to build \( F \) for the \( \beta \)-sheet coil transition, one must first define the transformation and generator matrices for a sheet. The nature of this formulation is to project the three-dimensional structure of the sheet onto a two-dimensional plane. Even more strictly, the progress of the chain is limited to its reflection onto a straight line that best defines its path. In essence, the sheet is created by a strand which travels in a straight line, and bends which are two right-angle turns oriented so as to send the line back in the same direction from which it came, parallel to the previous line (that is, parallel to the previous strand).

Defining the Transformation Matrices

To determine the transformation matrices necessary to accomplish this feat, a simple model is presented. The illustration in Figure 9 is within the restrictions defined in the preceding paragraph. It has
Figure 9. A chain in which every residue is in a $\beta$ sheet. This sheet adheres to restrictions defined in the text. Each dot represents an $\alpha$ carbon and each connecting line is a virtual bond of $l = 1.0$. 
the added restriction that all virtual bonds are

\[ l_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \]  \hfill (34)

This restriction on \( l \) will be removed later. Clearly, the end-to-end distance of this sheet can be defined as a vector which travels two units in the direction of the original virtual bond, and four units in a direction perpendicular to that. Since the sheet is limited--by definition--to the plane of the paper, the vector of the end-to-end distance is

\[ r = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} \]  \hfill (35)

The equation for calculating \( r \) is

\[ r = l_1 + T_1 l_2 + T_1 T_2 l_3 + \ldots + T_1 T_2 \ldots T_{14} T_{15} l_{16}. \]  \hfill (36)

Now it is necessary to determine what transformation matrices will give the correct value of \( r \). Looking back to the definition of the transformation matrix, it is apparent that values for \( \theta \) and \( \phi \) must be known. In this model, the choices are limited to

\[ \theta = 0^\circ \text{ or } 90^\circ \]

\[ \phi = 0^\circ \text{ or } 180^\circ \]

This problem has been solved for sheets of various geometries. From these calculations, it is found that the two bend matrices can be defined as
\[ T_{b1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \phi = 0^\circ, \ \theta = 90^\circ \] (37)

and

\[ T_{b2} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \phi = 180^\circ, \ \theta = 90^\circ \] (38)

\( T_{b1} \) is always immediately followed by \( T_{b2} \).

The matrix which sends the chain along in a straight line, the strand matrix, is a 3 x 3 identity matrix.

\[ T_s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \phi = 180^\circ, \ \theta = 0^\circ \] (39)

The matrix chosen for those residues in a random configuration is not defined by Figure 9. Instead the transformation matrix used is one determined for random poly (L-alanine) by Brant, Miller, and Flory (1967).

### Constructing The Generator Matrices

The necessary generator matrices can now be constructed. The distance progressed by each residue in one strand is determined from the distance between two \( \alpha \) carbons in a strand in one dimension as determined by x-ray diffraction studies on poly (L-alanine) (Fraser 1973). The values in Table 1 (page 9) as dimension 'v' are 0.383 and 3.828. The difference is 3.44 Å. The distance traversed between strands as the bend is made is the distance between strands of a poly (L-alanine) sheet, 4.73 Å. This number is the average of the distances between the two sets of \( \alpha \)
carbons in the up and down chains in dimension 'u'; more specifically, \[\frac{(7.235 - 2.225) + (6.955 - 2.505)}{2}\].

The interstrand loop is handled by taking the chain to the position at which the next strand is to start and then going nowhere for as many residues as are in the loop. When determining the end-to-end distance, the point of interest is the distance covered by the chain. Since the definition of a loop is a series of residues that take off and then come back to where they started from, determining the end-to-end distance of a loop of \(n\) residues is the same as going nowhere for \(n\) residues. The generator matrix for an interstrand loop is simply a 5 x 5 identity matrix.

Solving the problem of the interstrand loop is not so easy if one must keep track of what is actually happening in the loop. For example, in order to determine the mean square radius of gyration, one must know the location of each of the residues. Another example is the calculation of the dipole moment. In this calculation one must know the orientation of the dipole moment of each of the residues. In these two instances, a more precise loop generator matrix must be defined that allows random motion with the restriction that the final residue in the loop have the same location as the initial residue in the loop. The \(G\) matrices used for \(<r^2>\) are written out in Figure 10.

Building the \(F\) matrix

The next step in this development is the building
Figure 10. The generator matrices used in calculations.
of the $F$ matrix. Before proceeding any further it is necessary to define which generator matrix goes with which statistical weight. The $G$ matrix defines the position of the $\alpha$ carbon and points the direction the bond to the next $\alpha$ carbon will take. Figure 11 illustrates a chain with a defined conformation. The first column lists the letters that define the conformation of each residue. The second column lists the statistical weights that have been assigned for each of these conformations. The third column lists the transformation that the coordinate system must undergo in order to define the drawn conformation. The fourth column is how the program actually defines these transformations.

Everything is defined in terms of the symbol in column 1. Since one cannot predict the appearance of a bend from the conformation labelled 'b', the fact that it has happened is defined by the conformation labelled $B_{1j}$. This means that all the transformations are actually shifted one residue. In the same vein, one cannot predetermine when the sheet will end, so there appears to be an extra 'sheet' transformation. This is handled by assigning the first residue of the sheet, $b_1$, a random coil transformation matrix. The resulting $F$ matrix is written out in Figure 12.

In order to determine the final column for $F$, one must take into account which conformations are allowed for the last element of the chain as well as which element of
Figure 11. A chain with a defined conformation. This model is used to define which generator matrix should go with which statistical weight in the program. See the text for discussion.
<table>
<thead>
<tr>
<th>conformation</th>
<th>statistical weight</th>
<th>$T$</th>
<th>$T$ in program</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$b_1$</td>
<td>$tt$</td>
<td>$T_s$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$b_2$</td>
<td>$tt$</td>
<td>$T_s$</td>
<td>$T_s$</td>
</tr>
<tr>
<td>$b_3$</td>
<td>$tt$</td>
<td>$T_s$</td>
<td>$T_s$</td>
</tr>
<tr>
<td>$b_4$</td>
<td>$tt$</td>
<td>$T_b1$</td>
<td>$T_s$</td>
</tr>
<tr>
<td>$B_{14}$</td>
<td>$\delta t$</td>
<td>$T_b2$</td>
<td>$T_b1$</td>
</tr>
<tr>
<td>$B_{24}$</td>
<td>$t$</td>
<td>$T_s$</td>
<td>$T_b2$</td>
</tr>
<tr>
<td>$B_{34}$</td>
<td>$t$</td>
<td>$T_b1$</td>
<td>$T_s$</td>
</tr>
<tr>
<td>$B_{13}$</td>
<td>$\delta t$</td>
<td>$T_b2$</td>
<td>$T_b1$</td>
</tr>
<tr>
<td>$B_{23}$</td>
<td>$t$</td>
<td>$T_b1$</td>
<td>$T_b2$</td>
</tr>
<tr>
<td>$B_{12}$</td>
<td>$\delta t$</td>
<td>$T_b2$</td>
<td>$T_b1$</td>
</tr>
<tr>
<td>$B_{22}$</td>
<td>$t$</td>
<td>$T_s$</td>
<td>$T_b2$</td>
</tr>
<tr>
<td>$b_3$</td>
<td>$tt$</td>
<td>$G_c$</td>
<td>$T_s$</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>$G_c$</td>
<td>$G_c$</td>
</tr>
</tbody>
</table>
Figure 12. The $F$ matrix for $I = 3$. Dimensions are $5J \times 5J$, 45 x 45. Also shown are the appropriate initial row and final column (dimensions are $1 \times 5J$ and $5J \times 1$, respectively). The subscript 'n' on the zeros indicates a field of $n$ zeros.
the final $G$ matrices retain information about the mean square end-to-end distance. Multiplication of the specific $G$'s for a defined conformation leads to the conclusion that the last column of the last $G$ matrix must be retained. The analysis required for calculating $Z$ has already revealed which conformations are allowed in the final position. These facts combine to yield the following definition for the last row where $u_i$ indicates the $i$th element of the row.

\begin{align*}
    u_5 &= 1 \\
    u_i &= 1 \quad \text{where } i = 5m + 5 \\
    &\quad \quad 2 < m \leq I \text{ and } 2I < m \leq J \\
    \text{all other } u &= 0.
\end{align*} \hspace{1cm} (40)
CHAPTER FOUR

COMPUTATION

In order to calculate the mean square end-to-end distance, one must be able to calculate the conformation partition function. This is true because the equation for the mean square end-to-end distance includes the conformation partition function.

\[ <r^2> = Z^{-1} \text{(row)} F^n \text{(column)} \]  \hspace{1cm} (41)

In applying this type of calculation to the problem for a \( \beta \)-sheet, recall that the \( U \) matrices very rapidly increase in size with increasing values of \( I \). The dimensions of the \( U \) matrices are \( J \times J \) where \( J = I(I + 3)/2 \). The dimensions of \( F \) are \( 5J \times 5J \). The dimensions of the rows are \( 1 \times J \) for calculating \( Z \) and \( 1 \times 5J \) for calculating \( F \) in Equation (41). Very large square matrices in which many elements are zero result from realistic values of \( I \). Since it is impractical either to store or manipulate these huge matrices, the program operates by starting with only the initial row matrix. The elements of this row are multiplied by their corresponding nonzero elements one column at a time. The columns are never really built, but rather, their nonzero elements are described by the series of formulae listed in Appendix IV.

In essence, the program sequentially generates a
series of successive rows, multiplying the old row onto the
next matrix to result in a new row. This final row is
multiplied by a column of 1's and 0's to result in a scalar
value.

For large values of \( n \), the values calculated for
\( Z \) can become very large. The purpose here is not to define
an exact value for \( Z \) but to define values which require the
differentiation of \( \ln Z \). For this reason, one can assign a
value for the statistical weight of the random coil (see
Equation 10) and define the weights of every other
conformation relative to that value. In the discussion
thus far, this value has been '1'. There is no reason one
cannot alter all the statistical weights by a weighting
factor other than '1'. From here on, this weighting factor
shall be 'wtran'. The value of wtran is set such that the
final calculated value of \( Z \) is manageable by the computer.

The Conformation Partition Function

A program has been written in the 'C' programming
language to calculate the value of \( Z \) (Appendix Five, pp.
221 - 223). This program has been tested in a variety of
ways. In the first test, all statistical weights are set
to unity. In this situation, the program should simply
calculate the total number of conformations possible.
Results are shown in Table 5. The table lists the results
of the program for values of \( N \) from 2 to 10 and for values
of \( I \) from 2 to 10. Recall that \( I \) is the maximum number of
\( \alpha \) carbons that can be assigned to one strand of a sheet.
The value of $n$ is the number of $\alpha$ carbons in the chain. A chain of 2 residues, $n = 2$ can have two conformations. The values of $\theta$ and $\phi$ can be 1) those required for a $\beta$-sheet or 2) anything else, i.e., a 'c' residue. A chain of three residues can have one of three conformations if $I = 2$ and one of four conformations if $I = 3$. Some of these conformations are illustrated in Figure 13.

To test the program when statistical weights are not set equal to one, the matrix multiplication the program is intended to do is determined rigorously with a hand-held calculator for small values of $I$. When the program is set

<table>
<thead>
<tr>
<th>Table 5.—Values of $Z$ when $\tau = \delta = t = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

to print each row as it is produced, its results can be compared easily to results found with a hand calculator. Table 6 demonstrates one such calculation for $I = 2$, $t = 2.0$, $\tau = 0.1$, $\delta = 0.5$, and $n = 2$ to $n = 8$. From this table it is apparent that the program runs as predicted.
Figure 13. Possible conformations of short chains. Each dot represents an α-carbon. A straight line connecting two dots signifies that the two dots are in a sheet conformation.
Table 6.—The Conformation Partition Function Calculated by Hand and by Computer: \( \tau = 0.1, \delta = 0.5, t = 2.0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>newrow*</th>
<th>( Z(\text{calc}) )</th>
<th>( Z(\text{comp}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[1.00 0.20 0.00 0.00 0.00]</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>[1.00 0.20 0.04 0.00 0.00]</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>3</td>
<td>[1.04 0.20 0.04 0.04 0.00]</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>4</td>
<td>[1.08 0.21 0.04 0.04 0.08]</td>
<td>1.20</td>
<td>1.20</td>
</tr>
<tr>
<td>5</td>
<td>[1.20 0.22 0.042 0.12 0.08]</td>
<td>1.32</td>
<td>1.32</td>
</tr>
<tr>
<td>6</td>
<td>[1.32 0.24 0.043 0.12 0.24]</td>
<td>1.60</td>
<td>1.60</td>
</tr>
<tr>
<td>7</td>
<td>[1.60 0.26 0.048 0.28 0.24]</td>
<td>1.90</td>
<td>1.90</td>
</tr>
<tr>
<td>8</td>
<td>[1.90 0.32 0.053 0.29 0.57]</td>
<td>2.52</td>
<td>2.52</td>
</tr>
</tbody>
</table>

*Calculated as follows: newrow = row \* U
For \( n = 1 \) row = [1.00 0.00 0.00 0.00 0.00]
else row\(_n\) = newrow\(_{n-1}\)

\( Z(\text{calc}) = u_1 + u_3 + u_5 \). The symbols \( u_1 \), \( u_3 \), and \( u_5 \) represent the first, third, and last elements of newrow.

\( Z(\text{comp}) = \) value determined by the computer
One additional means of testing the program involves calculating $Z$ by a very different method and comparing results. One can write out all possible conformations for chains of a given length for all possible values of $I$. This is done for values of $n$ up to 5 and $I = 2, 3, 4, \text{ and } 5$. The statistical weight of each possible conformation is determined and multiplied by the appropriate degeneracy. Two sets of statistical weights are used to calculate the partition function. These same values are then put into the program. The agreement is very good.

The Mean Square End-to-End Distance

Once a method of calculating $Z$ has been found, the same general formulae can be used to calculate the numerator expressed in Equation (41) for $<r^2>$. If one imagines a third and fourth dimension to the large array, one can see that the same formulae can be used, but within each call for multiplication of matrix elements, there must be an inner loop that includes the appropriate column of the $G$ matrix. Since each $G$ matrix is a five by five, the dimensions of the original row are five times the dimension of the original row for the calculation of $Z$, that is $5*J$.

Modifications

There are a number of modifications required for the general construction of the equation for $<r^2>$. In the formulation of $Z$, $n$ is defined as the number of alpha carbons. However, in the formulation of the $F$ matrix the
number of transformations, which is equal to the number of virtual bonds between α carbon atoms, is used. Since each transformation takes the chain a distance and defines in what direction the next bond will point, after one matrix multiplication, one bond has been defined. In other words, the power of the F matrix is N, the number of virtual bonds, where \( N = n - 1 \). If one tries to define \( n = N \), the F matrices calculate the average end-to-end distance for one bond too many.

Since the number of virtual bonds will always be one less than the number of alpha carbons it may seem that the best approach to solving this problem is to simply multiply the F matrices by each other \( N - 1 \) times and the statistical weight matrices by each other n times. However this results in a difference in statistical weights in the numerator and the denominator. The best way to account for this discrepancy is by defining the length of an imaginary Nth bond in the chain to be zero. In this way, the statistical weights are constant in numerator and denominator and the calculated length is not affected.

The effect of this is that an "unallowable" conformation is allowed. According to the tables defined in the previous chapter, a residue in a position defined as \( B_{1x} \) must be followed by a residue in a position defined as \( B_{2x} \). The conformation shown in Figure 14a is allowed while the conformation in 14b is not. However, if the length of the last bond is set equal to zero, the conformations shown
Figure 14. End effects. (a) Expected allowable conformations. (b) Conformations that become allowable when the length of the Nth bond is set equal to zero. (c) Conformation constructed by the F matrix in place of corresponding conformations illustrated in (a) due to the 'begin' effect.
Figure 14. (cont.)
in Figure 14b are allowed.

There is one additional modification required. The construction of the statistical weight matrix, \( U \) is such that when the initial row is multiplied onto \( U \) the statistical weight, \( \delta \), is multiplied by zero for the first four cycles. That is, bends are not figured into the end-to-end distance until after the fourth row-onto-matrix multiplication. For this reason when the statistical weights favor sheet formation, the conformations actually drawn by the computer cannot have a bend before the fourth residue. When \( t \) is set very high, the end-to-end distances calculated by the program agree well with those shown in Figure 14b. This affects the radius greatly when \( N \) is small, but less so as \( N \) goes to infinity. This effect shall be referred to as the 'begin' effect. The results of the begin effect will be discussed later.

**Testing the Program**

The program to determine mean square end-to-end distance has been written and tested. A listing of the program is given in Appendix V. The final three pages are the program to calculate \( Z \), since this value is required for determining the value of \( \langle r^2 \rangle \). In the program are generator matrices with transformation matrices for every potential transformation required for a \( \beta \)-sheet and for a random coil. As with the calculation of \( Z \), the program was tested for small values of \( n \) and \( I \) to demonstrate that it does what it is intended to do.
In the first test, all the transformation matrices are set to the identity matrix, and the bond length is set to 2.0. In this case, the only conformation allowed is that of a rod. There ought to be no dependence on values of \( r, \delta, \) or \( t \) since there is only one conformation allowed. The program was run for values of \( n \) from 4 to 11. As can be seen from Table 7, the value of \( <r^2> \) is always \([2.0 \times (n - 1)]^2\).

To determine if the bend matrices are working as predicted, the value of \( \delta \) is pushed very high and the penalty accorded to \( r \) is increased. Also, all residues are encouraged to be in a sheet by increasing \( t \). In this case, one would suspect the conformation illustrated in Figure 14a and 14b. In fact, the program calculates end-to-end distances more in agreement with the conformation.

### Table 7.—Mean Square End to End Distance for a Straight Line Transformation

<table>
<thead>
<tr>
<th>( I )</th>
<th>( n )</th>
<th>( &lt;r^2&gt; )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau = 1.0, \delta = 0.1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>400</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>400</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>400</td>
</tr>
<tr>
<td>( \tau = 0.1, \delta = 1.0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>400</td>
</tr>
</tbody>
</table>

\( \tau = 2.0, t = 2.0 \)
illustrated in Figure 14c. The reason for this difference is mentioned in a preceding paragraph.

The begin effect is most problematic at small values of n, and is minimized for increasing n. For even n, the begin effect can be quantitatively defined. For the instance in which all virtual bond lengths are set to two, the statistical weights are set to favor the formation of a sheet of several strands all of which are two residues long, e.g., I = 2, t = 40, δ = 2.0, f = 0.0, and r = 0.1. One can draw the structures expected from this set of statistical weights and the structures computed due to the begin effects. All the expected structures have two residues in the first strand and all the computed structures have four residues in the first strand. For a virtual bond length of two, the following equation accurately describes the ratio of these two end-to-end distances as a function of even values of N.

\[
\frac{\langle r^2 \rangle (\text{expected})}{\langle r^2 \rangle (\text{computed})} = \frac{\left[ \frac{(n-2)}{2} \right]^2 + i}{\left[ \frac{(n-4)}{2} \right]^2 + (3 - i)^2}
\]

\[ i = 0 \text{ if } n \text{ is divisible by } 4 \]
\[ i = 1 \text{ if } n \text{ is not divisible by } 4 \]

The plot in Figure 15 demonstrates the degree of this difference as a function of N for even values of N. As 1/N goes to zero (N goes to infinity), the ratio of the end-to-end distance of a molecule with a conformation like that in Figure 14a to the end-to-end distance of a molecule with a conformation like that in Figure 14b goes to one.

When N is odd, understanding what is happening
Figure 15. The ratio of values expected (column (a) in Figure 14) to the values calculated by the program (column (c) in Figure 14). This is quantitatively described when $I = 2$, $n$ is even, and $l$ is set to be equal to 2.0 A (see Equation (42) in text).
becomes a bit more complex. This is because there is more
than one conformation of the same statistical weight that
the polymer can assume. In order to allow the formation of
the conformation shown in Figure 14a, one allows $I = 3$.
Setting $r$ to a small value and $s$ and $t$ to large values
causes this conformation to be the most probable. The
discussion in the previous paragraphs would lead one to
conclude that the conformations drawn by the program are
closer to those shown in Figure 14c. When $r$ is increased
without becoming greater than one, the conformations that
have their statistical weights defined in column a are
expected. The value of $<r'>$ calculated by the program
approximates the value expected from the conformations in
Figure 14c, i.e., structures numbered 8, 15, 16, and 17.
When $s$ is increased without becoming greater than one, the
conformations that have their statistical weights defined
in column b are expected. The value of $<r'>$ calculated by
the program approximates the distances defined by the
structures numbered 9 and 18.
CHAPTER FIVE
RESULTS OF CALCULATIONS

It has been demonstrated that the program listed in Appendix V calculates expected values of $\langle r^2 \rangle$ for carefully defined and limited conditions. Now, therefore, it is appropriate to discuss results obtained when variables are set to values indicative of more realistic conditions. For production run calculations the generator matrices used are those defined in Figure 10. The technique used to determine the transformation matrices has been discussed. Values used for virtual bond lengths are taken from the x-ray diffraction data on $\beta$ poly (L-alanine) (Fraser 1973).

An additional loop is incorporated into the program which allows the calculation of the value of $\langle r^2 \rangle$ for a range of $t$ with a fixed set of $\delta$ and $\tau$. The range of $t$ for each of the sets of $\delta$ and $\tau$ was chosen such that the value of $\langle r^2 \rangle$ began to plateau at high values of $t$. In most cases $t = 10$ is found to be sufficiently high.

The Value of $I$

One of the variables in the program is the value of $I$. While it is true that the ideal value for $I$ is approached as $I$ approaches $n$, this is not practicable and is often unnecessary for large values of $n$. The amount of
time required for the computer to calculate is greatly increased as \( I \) increases. Referring to the equations for the positions of nonzero elements in the \( U \) matrix in Appendix IV, one can see that the number of multiplications to be done increases rapidly with increasing \( I \). For this reason it is advisable to choose \( I < n \). Yet if the value of \( I \) is too small, the length of strands will be skewed toward a smaller value than would be calculated if \( I = n \).

One approach to this problem is to do calculations for increasing values of \( I \) until the value of \( \sqrt{\langle r^2 \rangle} \) no longer changes. At that value of \( I \) one can believe that the conformations allowed by the restriction set by \( I < n \) include those with the highest probability of existing. In this case \( I < n \) is no restriction. A difficulty with this is the time involved. If it were possible to estimate a value of \( I \) from which to start the approximation, the problem would be much more quickly solved.

Such a solution has been presented and defended by Mattice and Scheraga (1984b). An equation is given in terms of \( \delta \) and \( r \) to determine the value of \( I_\infty \) which is defined as the maximum number of residues per strand in that single sheet of \( n \) residues that has the largest statistical weight as \( t \) becomes infinite. The equation used is

\[
I_\infty = [(\ln \delta / \ln r) n]^{0.5}.
\]  

(43)

This equation is not helpful when \( \delta = r \) or when either is
equal to one. In these cases starting estimates were made by slightly decreasing $\delta$ and/or slightly increasing $r$ to make them work in the equation. These adjustments result in a larger number for the estimate of $I$ and are therefore reasonable approximations.

When a starting value of $I$ is found, calculations are done at values of $t$ which are increased until the value of $\langle r^2 \rangle$ plateaus. Calculations with $I$ increased by five are done at the highest values of $t$. This is repeated until a plot of $\langle r^2 \rangle$ against $t$ overlaps for two values of $I$ that differ by five. The lower value of $I$ is used in subsequent calculations. An example of this is shown for $r = 0.3$, $\delta = 0.9$, and $n = 300$ in Figure 16. When these values are put into Equation 43, $I_0$ is 5.12. Plotting $\langle r^2 \rangle$ against $t$ shows a change in the values of $\langle r^2 \rangle$ from $I = 5$ to $I = 10$, but no additional difference when $I = 15$ is used for the calculation of $\langle r^2 \rangle$ (the crosses and diamonds are indistinguishable). For this set of statistical weights, $I = 10$ is used in all calculations. This same analysis was used for every set of statistical weights discussed in this paper.

**Values of Statistical Weights**

Results from calculations on eight sets of $\delta$ and $r$ are reported. These can be divided into two groups. Each of the groups includes one set in which the value of $\delta$ and $r$ are equal and sets in which one of the two is varied from this value. Using such values it should be possible
Figure 16. The effect of increasing $I$. Shown is data for $n = 300$, $r = 0.3$, $\delta = 0.9$, $I = 5$ (□); $I = 10$ (+); $I = 15$ (●).
to determine something about the effect of varying the statistical weight for one of the edge effects while holding the other constant.

One of the groups is meant to illustrate two extreme possible values. In this group the sets are: 1) \( \delta = 1.0, r = 1.0; \) 2) \( \delta = 1.0, r = 0.1; \) and 3) \( \delta = 0.1, r = 1.0. \) The set in which both statistical weights are equal to one demonstrates the situation in which no architecture for the sheet is favored. Whether or not a sheet forms depends solely on the value of \( t \) because \( \delta \) and \( r \) have no effect on the statistical weight of the sheet. Setting one of these equal to a much smaller value has the effect of penalizing either the first strand (if \( r = 0.1 \)) or each bend (if \( \delta = 0.1 \)). The sheets that form will have a shape that depends on the relative value of these two statistical weights. When \( \delta \) is very small and bends are penalized, it is suspected that a few very long strands will make a sheet. When \( r \) is very small it is suspected that many short strands will make the sheet, minimizing the number of residues with no corresponding residues in the preceding strand.

The second group has five sets of \( \delta \) and \( r \) and includes 1) \( \delta = r = 0.3; \) 2) \( \delta = 0.3, r = 0.6; \) 3) \( \delta = 0.3, r = 0.9; \) 4) \( \delta = 0.6, r = 0.3; \) and 5) \( \delta = 0.9, r = 0.3. \) This is similar to the previous set in that the values of \( \delta \) and \( r \) are equal in one case and the statistical weights are varied one at a time in the other cases. Moreover, this
group avoids the special case in which neither $\delta$ nor $\tau$ penalize sheet formation. It is also possible to do some analyses that cannot be done for the more extreme penalties of $\delta = 0.1$ ascribed in the first group. It is expected that the trends seen from calculations done on this group parallel trends seen in the first group.

**Calculation of Other Parameters**

A greater understanding of the architecture of the sheet can be obtained if certain parameters other than the mean square end-to-end distance are known. These include the fraction of residues in the sheet, the average number of residues in a sheet, the average number of residues per strand, and the average number of strands in a sheet. These values can be calculated by the formula stated in the previous chapter. They were computed by the program used for data published by Mattice and Scheraga (1984a; 1984b; 1985).

**Group I: $\delta, \tau = 1.0, 0.1$**

When comparing the mean square end-to-end distance of an $\alpha$-helix with the mean square end-to-end distance for a random coil of the same number of residues, it is found that small helices have a smaller end-to-end distance than a random coil of the same number of residues. As the number of residues increases the end-to-end distance of an $\alpha$-helix is larger than for a random coil of the same number of residues. This is understandable since the helix is like a rod when it has many residues in it, the rod
having a greater end-to-end distance than an average random coil. Small helices, on the other hand, force the chain into a very compact space in a very ordered structure.

It is interesting to consider the end-to-end distance for a chain in a random coil conformation as compared to the end-to-end distance for an average $\beta$-sheet which contains the same number of residues. A comparison of these distances for the $\beta$-sheet and random coil is reported when $\delta = r = 1.0$. The program is run as a function of $n$ for $n \leq 60$. The value of $t$ is fixed at 0.0001 to favor the random coil conformation. The value of the end-to-end distance for one of the average sheets (as opposed to the end-to-end distance for the chain of 300 with many such average sheets) is calculated by the following equation.

$$\left\{( (<\text{residue/strand}>)_\beta - 1 \right\} \cdot 3.4 \AA^2 + \left\{( (<\text{strand/sheet}>)_\beta - 1 \right\} \cdot 4.7 \AA^2 \right\}^{1/2}$$

(44)

This is plotted against the corresponding average number of residues per sheet which is corrected for the number of residues in interstrand loops. The correction is made according to the equation

$$<\text{residues/sheet}>_\beta - ((<\text{strands/sheet}>)_\beta - 1) \cdot <\text{residues/loop}>_\beta$$

(45)

The resulting plot is shown in Figure 17. Although the two lines do not cross, it appears that they will do so. At some point a planar cross $\beta$-sheet will have a larger end-to-end distance than a random coil of the same number.
Figure 17. A comparison of the mean square end-to-end distance as a function of \( n \) for a random coil (□) and the average sheet (+) when \( \delta = \tau = 1.0 \).
of residues because, for the sheets in Figure 17, $\sqrt{\langle r^2 \rangle}$ is very nearly proportional to $n$.

Table 8 summarizes the results of all calculations done for a chain of 300 residues with $\delta = 1.0$, $r = 1.0$ as $t$ is varied. Table 9 lists the same results when $r$ is much less than 1.0 and Table 10 lists results when $\delta = 0.1$ and $r = 1.0$. These tables shall be referred to in the following paragraphs.

When $\delta = r = 1.0$, no particular architecture is favored for the sheet. The average number of residues in a sheet and the mean square end-to-end distance depend only on the value of $t$. The effect of $t$ on the mean square end-to-end distance is shown in Figure 18 for three sets of statistical weights. The line for $\delta = r = 1.0$ starts at $\sqrt{\langle r^2 \rangle} = 182$ with $t = 0.25$. As $t$ increases the value of $\sqrt{\langle r^2 \rangle}$ decreases to a minimum at $t = 1.25$, then increases steadily. The decrease can be explained by considering the formation of a large number of very small sheets which have the effect of compacting the random coil into a smaller area. Reference to Table 8 leads to a picture of the architecture of the sheet as $t$ increases. The fifth and sixth columns list the average numbers of residues per strand and the average number of strands per sheet respectively. Column three lists the average number of residues per sheet. As the average number of residues in the sheet increases, it is apparent that the sheets go from having one or two strands of two or three residues each to
Table 8.—Sheet Coil Transition. \( \delta = 1.0, \tau = 1.0, I = 12, \)
\( n = 300 \)

<table>
<thead>
<tr>
<th>( t )</th>
<th>( f(\text{sheet}) )</th>
<th>( \text{ave } \sqrt{\langle r^2 \rangle} )</th>
<th>( \text{ave } \text{res} \text{ strand} )</th>
<th>( \text{ave } \text{res} \text{ sheet} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.1880</td>
<td>3.01</td>
<td>184</td>
<td>2.30</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5021</td>
<td>4.76</td>
<td>140</td>
<td>2.56</td>
</tr>
<tr>
<td>0.75</td>
<td>0.6531</td>
<td>6.36</td>
<td>124</td>
<td>2.74</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7334</td>
<td>7.86</td>
<td>118</td>
<td>2.87</td>
</tr>
<tr>
<td>1.25</td>
<td>0.7829</td>
<td>9.30</td>
<td>117</td>
<td>2.96</td>
</tr>
<tr>
<td>1.50</td>
<td>0.8166</td>
<td>10.70</td>
<td>117</td>
<td>3.03</td>
</tr>
<tr>
<td>1.75</td>
<td>0.8411</td>
<td>12.08</td>
<td>119</td>
<td>3.09</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8596</td>
<td>13.44</td>
<td>121</td>
<td>3.13</td>
</tr>
<tr>
<td>3.00</td>
<td>0.9041</td>
<td>18.71</td>
<td>126</td>
<td>3.25</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9270</td>
<td>23.75</td>
<td>137</td>
<td>3.31</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9410</td>
<td>28.60</td>
<td>148</td>
<td>3.36</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9505</td>
<td>33.28</td>
<td>157</td>
<td>3.39</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9574</td>
<td>37.80</td>
<td>166</td>
<td>3.41</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9625</td>
<td>42.16</td>
<td>174</td>
<td>3.43</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9666</td>
<td>46.38</td>
<td>181</td>
<td>3.55</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9699</td>
<td>50.46</td>
<td>187</td>
<td>3.45</td>
</tr>
<tr>
<td>20.00</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.00</td>
<td>262</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9.—Sheet Coil Transition. $\delta = 1.0$, $r = 0.1$, $I = 6$, $n = 300$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f$(sheet)</th>
<th>ave res sheet</th>
<th>$\langle r^2 \rangle$</th>
<th>ave res strand</th>
<th>ave strand sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.0012</td>
<td>2.47</td>
<td>200</td>
<td>2.02</td>
<td>1.22</td>
</tr>
<tr>
<td>0.25</td>
<td>0.0024</td>
<td>2.84</td>
<td>200</td>
<td>2.03</td>
<td>1.40</td>
</tr>
<tr>
<td>0.30</td>
<td>0.0054</td>
<td>3.48</td>
<td>199</td>
<td>2.04</td>
<td>1.71</td>
</tr>
<tr>
<td>0.40</td>
<td>0.0424</td>
<td>7.71</td>
<td>189</td>
<td>2.07</td>
<td>3.72</td>
</tr>
<tr>
<td>0.45</td>
<td>0.1675</td>
<td>16.88</td>
<td>163</td>
<td>2.10</td>
<td>8.03</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3314</td>
<td>33.58</td>
<td>153</td>
<td>2.13</td>
<td>15.77</td>
</tr>
<tr>
<td>0.60</td>
<td>0.4679</td>
<td>61.19</td>
<td>188</td>
<td>2.17</td>
<td>28.18</td>
</tr>
<tr>
<td>0.70</td>
<td>0.5443</td>
<td>81.05</td>
<td>219</td>
<td>2.21</td>
<td>36.69</td>
</tr>
<tr>
<td>0.75</td>
<td>0.5748</td>
<td>89.44</td>
<td>232</td>
<td>2.23</td>
<td>40.15</td>
</tr>
<tr>
<td>0.80</td>
<td>0.6017</td>
<td>97.14</td>
<td>242</td>
<td>2.25</td>
<td>43.27</td>
</tr>
<tr>
<td>1.00</td>
<td>0.6850</td>
<td>123.09</td>
<td>276</td>
<td>2.31</td>
<td>53.28</td>
</tr>
<tr>
<td>1.25</td>
<td>0.7535</td>
<td>148.18</td>
<td>301</td>
<td>2.38</td>
<td>62.29</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7993</td>
<td>167.74</td>
<td>316</td>
<td>2.43</td>
<td>68.89</td>
</tr>
<tr>
<td>1.75</td>
<td>0.8316</td>
<td>183.30</td>
<td>326</td>
<td>2.48</td>
<td>73.88</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8554</td>
<td>195.89</td>
<td>333</td>
<td>2.52</td>
<td>77.75</td>
</tr>
<tr>
<td>3.00</td>
<td>0.9087</td>
<td>228.20</td>
<td>345</td>
<td>2.62</td>
<td>87.03</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9337</td>
<td>245.67</td>
<td>349</td>
<td>2.68</td>
<td>91.62</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9481</td>
<td>256.46</td>
<td>351</td>
<td>2.72</td>
<td>94.30</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9574</td>
<td>263.72</td>
<td>352</td>
<td>2.75</td>
<td>96.03</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9638</td>
<td>268.94</td>
<td>353</td>
<td>2.77</td>
<td>97.24</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9686</td>
<td>272.85</td>
<td>353</td>
<td>2.78</td>
<td>98.12</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9723</td>
<td>275.90</td>
<td>353</td>
<td>2.79</td>
<td>98.80</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9752</td>
<td>278.33</td>
<td>353</td>
<td>2.80</td>
<td>99.33</td>
</tr>
</tbody>
</table>
Table 10.—Sheet Coil Transition. $\delta = 0.1$, $r = 1.0$, $I = 15$, $n = 300$

<table>
<thead>
<tr>
<th>t</th>
<th>$f$(sheet)</th>
<th>ave</th>
<th>ave</th>
<th>ave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>res</td>
<td>/&lt;r^2&gt;</td>
<td>strand</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>0.25</td>
<td>0.1371</td>
<td>2.36</td>
<td>205</td>
<td>2.30</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3711</td>
<td>3.27</td>
<td>209</td>
<td>2.64</td>
</tr>
<tr>
<td>0.75</td>
<td>0.5208</td>
<td>3.27</td>
<td>210</td>
<td>2.96</td>
</tr>
<tr>
<td>1.00</td>
<td>0.6104</td>
<td>3.69</td>
<td>209</td>
<td>3.24</td>
</tr>
<tr>
<td>1.25</td>
<td>0.6685</td>
<td>4.09</td>
<td>209</td>
<td>3.50</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7093</td>
<td>4.46</td>
<td>208</td>
<td>3.72</td>
</tr>
<tr>
<td>1.75</td>
<td>0.7395</td>
<td>4.82</td>
<td>206</td>
<td>3.93</td>
</tr>
<tr>
<td>2.00</td>
<td>0.7629</td>
<td>5.16</td>
<td>205</td>
<td>4.11</td>
</tr>
<tr>
<td>3.00</td>
<td>0.8208</td>
<td>6.42</td>
<td>191</td>
<td>4.70</td>
</tr>
<tr>
<td>4.00</td>
<td>0.8524</td>
<td>7.54</td>
<td>184</td>
<td>5.12</td>
</tr>
<tr>
<td>5.00</td>
<td>0.8729</td>
<td>8.57</td>
<td>179</td>
<td>5.44</td>
</tr>
<tr>
<td>6.00</td>
<td>0.8875</td>
<td>9.55</td>
<td>174</td>
<td>5.69</td>
</tr>
<tr>
<td>7.00</td>
<td>0.8985</td>
<td>10.48</td>
<td>169</td>
<td>5.89</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9073</td>
<td>11.37</td>
<td>165</td>
<td>6.05</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9145</td>
<td>12.24</td>
<td>162</td>
<td>6.19</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9206</td>
<td>13.10</td>
<td>159</td>
<td>6.30</td>
</tr>
</tbody>
</table>
Figure 18. Effect of $t$ on the square root of the mean square end to end distance when $n = 300$: $\delta = \tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (+); and $\delta = 0.1, \tau = 1.0$ (◇). (a) $f_1 = 0$; (b) $f_1 = 0$. 
sheets of fifteen strands of three or four residues each. When the average sheet is one of approximately four strands of three residues each, the overall size of the chain begins to increase. When the average sheet is one of fifteen strands of three or four residues each, the formation of sheets with less than \( n \) residues no longer results in a compacting of the chain. Notice, too, that the average number of residues per sheet is much less than 300 while the fraction of residues in a sheet is very high. Clearly, more than one sheet is forming.

When the initiation of new strands is penalized by setting \( r = 0.1 \), there is a drastic change in the effect of \( t \) on \( \sqrt{\langle r' \rangle} \) (Figure 18). The initial decrease in the end-to-end distance can be explained with the same discussion used for \( \delta = r = 1.0 \), but the following increase occurs at a lower value of \( t \) and is sharper. This can be accounted for by considering the effect of a penalty on the initiation of new strands. If one strand of minimal length has formed, creating any additional new strands is not favored, nor is it necessary. As \( t \) increases additional sheet is formed by adding to the existing sheet. The result is one sheet very many strands all of which have very few residues. This can be confirmed by looking at columns five and six in Table 9. Notice, too, that when \( t = 10.0 \), the average sheet occupies a very large portion of the 300 residues available. Such a sheet resembles a rod and is expected to have a very large end-to-end
distance.

The third line plotted in Figure 18 represents the results when $\delta$ rather than $r$ is penalized. In this case the noncooperative formation of very large strands with very few bends connecting them is expected. The path of the curve for this set of statistical weights is unorthodox. The decrease in $\langle r' \rangle$ at high values of $t$ is unexpected. This is because the value of $I$ used is falsely low. Support for this can be gained by referring to Table 10. When $I = 15$, the highest average number of residues in a sheet is $13.10$. Although the average sheets formed two strands of six residues each, it is possible that by limiting the value of $I$ to 15, the average is being skewed toward a falsely low value. In theory this calculation should be repeated with larger and larger values of $I$. In fact, when $\delta$ is penalized this much the most appropriate value of $I$ is $I \approx 262$. In this case, the time required to do the calculation is much greater than interest in the actual results. Another difficulty with this approach is that the C compiler used refused to permit formation of the large rows required for $I > 55$.

Part b of Figure 18 shows results when interstrand loops are not allowed, i.e., $f_1 = 0$. The effect of this transition is to shift the minimum end-to-end distance to a higher value of $t$. This can be justified as follows. First consider small values of $t$, that is, $t$ less than the value that results in the minimum in the mean
square end-to-end distance. Assume that, with $\delta$ and $\tau$ constant, a given value of $t$ results in an average sheet of a specific number of residues. For example, for a chain of eighteen residues, a chosen value of $t$ results in an average sheet with four strands of three residues each. If interstrand loops are allowed, it is possible to picture a conformation in which the six residues not in the sheet are in loops between strands. Then the end-to-end distance for the chain is the end-to-end distance for a $4 \times 3$ sheet. On the other hand, if interstrand loops are not allowed, the remaining six residues must be at one end or the other. In this case it is possible to picture a conformation for which the end-to-end distance is greater than it is for the $4 \times 3$ sheet. As $t$ increases, the statistical weight of interstrand loops becomes less than the statistical weight for starting a new strand ($f_1 < \delta t$), and so the line for $f_1 = 0$ joins the line for $f_1 = 0.78$.

For all three sets of statistical weights, regardless of the value of $f_1$, the value of $\langle r^2 \rangle$ approaches 200 as $t$ goes to zero. This number is the mean square end-to-end distance for a randomly coiled homopolypeptide of 300 residues. A similar behavior is seen for the $\alpha$-helix coil transition. The mean square end-to-end distance for the chain is the same when the chain is completely random and when the chain is completely helical regardless of differences during transition due to the values of the statistical weights. In the random coil to
\( \beta \)-sheet transition the mean square end-to-end distance is the same for all random coils, but not for all chains in which every residue is in a \( \beta \)-sheet. This is because of the different potential architectures of a sheet which are dependent on the values of the end effects, \( \delta \) and \( \tau \).

It is interesting to consider the value of \( \sqrt{\langle r' \rangle} \) when plotted against the average number of residues in a sheet for a given set of statistical weights. These are plotted for this group in Figure 19. When both \( \delta \) and \( \tau \) are unity, the plot supports the idea expressed above that an initial increase in the average number of residues per sheet results in closer packing of the chain due to the more ordered structure. However, at some number of residues per sheet--approximately ten in this case--the addition of residues results in an increase in the end-to-end distance due to the increased size of the sheet being formed and also, possibly, due to an increase in the average number of sheets being formed.

When \( \delta \) is penalized (\( \delta = 0.1 \)), and strand length is limited to 15 residues, one can imagine a large number of sheets of two strands of the same length (and therefore of a very small end-to-end distance) packing closely to decrease the end-to-end distance. Penalizing \( \tau \) results in the formation of one sheet with many strands occupying over 270 residues. This sheet has a very large end-to-end distance because it is a rod.

In an analysis of the helix coil transition, the
Figure 19. A plot of the square root of the mean square end-to-end distance as a function of the average number of residues per sheet for n = 300: $\delta = 1.0, \tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (★); and $\delta = 0.1, \tau = 1.0$ (◇).
fraction of residues in a helix is often plotted as a function of the value of $s$, the statistical weight parameter for a helix. In these plots, it is found that the steepness of the plot increases as smaller values of the initiation parameter, $\sigma$, are used. The steepness of this plot at the midpoint of transition ($f_h = 0.5$) represents the degree of cooperativity in the transition; the steeper the plot the more cooperative the transition. A highly cooperative transition means that once a helix has begun to form, it favors its own growth.

When a similar plot is made to gauge the sheet coil transition, Figure 20, it is seen that all transitions are quite cooperative. Penalizing $r$, the sheet initiation factor, slightly increases the steepness of the plot. Surprisingly, altering the value of $r$ shifts the entire transition curve, the two lines not meeting until over 80% of the residues are in sheet conformation. If this transition were analagous to the helix coil transition, one might expect these lines to cross at the midpoint of transition. The plot for $\delta = 0.1$, $r = 1.0$ indicates that penalizing $\delta$ shifts the entire transition to a higher value of $t$. This line is not plotted with a high level of confidence because calculations were made with $I = 15$, and, as mentioned earlier, the appropriate value of $I$ is 262.

One additional analysis of these data is interesting and informative because of its relationship to experimentally obtainable results. Using circular
Figure 20. The effect of $t$ on the fraction of residues in a sheet with $n = 300$ and $\delta = \tau = 1.0$ ($\square$); $\delta = 1.0$, $\tau = 0.1$ ($+$); $\delta = 0.1$, $\tau = 1.0$ ($\diamond$).
dichroism, one can determine the fraction sheet in a polypeptide chain. Small angle light scattering yields information about the end-to-end distance of large molecules. In principle, it is possible to measure both of these quantities in identical systems and observe how the end-to-end distance varies with the fraction of residues in a $\beta$-sheet. The major practical difficulty is finding a system where aggregation will not be a problem.

A theoretical prediction of these results for this group of statistical weights ($\delta, \tau = 1.0, 0.1$) is shown in Figure 21. When there is no penalty for edge effects ($\delta = \tau = 1.0$), the square root of the mean square end-to-end distance decreases to a minimum at approximately 80% $\beta$-sheet. The addition of residues to existing sheets results in an increase in the end-to-end distance. When $\tau$ is penalized, the initial decrease is followed by a steady and continuous increase in the end-to-end distance as more and more of the residues go into sheets. This, in agreement with previous evidence, supports the formation of a cross $\beta$-sheet; a sheet with many strands of few residues. The third line in Figure 21 is the result of penalizing $\delta$. Since a falsely low value of $I$ is used in these calculations, this line is not totally correct. It would appear from this plot that when more than 75% of the residues are in a $\beta$-sheet, the low value of $I$ is forcing bends to form. This results in a delayed decrease in the end-to-end distance.
Figure 21. The relationship between $\langle r^2 \rangle$ and the fraction of residues in a sheet for $n = 300$: $\delta = 1.0, \tau = 1.0$ (□); $\delta = 1.0, \tau = 0.1$ (♦); $\delta = 0.1, \tau = 1.0$ (◊).
Group II: \( \delta, \tau = 0.3, 0.6, 0.9 \)

For the five sets of statistical weights in this group, the initial effect of an increase in \( t \) is to decrease \( \langle r^2 \rangle \). Reference to Tables 11 - 15 reveals that at the smallest values of \( t \) the average numbers of residues per sheet is two or three. This suggests that the formation of small sheets results in a decrease in the overall end-to-end distance. When \( \delta \) is increased, the minimum is shifted to a lower value of \( t \). By looking at Tables 11, 14, and 15 it can be seen that as \( \delta \) goes from 0.3 to 0.9, the average number of strands per sheet increases more rapidly with \( t \). Moreover, the average number of residues per strand increases less with increasing \( t \) as the penalty for bends is decreased from 0.3 to 0.9. The increase in the favorability of bend formation results in a cross \( \beta \)-sheet, a rod-like formation, at lower values of \( t \) than when \( \delta \) is more heavily penalized.

When \( \tau \) is increased from 0.3 to 0.9 while holding \( \delta \) constant, the value of \( \langle r^2 \rangle \) for a chain 300 residues long is nearly the same when the average number of residues per sheet is the same. Compare, for example, the entry for \( t = 1.75 \) in Table 11 to the entry for \( t = 6 \) in Table 12. The shape of the sheets being formed is not changed very much by the change in the penalty for formation of a first strand. However, there is a large difference in the value of \( t \) required for the formation of sheets of similar size and shape, and the sheet content of the chains differ
Table 11.—Sheet Coil Transition. $\delta = 0.3$, $r = 0.3$, $I = 20$, $n = 300$.

<table>
<thead>
<tr>
<th>t</th>
<th>f(sheet)</th>
<th>ave res sheet</th>
<th>$&lt;r^2&gt;$</th>
<th>ave res strand</th>
<th>ave strand sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.0146</td>
<td>2.28</td>
<td>199</td>
<td>2.08</td>
<td>1.10</td>
</tr>
<tr>
<td>0.50</td>
<td>0.1020</td>
<td>3.36</td>
<td>186</td>
<td>2.21</td>
<td>1.52</td>
</tr>
<tr>
<td>0.75</td>
<td>0.3363</td>
<td>6.33</td>
<td>140</td>
<td>2.45</td>
<td>2.58</td>
</tr>
<tr>
<td>1.00</td>
<td>0.5500</td>
<td>11.57</td>
<td>113</td>
<td>2.78</td>
<td>4.17</td>
</tr>
<tr>
<td>1.25</td>
<td>0.6842</td>
<td>19.10</td>
<td>114</td>
<td>3.13</td>
<td>6.10</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7684</td>
<td>29.16</td>
<td>125</td>
<td>3.49</td>
<td>8.35</td>
</tr>
<tr>
<td>1.75</td>
<td>0.8253</td>
<td>41.67</td>
<td>137</td>
<td>3.84</td>
<td>10.85</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8630</td>
<td>56.14</td>
<td>146</td>
<td>4.17</td>
<td>13.46</td>
</tr>
<tr>
<td>3.00</td>
<td>0.9331</td>
<td>118.72</td>
<td>158</td>
<td>5.22</td>
<td>22.74</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9579</td>
<td>187.31</td>
<td>156</td>
<td>5.90</td>
<td>28.36</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9697</td>
<td>199.01</td>
<td>153</td>
<td>6.34</td>
<td>31.39</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9764</td>
<td>219.73</td>
<td>151</td>
<td>6.64</td>
<td>33.08</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9807</td>
<td>233.87</td>
<td>149</td>
<td>6.86</td>
<td>34.10</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9837</td>
<td>243.99</td>
<td>147</td>
<td>7.02</td>
<td>34.75</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9859</td>
<td>251.52</td>
<td>146</td>
<td>7.15</td>
<td>35.19</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9876</td>
<td>257.32</td>
<td>145</td>
<td>7.25</td>
<td>35.51</td>
</tr>
</tbody>
</table>
Table 12.—Sheet Coil Transition: $\delta = 0.3$, $r = 0.6$, $I = 25$, $n = 300$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f$(sheet)</th>
<th>$\bar{r}^2$</th>
<th>$\sqrt{\langle r^2 \rangle}$</th>
<th>strand</th>
<th>sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.0588</td>
<td>2.38</td>
<td>199</td>
<td>2.17</td>
<td>1.09</td>
</tr>
<tr>
<td>0.50</td>
<td>0.2561</td>
<td>3.26</td>
<td>177</td>
<td>2.39</td>
<td>1.36</td>
</tr>
<tr>
<td>0.75</td>
<td>0.4560</td>
<td>4.44</td>
<td>148</td>
<td>2.62</td>
<td>1.69</td>
</tr>
<tr>
<td>1.00</td>
<td>0.5899</td>
<td>5.75</td>
<td>130</td>
<td>2.84</td>
<td>2.09</td>
</tr>
<tr>
<td>1.25</td>
<td>0.6780</td>
<td>7.17</td>
<td>120</td>
<td>3.04</td>
<td>2.36</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7389</td>
<td>8.72</td>
<td>114</td>
<td>3.23</td>
<td>2.70</td>
</tr>
<tr>
<td>1.75</td>
<td>0.7829</td>
<td>10.38</td>
<td>111</td>
<td>3.39</td>
<td>3.06</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8157</td>
<td>12.13</td>
<td>109</td>
<td>3.54</td>
<td>3.43</td>
</tr>
<tr>
<td>3.00</td>
<td>0.8894</td>
<td>19.78</td>
<td>112</td>
<td>3.99</td>
<td>4.95</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9231</td>
<td>27.89</td>
<td>119</td>
<td>4.29</td>
<td>6.50</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9417</td>
<td>35.99</td>
<td>125</td>
<td>4.49</td>
<td>8.01</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9710</td>
<td>43.89</td>
<td>132</td>
<td>4.64</td>
<td>9.46</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9611</td>
<td>51.47</td>
<td>137</td>
<td>4.75</td>
<td>10.84</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9668</td>
<td>58.72</td>
<td>141</td>
<td>4.84</td>
<td>12.14</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9710</td>
<td>65.62</td>
<td>145</td>
<td>4.90</td>
<td>13.39</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9743</td>
<td>72.18</td>
<td>148</td>
<td>4.96</td>
<td>14.56</td>
</tr>
</tbody>
</table>

Table 13.—Sheet Coil Transition: $\delta = 0.3$, $r = 0.9$, $I = 45$, $n = 300$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f$(sheet)</th>
<th>$\bar{r}^2$</th>
<th>$\sqrt{\langle r^2 \rangle}$</th>
<th>strand</th>
<th>sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.1235</td>
<td>2.46</td>
<td>198</td>
<td>2.27</td>
<td>1.09</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3727</td>
<td>3.20</td>
<td>177</td>
<td>2.57</td>
<td>1.25</td>
</tr>
<tr>
<td>0.75</td>
<td>0.5373</td>
<td>3.94</td>
<td>159</td>
<td>2.83</td>
<td>1.39</td>
</tr>
<tr>
<td>1.00</td>
<td>0.6349</td>
<td>4.65</td>
<td>148</td>
<td>3.05</td>
<td>1.53</td>
</tr>
<tr>
<td>1.25</td>
<td>0.6978</td>
<td>5.35</td>
<td>141</td>
<td>3.23</td>
<td>1.66</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7418</td>
<td>6.04</td>
<td>136</td>
<td>3.39</td>
<td>1.78</td>
</tr>
<tr>
<td>1.75</td>
<td>0.7744</td>
<td>6.74</td>
<td>133</td>
<td>3.54</td>
<td>1.90</td>
</tr>
<tr>
<td>2.00</td>
<td>0.7996</td>
<td>7.43</td>
<td>130</td>
<td>3.66</td>
<td>2.03</td>
</tr>
<tr>
<td>3.00</td>
<td>0.8614</td>
<td>10.19</td>
<td>125</td>
<td>4.05</td>
<td>2.51</td>
</tr>
<tr>
<td>4.00</td>
<td>0.8942</td>
<td>12.95</td>
<td>123</td>
<td>4.33</td>
<td>3.00</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9145</td>
<td>15.70</td>
<td>123</td>
<td>4.53</td>
<td>3.47</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9284</td>
<td>18.43</td>
<td>124</td>
<td>4.68</td>
<td>3.94</td>
</tr>
</tbody>
</table>
Table 14.—Sheet Coil Transition: $\delta = 0.6$, $r = 0.3$, $I = 12$, $n = 300$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f$(sheet)</th>
<th>ave res sheet</th>
<th>$/r^1$</th>
<th>ave res strand</th>
<th>ave strand sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.0178</td>
<td>2.53</td>
<td>199</td>
<td>2.08</td>
<td>1.21</td>
</tr>
<tr>
<td>0.50</td>
<td>0.2078</td>
<td>5.80</td>
<td>159</td>
<td>2.25</td>
<td>2.58</td>
</tr>
<tr>
<td>0.75</td>
<td>0.5099</td>
<td>13.87</td>
<td>118</td>
<td>2.47</td>
<td>5.62</td>
</tr>
<tr>
<td>1.00</td>
<td>0.6584</td>
<td>22.84</td>
<td>132</td>
<td>2.65</td>
<td>8.61</td>
</tr>
<tr>
<td>1.25</td>
<td>0.7433</td>
<td>32.08</td>
<td>152</td>
<td>2.80</td>
<td>11.44</td>
</tr>
<tr>
<td>1.50</td>
<td>0.7972</td>
<td>41.42</td>
<td>169</td>
<td>2.93</td>
<td>14.15</td>
</tr>
<tr>
<td>1.75</td>
<td>0.8339</td>
<td>50.70</td>
<td>183</td>
<td>3.03</td>
<td>16.74</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8601</td>
<td>59.76</td>
<td>195</td>
<td>3.11</td>
<td>19.20</td>
</tr>
<tr>
<td>3.00</td>
<td>0.9160</td>
<td>92.45</td>
<td>222</td>
<td>3.34</td>
<td>27.66</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9406</td>
<td>118.87</td>
<td>235</td>
<td>3.48</td>
<td>34.19</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9543</td>
<td>139.94</td>
<td>242</td>
<td>3.56</td>
<td>39.27</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9629</td>
<td>156.91</td>
<td>246</td>
<td>3.62</td>
<td>43.30</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9688</td>
<td>170.77</td>
<td>249</td>
<td>3.67</td>
<td>46.54</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9731</td>
<td>182.26</td>
<td>251</td>
<td>3.70</td>
<td>49.22</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9764</td>
<td>191.93</td>
<td>252</td>
<td>3.73</td>
<td>51.45</td>
</tr>
<tr>
<td>10.00</td>
<td>0.9790</td>
<td>200.16</td>
<td>253</td>
<td>3.75</td>
<td>53.34</td>
</tr>
</tbody>
</table>

Table 15.—Sheet Coil Transition $\delta = 0.9$, $r = 0.3$, $I = 10$, $n = 300$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f$(sheet)</th>
<th>ave res sheet</th>
<th>$/r^1$</th>
<th>ave res strand</th>
<th>ave strand sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.0220</td>
<td>2.83</td>
<td>197</td>
<td>2.09</td>
<td>1.36</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3358</td>
<td>10.01</td>
<td>128</td>
<td>2.26</td>
<td>4.43</td>
</tr>
<tr>
<td>0.75</td>
<td>0.5897</td>
<td>22.37</td>
<td>133</td>
<td>2.42</td>
<td>9.23</td>
</tr>
<tr>
<td>1.00</td>
<td>0.7057</td>
<td>33.68</td>
<td>163</td>
<td>2.54</td>
<td>13.25</td>
</tr>
<tr>
<td>1.25</td>
<td>0.7734</td>
<td>44.33</td>
<td>188</td>
<td>2.63</td>
<td>16.84</td>
</tr>
<tr>
<td>1.50</td>
<td>0.8171</td>
<td>54.41</td>
<td>207</td>
<td>2.70</td>
<td>20.14</td>
</tr>
<tr>
<td>1.75</td>
<td>0.8474</td>
<td>63.92</td>
<td>222</td>
<td>2.76</td>
<td>23.19</td>
</tr>
<tr>
<td>2.00</td>
<td>0.8694</td>
<td>72.88</td>
<td>234</td>
<td>2.80</td>
<td>26.02</td>
</tr>
<tr>
<td>3.00</td>
<td>0.9180</td>
<td>103.51</td>
<td>264</td>
<td>2.92</td>
<td>35.50</td>
</tr>
<tr>
<td>4.00</td>
<td>0.9405</td>
<td>127.35</td>
<td>279</td>
<td>2.98</td>
<td>42.73</td>
</tr>
<tr>
<td>5.00</td>
<td>0.9534</td>
<td>146.21</td>
<td>288</td>
<td>3.02</td>
<td>48.39</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9618</td>
<td>161.45</td>
<td>293</td>
<td>3.05</td>
<td>52.94</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9676</td>
<td>173.97</td>
<td>298</td>
<td>3.07</td>
<td>56.66</td>
</tr>
<tr>
<td>8.00</td>
<td>0.9719</td>
<td>184.45</td>
<td>300</td>
<td>3.09</td>
<td>59.76</td>
</tr>
<tr>
<td>9.00</td>
<td>0.9752</td>
<td>199.33</td>
<td>303</td>
<td>3.10</td>
<td>62.39</td>
</tr>
</tbody>
</table>
(f(sheet) = 0.83 and 0.97 for the table entries cited).

When the value of $r$ is increased from 0.3 to 0.9, the formation of many small sheets becomes more favored. The third column of Tables 12 and 13 reveals that the average number of residues per sheet is consistently less for any given value of $t$ than the corresponding number for the other sets of statistical weights.

A different situation arises when $\delta$ is increased from 0.3 to 0.9 while $r$ is held constant. The penalty on $r$ discourages the formation of new sheets. Decreasing the penalty ascribed to bends results in a more rapid and exaggerated increase in $\langle r^2 \rangle$. The formation of sheets with more bends leads to sheets which begin to resemble rigid strands of cross $\beta$-sheet.

Considering the relative values of the average number of residue per sheet and $\langle r^2 \rangle$, the most obvious and palatable observation is that the greater the penalty for edges, the more likely it is that one large sheet will form. This observation arises from the fact that when $\delta = r = 0.3$ the average number of residues per sheet at the highest value of $t$ calculated (Table 11) is larger than for the other sets (Tables 12 - 15). This fact suggests that when the cost of initiation is high, initiation will occur less often.

The square root of the mean square end-to-end distance decreases slightly at high values of $t$ when $\delta$ and $r$ are equally penalized. Analysis of the values in the
fifth and sixth columns of Table 11 leads one to conclude that the average sheet is becoming more square and less of a cross \( \beta \)-sheet. The ratio of the values in column six to those in volume five decreases for \( t > 6 \).

Figure 22 and 23 are plots of \( \langle r^2 \rangle \) against \( t \) for the sets of \( \delta \) and \( r \) in this group. The effect of increasing the value of \( \delta \) is to shift the value of \( t \) at which \( \langle r^2 \rangle \) reaches a minimum to a slightly lower number and to allow a larger \( \langle r^2 \rangle \) at high values of \( t \). Increasing the value of \( r \)--that is, decreasing the penalty ascribed to the first strand--has the effect of shifting the minimum \( \langle r^2 \rangle \) to a higher value of \( t \). The increase in \( \langle r^2 \rangle \) with \( t \) after the minimum is more gradual than when \( r \) is smaller.

Also shown in Figures 22 and 23 are the results when \( f_1 = 0 \). The minimum value of \( \langle r^2 \rangle \) occurs at a higher value of \( t \), as seen when \( \delta, r = 1.0, 0.1 \). Since the results show the same trend as is seen in the first group of statistical weights, the reader is referred to the earlier discussion.

Figures 24 and 25 are plots of the fraction of residues in a sheet against \( t \). In all cases a plateau is reached at high values of \( t \), in some cases higher than is shown on the figure. The plateau indicates that the transition has gone as far as it will go for a given set of statistical weights. When \( \delta \) is increased from 0.3 to 0.9 while keeping \( r \) constant, the transition is shifted to
Figure 22. The effect of $t$ on the square root of the mean square end to end distance when $n = 300$ for $\delta = 0.3$, $\tau = 0.3$ (□); $\delta = 0.6$, $\tau = 0.3$ (+); $\delta = 0.9$, $\tau = 0.3$ (Φ). (a) $f_1 = 0.78$, (b) $f_1 = 0$. 
Figure 23. The effect of $t$ on the square root of the mean square end to end distance for $\delta = \tau = 0.3$ (□); $\delta = 0.3$, $\tau = 0.6$ (+); $\delta = 0.3$, $\tau = 0.9$ (○). (a) $f_1 = 0.78$; (b) $f_1 = 0$. 
Figure 24. The effect of $t$ on the fraction of residues in a sheet for $\delta = \tau = 0.3$ (□); $\delta = 0.6$, $\tau = 0.3$ (＋); $\delta = 0.9$, $\tau = 0.3$ (◇).
Figure 25. The effect of $t$ on the fraction of residues in a sheet for $\delta = \tau = 0.3$ (■); $\delta = 0.3, \tau = 0.6$ (♦); $\delta = 0.3, \tau = 0.9$ (◇).
lower values of \( t \).

When \( t \) is increased from 0.3 to 0.9, the cooperativity of the transition is lessened. The transition is less sharp. This behavior agrees with that observed by Zimm and Bragg for the relationship between the value of \( \sigma \) and the helix coil transition. Since \( \sigma \) and \( r \) have already been compared as initiation parameters, it is reassuring that they have the same effect on the transition. Any decrease in the edge penalties will result in a more cooperative transition. There is, however, an interesting difference. In Figure 25 there appears to be a point, or at least a region for which one value of \( t \) results in a given fraction of sheet, regardless of the value of \( r \). For the \( \alpha \)-helix coil transition, this point occurs at the midpoint of transition, i.e., \( f(\text{sheet}) = 0.5 \). The fact that there is a common point encourages the analogy with the helix coil transition. The fact that it is not at the midpoint of transition, causes one to suspect that the role of \( \delta \) is important. It would be interesting to determine if a value of \( \delta \) can be found that results in shifting this common point in the transition closer to the midpoint.

Plotting the effect of changing fraction of residues in \( \beta \)-sheet conformation on the square root of the mean square end-to-end distance gives a connection to results that can be confirmed experimentally. These plots are shown in Figures 26 and 27. In Figure 26 the three
lines represent a decrease in the penalty ascribed to bends as \( \delta \) is increased from 0.3 to 0.9. Reducing the penalty of bends results in the minimum end-to-end distance occurring when a lower fraction of the residues is in sheets. In other words, the more easily bends can form, the more quickly sheet formation results in an increase in the end-to-end distance due to the formation of rod-like cross-\( \beta \)-sheets.

As \( r \) is increased (and the penalty for new strand formation is lessened), the minimum end-to-end distance occurs when the fraction of residues in sheets is higher. Decreasing the penalty for new strand formation encourages the formation of sheets of one very long strand to occur. As long as the end-to-end distance is decreasing, it is probably the case that several unconnected sheets of a single strand each are forming. In fact, the average sheet has more than one strand (Tables 12 and 13). It is true, however, that as \( t \) goes from 1.75 to 3.0 in for \( \delta = 0.3, \ r = 0.9 \) (Table 12), the increase in the average number of residues per sheet jumps. This corresponds to the minimum in the square root of the mean square end-to-end distance and supports the suggestion that as the average chain goes from one of many small sheets to one of fewer larger sheets, the end-to-end distance must increase. For small values of \( r \) this occurs at higher values of \( t \) and when a much larger portion of the chain is in sheets.

These final two plots suggest a large series of
Figure 26. The relationship between $\langle r^2 \rangle$ and fraction of residues in a sheet for $n = 300$: $\delta = 0.3, \tau = 0.3$ (□); $\delta = 0.6, \tau = 0.3$ (⧫); $\delta = 0.9, \tau = 0.3$ (◊).
Figure 27. The relationship between $\langle r^2 \rangle$ and fraction of residues in a sheet for $n = 300$: $\delta = 0.3$, $\tau = 0.3$ (□); $\delta = 0.3$, $\tau = 0.6$ (+); $\delta = 0.3$, $\tau = 0.9$ (◊).
experiments. There are currently three homopolypeptides for which the sheet coil transition has been measured via circular dichroism. These are poly (L-lysine) (Greenfield and Fasman 1969), poly (L-tyrosine (Auer and Patton 1976), and a modified homopolypeptide, poly (S-carboxymethyl cysteine) (Maeda, et al. 1982). Comparison of results of cd and light scattering experiments with theoretical plots of the sort seen in Figures 21, 26, and 27 may be a means of determining values of $\delta$ and $r$ for these amino acids.

It must be emphasized, however, that avoiding the formation of aggregates is not a trivial problem. The formation of $\beta$-sheets is often stabilized by the stacking of these sheets. This type of aggregation is not accounted for in this theory. To avoid aggregation, very dilute solutions must be used. The theory is best approximated for long chains, but long chains in the sheet-coil transition have a high probability of intramolecular stacking. This very major technical problem is the primary reason the literature is full of studies on the helix-coil transition, but has very few reports on the sheet-coil transition.
PART II:

COMPLEX FORMATION OF POLYPEPTIDES

AND PROANTHOCYANIDINS
CHAPTER SIX
BACKGROUND AND LITERATURE REVIEW

Catechins and Procyanidins

Tannins are a group of molecules originally defined according to what they do rather than by their structure. The name arises from the ability of members of this class of molecules to tan animal hide. It has been suggested that the name tannin be reserved for flavanoid oligomers with molecular weights between 500 and 3000 since these boundaries neatly define the limits of size necessary for the ability to cross-link collagen (van Sumere 1975). The ability to tan leather requires a minimum degree of polymerization of three, i.e., the trimer (Roux 1972)—and is no longer apparent for very large oligomers. A general description designates an essential property of tannins as the ability to combine with proteins and other polymers.

Structure

Tannins are oligomers of flavanoids, illustrated in Figure 28. This figure also introduces the numbering convention used for nonhydrogen atoms of the ring systems. The monomer subunit of the flavanoids discussed herein are flavan-3-ols. In terms of Figure 28, this means that there is a hydroxyl off C-3 either above or below the plane of the paper. The monomers are known collectively as
Figure 28. The skeletal structure of flavonoids including the numbering system. Flavonoids differ by the constituents off the A ring and the heterocyclic ring. The presence of double bonds in the heterocyclic ring is dependent on its constituents. An alternative numbering system defines the C-11 to C-18 carbons in ring B as C-1' to C-6'. The numbering system illustrated here is used throughout the text.
catechins. All the flavan-3-ol monomers have two asymmetric carbon atoms, C-2 and C-3. This gives rise to the four optical isomers shown in Figure 29. There is isomerization between (+)-catechin and (-)-epicatechin and also between (-)-catechin and (+)-epicatechin induced by either brief treatment with a strongly basic solution followed by rapid quenching or by prolonged heating in neutral solution (Harbourne 1975). When purified from plant material such as wattle bark it has been found that (+)-catechin and (-)-epicatechin are more common than their isomers.

There exist many reports of studies of the structure of the monomers. An x-ray analysis of crystalline (-)-epicatechin leads to the conclusion that the aromatic ring bonded to C-2 is in an equatorial position while the hydroxyl group bonded to C-3 is in an axial position (Fronczek, et al. 1984). The conformation of the heterocyclic ring is a half-chair with C-2 above and C-3 below the plane of the adjacent aromatic ring (Spek 1984). Free (+)-catechin has not been crystallized in a form useful for x-ray study, however 8-bromotetra-O-methyl-(+)-catechin has been crystallized. Its x-ray analysis leads to the proposal of a slightly different structure, but this difference may be due to crystal packing (Engel, et al. 1978). The heterocyclic oxygen and C-4 are above the plane of the adjacent benzene ring. The heterocyclic ring is more of a half-chair in (-)-epicatechin and is
Figure 29. The structure of the flavan-3-ol monomers.
between a C-2 sofa and a C-2, C-3 half-chair in 8-bromotetra-0-methyl-(+)-catechin. Proton magnetic resonance combined with x-ray studies on 6-bromo-3, 5, 7, 13, 14-penta-0-methyl-(+)-catechin resulted in elucidation of two structures. In both, C-2 is above and C-3 below the plane of ring A. However, the attached heavy atoms (C-11 at C-2 and O at C-3) are sofa in one molecule and half-chair in the other (Einstein, et al. 1985). Using MM2 calculations, Viswanadhan and Mattice (1986) and Porter, et al. (1986) found the heterocyclic rings to be usually half-chairs with distortions toward C-2 sofa or C-3 sofa.

When the two monomers (+)-catechin and (-)-epicatechin are linked from C-4 of one to C-8 of the other, eight dimers can be formed. The dimers are known as the procyanidins. Acid catalyzed decomposition gives the pigment cyanidin from the upper half and a catechin monomer from the lower half (Harbourne 1975). This is the reason for the name procyanidin.

The formation of this interflavin bond generates a new asymmetric center at C-4. If the linkage is a β-4,8 linkage, where 4,8 refers to the C-4 on the first monomer connected to the C-8 on the second monomer, the configurations are those in which the bond to the lower unit comes out of the plane of the diagram. An α linkage is defined to be one in which the same bond goes into the plane of the diagram. The β dimers—entitled B-1, B-2,
B-3, and B-4--are shown in Figure 30.

The use of proton magnetic resonance enabled Weinges (1968) to assign the configurations of the four diasteriomereric deca-acetylprocyanidins. The various forms were differentiated by determination of the spin-spin coupling constants of the hydrogen atoms at C-2, 2', 3, 3', and 4. Here the primed numbers refer to members of the second subunit, i.e., the unit connected by its C-8. It was assumed that the conformations of the two pyran rings approximated to half-chair forms with the bulky phenyl substituents at C-2 and C-2' in equatorial positions.

There is a two-fold rotation about the bond between monomer units. Nuclear magnetic resonance studies on deca-acetate derivatives of the four dimers in Figure 30 in chloroform at 30°C has demonstrated the presence of two rotational isomers (Foo and Porter 1983). The A rings of each of the two subunits of the dimers form an angle of approximately 90 degrees to each other. The differences in the energies at the minima are sufficiently small so that both minima can be populated to a significant extent in several of the dimers (Viswanadhan and Mattice 1986).

Circular Dichroism

Chiroptical studies on tannin monomers and dimers have been done in methanol and acetone (Barrett, et al. 1979), as well as in acetonitrile and methanol (Gaffield, et al. 1985). In the latter case a parallel $^1$H-NMR study on the same solutions allows an understanding of changes in
Figure 30. The four 4β-8 dimers of (+)-catechin and (-)-epicatechin.
the circular dichroism bands due to conformational differences. The strong negative or positive band at short wavelengths (190-220 nm) is due to an intense $\pi-\pi^*$ absorption band and the interaction of coupled chromophores (ring A and A') located in chiral positions with respect to each other. The magnitude of low-wavelength cd couplets of dimeric procyandin decap-acetates is similar even when the NMR of the dimers shows that they contain differing amounts of conformational isomers. For this reason the transition moment vectors of the A and A' rings may be postulated to be directed along the C-4a to C-7 and C-8' to C-5' axes, respectively. Roux and his coworkers (1982) have formulated a rule based on the cd band at 210-240 nm which defines the absolute stereochemistry at C-4 of flavenoid units. Korver and Wilkins (1971) correlate the sign of the 280 nm region cd bands with flavenol stereochemistry at C-2. Circular dichroism of tannin monomers and dimers in water has not been reported.

**Fluorescence**

Tannin monomers and dimers fluoresce when excited within the absorption band centered at 280 nm. Bergmann, Barkley, and Mattice (1986) report the quantum yield and the lifetime of (+)-catechin and (-)-epicatechin in dioxane and water. The effect of a more hydrophobic solvent is to increase the quantum yield three-fold (0.3 in dioxane and 0.1 in water). The lifetime of the excited state is also much greater in dioxane. In water the lifetime is 0.74 ns
for (+)-catechin and 0.61 ns for (-)-epicatechin. The lifetime for both in dioxane is 2.0 ns.

**Interaction of Catechins and Procyanidins with Polymers**

There are many instances of interaction between plant phenols and proteins in both nature and industry. A short list includes the inhibition or activation of enzymes, providing resistance of fruits to microbial attack, formation of hazes in beer or wine, and—as mentioned previously—tanning of hides.

The chemical structure of catechin and procyanidins allows one to propose three potential modes of interaction. These are covalent bonding, hydrogen bonding, and hydrophobic interactions. Covalent bonding is essentially irreversible. While this is not without interest, the reversible reactions are the focus of the following discussion.

Tannins are believed to bind proteins primarily through the formation of multiple hydrogen bonds. These bonds form between the phenolic hydroxyl groups of tannins and the carbonyl functions of the peptide linkages of proteins (Gustavson 1954; Cannon 1955; Goldstein and Swain 1963; Morawiecki 1963; Loomis and Battaile 1966; van Sumere, et al. 1975; Synge 1975). The emphasis is placed on hydrogen bonding because tannins are bound by synthetic polymers (nylons) which contain -CONH- as the only reactive group. Moreover, the exposed carbonyl oxygens of loosely coiled proteins which are available for hydrogen bonding
are more likely to interact with tannins than their tightly coiled counterparts. For example, oxidized ribonuclease A forms complexes with tannins while native ribonuclease A does not (Hagerman 1978). Further examples of proteins known to be loosely coiled which form complexes with tannins are bovine serum albumin, histone F1, and α-lactalbumin (Hagerman 1978).

There is also evidence that hydrophobic bonding is involved in complex formation. Goldstein and Swain (1965) observed that catechin/protein complexes can be dissociated by detergents. Loomis (1969) has cited several instances in which the same dissociation was achieved with organic solvents such as esters and ketones.

In attempting to prove the importance of hydrophobic interactions to tannin/protein interactions, Hoon, et al., (1980) reports the effect of certain substances on the interaction of tannins with a number of polymers. Anionic and nonionic and to a lesser extent cationic detergents effectively dissociate tannin/protein complexes. However, this may be due to protein-detergent interactions rather than tannin-detergent interactions since the detergents remove the proteins from tannins bound to a column. One interesting observation of this paper is the absorption of procyanidins onto an uncharged polystyrene resin, Amberlite XAD-10. This is contrary to an earlier report by Gray (1978) which reports no absorption of procyanidin. Hoon clearly demonstrated that
procyanidins are capable of hydrophobic interactions.

There is some additional support for the idea that hydrophobic interactions may play an important role in hydrophilic environments. Tanford (1973) has pointed out that in a hydrophilic environment the primary functional group involved in hydrogen-bonding, the carbonyl of the peptide bond, would be solvated. The initial disruption required to form new hydrogen bonds would be thermodynamically unfavorable. Hydrophobic interactions between the solvated hydrophobic regions would, however, be favored.

A mechanism which supports both hydrogen bonding and hydrophobic interaction is presented by McManus (1983). It is possible that proanthocyanidins hydrogen bond to proteins as described above. Since the tannins are capable of hydrophobic interaction, they may interact with each other, clustering together. The proteins which are hydrogen bound to the tannins join in the aggregation and eventually precipitate.

The main purpose of this study is to determine the nature of the interaction of simple catechins and procyanidins with polymers in dilute solution. The study is primarily the interaction of tannin monomers and dimers with polymers that vary in conformation, charge, and the number of methylenes in their side chains. The hope is that by looking at very simple systems of monomers and dimers and homo- or co-polypeptides, the type of
interactions that are important to tannin-protein complex formation can be clarified.

**Polymers**

In an attempt to isolate the requirements of interaction, a number of synthetic polymers are examined. Poly vinyl pyrrolidone has a strong interaction with catechin and procyanidin B-1 (Armstrong 1983; Bergmann 1986). Its side chains have some hydrophobic nature as well as an ability to form hydrogen bonds. It is a very flexible polymer.

Bergmann has observed that in an aqueous solution of poly vinyl pyrrolidone and catechin, the quantum yield of catechin is greater than in a solution of catechin in water alone. Since the solution of poly vinyl pyrrolidone and catechin also shows a decrease in viscosity relative to a solution of poly vinyl pyrrolidone without catechin present, Bergmann concludes that the poly vinyl pyrrolidone wraps around the catechin, thus providing it with a hydrophobic environment. This would increase the quantum yield of catechin and decrease the viscosity of poly vinyl pyrrolidone.

Poly (ethylene glycol) is a flexible hydrogen bond acceptor with no side chains, thereby providing a means of determining if hydrogen-bonds can form with the backbone of the polymer in the absence of side chains. Poly acrylamide, which is capable of hydrogen bonding via side chains, is flexible enough to collapse around the
small molecules. Hydroxypropyl cellulose is a stiff
polymer unable to collapse around the small molecules but
capable of providing either a hydrophobic environment or
accepting hydrogen bonds. The structures of these are
shown in Figure 31.

A number of poly [hydroxyalkyl glutamine] homo-
and co-polymers are examined to determine the effect of
varying degrees of hydrophobicity in side chains on tannin-
protein interaction. These include poly [hydroxybutyl
glutamine] (PHBG), poly [hydroxypropyl glutamine] (PHPG),
poly [bis-hydroxyethyl glutamine] (PdHEG), poly
[hydroxy(ethyl:butyl) glutamine] (PH(E:B)G) where the ratio
of ethyl to butyl is 3.7 to 6.3, and a copolymer of PHBG
and poly (L-Arginine)--(P(HBG:Arg))--where the ratio of HBG
to Arginine is 7 to 1. These are drawn in Figure 32.

The structural transition of PHBG has been well
characterized (Lotan, et al. 1966). This polymer
undergoes a temperature-induced helix to coil transition in
water. The temperature-induced transition of PHPG is
similar to that of PHBG, but occurs at lower temperature
1971). The transition of PH(E:B)G follows the same basic
pattern as for PHBG, but this copolymer is never as
strongly helical as is PHBG. The helix to coil transition
of P(HBG:ARG) is also very similar to that of PHBG. The
temperature-induced helix to coil transition of these four
polymers enables the study of the effect of polypeptide
Figure 31. The structure of synthetic polymers used in this study. (PrOH = CH₂CH₂CH₂OH)
poly [N-hydroxybutyl glutamine] $R_1 = R_3 = H$
$R_2 = R_4 = CH_2CH_2CH_2CH_2OH$

poly [N-hydroxypropyl glutamine] $R_1 = R_3 = H$
$R_2 = R_4 = CH_2CH_2CH_2OH$

copoly [N-hydroxy(butyl:ethyl) glutamine] $R_1 = R_3 = H$
$R_2 = CH_2CH_2CH_2CH_2OH$
$R_4 = CH_2CH_2OH$
$x/y = 2$

copoly [(N-hydroxybutyl glutamine):arginine] $R_5 = CH_2CH_2CH_2CH_2OH$
$x'/y' = 7$

Figure 32. The structures of poly (L-glutamines) used in this study. The sequence of the copolymers is uncontrolled and presumably random.
conformation on procyanidin/protein interaction. The polypeptides also vary in the number of methylene groups in their side chains, and therefore in the degree of hydrophobic interaction accessible to the procyanidins.

Three polymers which have a pH-induced transition are examined. They are poly (L-Glutamic acid) (p-L-Glu), poly (D-Glutamic acid) (p-D-Glu) which have helix coil transitions centered at pH = 5.7, and poly (S-carboxymethyl-L-cysteine) (PCMS) which undergoes a sheet coil transition centered near pH = 5.2 (Adler, et al. 1968; Myer 1969; Krimm and Tiffany 1974; Maeda, et al. 1972). The structures of these three polypeptides are illustrated in Figure 33.

The exact nature of the transition of PCMS is very dependent on polymer concentration and degree of polymerization. High concentration favors intermolecular sheet aggregation. Long chains allow sheet aggregation among sheets that may form from residues far removed from each other along the chain but nevertheless close enough to interact even in very dilute solutions (Maeda, et al. 1972). Both are complicating factors to be avoided.

The interaction of catechins and procyanidins with two additional homopolypeptides was observed. These are poly (L-lysine) and poly (L-arginine). Their structures are also shown in Figure 33.
poly (S-carboxymethyl cysteine) \[ R = \text{CH}_2\text{SCH}_2\text{COOH} \]

poly (glutamic acid) (L and D) \[ R = \text{CH}_2\text{CH}_2\text{COOH} \]

poly (L-lysine) \[ R = \text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2 \]

poly (L-arginine) \[ R = \text{CH}_2\text{CH}_2\text{CH}_2\text{NH}-\text{C}-\text{NH}_2 \]

Figure 33. Structure of homopolypeptides used in this study. Side chains are drawn in the unionized form.
CHAPTER SEVEN

THEORY OF EXPERIMENTAL METHODS

Ultraviolet-Visible Absorbance

A section on the theory of absorbance is most appropriately opened with a brief discussion of the Lambert-Beer Law. This law states that in dilute solutions, the intensity, $I$, of light that has traversed a distance $l$ through a solution of concentration $c$ is related to the intensity of the incident light, $I_0$, by

$$I = I_0 \cdot 10^{-ecl}. \quad (46)$$

This can be rewritten as

$$\frac{I}{I_0} = 10^{-ecl} \quad (47)$$

and since

$$A = \log \frac{I}{I_0} \quad (48)$$

$$A = ecl. \quad (49)$$

The polymers used in this work do not absorb in the region being studied by fluorescence; that is, 250 nm to 450 nm. At the lowest wavelength there is a tail of absorbance due to the peptide bond $n$ to $\pi^*$ transition at 220 nm, but for the most part this is cancelled out by subtracting it as a baseline. Catechins and procyanidins absorb strongly in the region defined.

Absorbance is measured on a Hewlett-Packard 8451A Diode Array Spectrophotometer. This single light path
instrument is capable of measuring the baseline and storing it in memory as a reference. The reference is automatically subtracted from all subsequent measurement until one chooses to change it. There is no temperature regulator on the Diode Array. A diagram of this instrument is shown in Figure 34.

Fluorescence

Absorbance is due to light of a select energy exciting an electron to a higher state. Fluorescence is a radiative transition between the same electronic levels but in the reverse direction. The most frequent means of illustrating these transitions in poly-atomic molecules is by way of a Jablonski diagram (Figure 35). The heavy lines represent the electronic states; \( S_0 \) is the ground state, \( S_1 \) is the first excited state, \( S_2 \) is the second excited state, and so on. The lighter lines represent vibrational states of each electronic state. In this diagram, transitions are drawn vertically to establish that there is no significant displacement of the nuclei. Because the transitions occur very quickly, on the order of \( 10^{-15} \) seconds, there is insufficient time for movement of the relatively large mass of the nuclei. The absence of nuclei displacement is known as the Franck-Condon Principle.

Molecules absorb light of the wavelength that corresponds to the spacing between electronic energy levels. Absorption can be to any of the vibrational states, however before any fluorescent radiation can occur,
Figure 34. Block diagram of the Hewlett-Packard 8451A Diode Array Spectrophotometer.
Energy intersystem crossing

Figure 35. Jablonski Diagram
there is a relaxation to the lowest vibrational level of $S_1$. From here a number of things can happen. Most notable among the possibilities are intersystem crossing, radiationless decay, and fluorescence.

Intersystem crossing refers to a radiationless transition connected with a change in multiplicity. This transition is often between the singlet $S_1$ and the triplet $T_1$. The triplet state cannot decay to $S_0$ until a highly forbidden spin reversal occurs. Consequently, the emission is not immediate. This delayed emission is known as phosphorescence.

Radiationless decay which does not involve a change in spin multiplicity is another path from the excited state. Interaction with the solvent may result in the release of energy without radiation. The excited moiety may collide with other molecules and thus lose its energy. It is also possible that the excited molecule undergoes an excited state reaction.

The path of most interest for this discussion is fluorescence. This is when the relaxation from $S_1$ to $S_0$ is accompanied with the emission of light. There are several important characteristics of fluorescent emission. Discussion of some of these follows.

Light emitted by an excited molecule is at a higher wavelength, a lower energy, than the light absorbed. Called the Stokes' shift, this phenomenon is due to several factors. One of the factors is the decay to the lowest
vibrational level of $S_1$ before emission to higher vibrational levels of $S_0$. A second factor arises from decay to the lowest vibrational level of this electronic energy level. The difference in the energy of transition is clearly less in path 2 of Figure 35 than in path 1. Other factors contributing to the Stokes' shift are solvent effects, excited state reactions, and collision with other molecules.

Another rule of fluorescence is that the emission spectra are independent of excitation wavelength. As long as the excitation wavelength is within the same band, i.e., causes the same transition, the shape of the spectra should be the same. The intensity of emission, however, may alter.

Since the same transitions are involved in both absorption and emission, and the vibrational energy levels of $S_0$ and $S_1$ are similar, the excitation and emission spectra of a given fluorophore will often be the mirror images of each other. That is, excitation from the lowest vibrational state of the electronic ground state to the various vibrational states of higher electronic states is the mirror image of the reciprocal action.

Fluorescence emitted from the sample is polarized in three dimensions. The detector is able to detect emitted light polarized in two dimensions. Emission polarized in the third dimension can be accounted for by orienting excitation and emission polarizers in the 'magic
angle' positions. The rational for this is presented in Appendix VI. These angles are found to be 0° for excitation and 54.7° for the emission polarizer or vice versa. The use of polarizers does greatly decrease the intensity of exciting light that reaches the sample and the intensity of emitted light that reaches the photomultiplier tube. If the exciting light is entirely unpolarized, one polarizer in the emission channel set at 35.3° accomplishes the same goal with a less drastic decrease in light intensity.

The quantum yield of a fluorophore is defined as the ratio of the number of photons emitted to the number absorbed. Looking again at the Jablonski Diagram, consider the rate of fluorescent emission to be \( \Gamma \) and the rate of all other decays to sum to \( k \). Then the quantum yield can be defined as

\[
Q = \frac{\Gamma}{\Gamma + k}.
\]  

(50)

Lifetime is defined as the average time the molecule spends in the excited state prior to its return to the ground state. That is equal to

\[
\tau = \frac{1}{\Gamma + k}.
\]  

(51)

Intrinsic lifetime is defined as the lifetime in the absence of any radiationless decay, that is

\[
\tau_o = \frac{1}{\Gamma}.
\]  

(52)

It is easy to see that

\[
Q = \frac{\tau}{\tau_o}.
\]  

(53)

Steady-state fluorescence is measured on the SLM
Figure 36. Block Diagram of the SLM 8000C Spectrofluorometer.
8000C Spectrofluorometer (Figure 36). Excitation is with a Xenon arc lamp, specifically an Osram XBO 450 W12 ozone free 450 watt lamp. There is a double monochromator on the excitation side. Each of the three slits can be adjusted to give a bandpass of 0.5, 1, 2, 4, 8, or 16 nm. The emission side has a single monochromator with slits to be adjusted as for the double monochromator. In general, the practise is to set the scanning monochromator for the desired resolution and use the slits on the monochromator not scanning for intensity control.

As the light leaves the excitation monochromator, it encounters a beam splitter. As a result of this, light is sent to a reference channel. This reference beam encounters the quantum counter rhodamine B which effectively absorbs light of all wavelengths and emits fluorescence. The light that impinges on the reference PMT is therefore a measure of the light intensity at each wavelength. The reference channel corrects for wavelength dependence of lamp intensity or monochromator efficiency as well as fluctuations in source intensity.

Light not split to the reference channel goes to the sample chamber which is thermostatted. Emission is measured perpendicular to the direction of excitation. The sample PMT is set to photon counting. Photons are counted at each wavelength for many seconds and averaged to yield counts per second. Counts are recorded as the ratio of the number of photons emitted per second to the intensity of
the light that reaches the reference channel. The PMT counting photons emitted by the sample is water-cooled to minimize dark current.

Circular Dichroism

The importance of optical rotation as a tool in structural chemistry lies in its extraordinary sensitivity to molecular geometry. Optical rotation has no primary source within the electronic structure of a molecule but springs directly from the relative orientation of different sets of orbitals.

The integrated intensity of an electronic absorption band may be shown to be directly proportional to the square of the magnitude of the electric dipole moment for the transition. Circular dichroism arises from an analogous quantity, the rotatory strength, \( R \). The factor which determines \( R \) for a transition from state \( \text{a} \) to state \( \text{b} \) is the imaginary part of the complex scalar product of transition matrix elements.

\[
R_{\text{ab}} = -\text{Im}(\langle \text{a} | \text{p} | \text{b} \rangle \cdot \langle \text{b} | \text{m} | \text{a} \rangle)
\]  

(54)

Using a vastly simplified model which has only two groups of electrons that make up a chromophore and only two possible electronic transitions, one can imagine three ways in which rotatory strength can be developed by the two groups. (1) Two transitions occur in one chromophore, one magnetic and the other electric. In other words, one electron moving in an asymmetric field leads to optical activity. (2) Both groups have a single electric transition
and because of their proximity these are coupled by their dipolar fields to produce a magnetic moment. (3) One group has a magnetic transition and the other an electronic transition which couple to produce optical activity.

The first possibility was proposed by Condon, Altar, and Eyring in 1937 and is called the single electron theory. The second is attributed to Kuhn and Kirkwood—also in 1937—and is called the coupled oscillator theory. When the two groups are identical, the excited states are degenerate and the two groups participate equally in the resulting coupled transitions. This modification was proposed by Moffitt and is called the exciton theory. The third theory was proposed by Woody and Tinoco (1967), Hohn and Weigang (1968), and Schellman (1968). It can be called a μ-m theory. The actual mechanism is probably some combination of these.

The peptide bond in polypeptides has a strong π to π* transition and a weaker n to π* transition. Since the π to π* transition has components parallel and perpendicular to the helix axis, the circular dichroism of a polypeptide in a helical conformation is expected to have three bands, two due to the π to π* transition and one due to the n to π* transition. However, the close proximity of the many peptide bonds results in an exciton interaction (as proposed by Moffitt) and the component of the π to π* transition that is perpendicular to the helix axis is split. Therefore four bands are predicted by Moffit's
theory (Charney 1974). If one assumes that the band due to the n to \( \pi^* \) transition is too weak to observe, three remain. A strong band at 222 nm is due to the component of the \( \pi \) to \( \pi^* \) transition that is parallel to the helix axis. The remaining bands appear as a peak at 195 nm and a shoulder at 175 nm. The theory concludes that they should be of opposite sign and that the sum of their rotational strengths should be approximately zero. CD spectra of helical polymers show that these last two bands are not of opposite sign. This discrepancy may lead one in search of a new interpretation. A modification of Moffit's theory is one presented by Holzwarth (1964) who assigns the band at 222 nm to the n to \( \pi^* \) transition, the band at 206 nm to the parallel component of the \( \pi \) to \( \pi^* \) transition, and a band at 190 nm to the perpendicular component of the \( \pi \) to \( \pi^* \) transition. The important point here is that helices do present a strongly positive circular dichroism band at 195 nm and a negative band at 222 nm. This knowledge is sufficient to identify the \( \alpha \)-helix conformation of polypeptides.

Calculated spectra for the \( \alpha \)-helix conformation agree well with experimental results (Figure 37). Spectra are less predictable, however, for the \( \beta \)-sheet. While all antiparallel \( \beta \)-sheets have hydrogen-bonds between strands that run in alternating directions, the length and number of strands is not fixed. Calculations that have been done for \( \beta \)-sheets of 2 strands each 10 residues long predict a
Figure 37. Comparison of the experimental and theoretical circular dichroism curves for "infinite" right-handed α helical poly (L-alanine) in trifluoroacetic acid, 98.5:1.5 v/v (Woody 1968).
positive band at 200 nm of $[\Theta] \approx 2 \times 10^4 \text{ deg} \cdot \text{cm}^2/\text{dmole}$ and a negative band at 222 nm of $[\Theta] \approx -1.5 \times 10^4 \text{ deg} \cdot \text{cm}^2/\text{dmole}$ (Figure 38, Pysh 1968).

A calculated spectrum for the random coil has been even more elusive. There are too many uncertainties surrounding the structural characterization of polypeptides which lack long-range order. Very typically, the random coil has a strong negative band near 197 nm and a less intense positive band near 217 nm. The cd for all three conformations for poly (L-lysine) is shown in Figure 39 (Greenfield and Fasman 1969).

The measurement of circular dichroism is accomplished by exciting the sample with plane polarized light which is actually the resultant of two co-terminus beams of right and left-circularly polarized beams. That plane polarized light is the resultant of these two beams can be shown by drawing the vectors and summing them. At a given wavelength, a disymmetric compound will affect each of these circularly polarized beams differently. There are two sources of this difference. First, the velocity of light is altered more for one direction of polarization than for another—that is, the index of refraction for light circularly polarized in one direction is different than the index of refraction for light circularly polarized in the opposite direction when the molecules being encountered are dissymmetric. The second source of difference is the molar absorbance. At a given wavelength
Figure 38. Comparison of experimental and theoretical circular dichroism curves for high molecular weight poly (L-lysine) and an antiparallel $\beta$ pleated sheet of two strands, each ten residues long (Pysh 1970).
Figure 39. Circular dichroism spectra of poly (L-lysine) in the α helical (1), β (2), and random (3) conformations (from Greenfield and Fasman 1969)
a disymmetric compound will absorb more from light circularly polarized in one direction than from light circularly polarized in the other direction. Because of these effects, the new resultant describes an ellipse instead of a line. The principle axis of this ellipse defines the plane of the rotated beam and the minor axis defines the absolute value, $|\varepsilon_1 - \varepsilon_r|$. This value is the circular dichroism. Circular dichroism is most often given in units of ellipticity, $[\Theta]$, which is approximately given by the following equation.

$$[\Theta] = 3300 |\varepsilon_1 - \varepsilon_r|$$  \hspace{1cm} (55)

The derivation of this equation is given in Appendix III.

Circular dichroism was measured on a Jasco J-500A Spectropolarimeter (Figure 40). The light source is a 450W Xenon arc lamp mounted in a water-cooled housing. Light from the lamp is focused by a spherical mirror onto the entrance slit. A segment of the optical system from the entrance slit to an intermediate slit is designated as the first monochromator. In between these two slits, light is reflected from a spherical mirror to a crystal quartz prism, then to another mirror. After passing through the second slit, light hits another mirror, another crystal quartz prism a fifth mirror, and finally through the third slit. The two quartz prisms have different axial direction with respect to each other, therefore the light beam passing through the monochromators is monochromatic and, at the same time, it is linearly polarized light oscillating
M₀, M₁, M₂, M₃, M₄, M₅: mirror
LS: Light source
S₁, S₂, S₃: Slit
P₁: First prism (Horizontal axis)
P₂: Second prism (Vertical axis)
O-ray: Ordinary ray
E-ray: Extra-ordinary ray

Note: For the Model J-500A, filter F is used. However, for the Model J-500C, Rochon prism PO is used instead of filter.

Figure 40. A diagram of the JASCO 500-A.
in the horizontal direction. This light beam is modulated to left and right circularly polarized light by a cd modulator. The cd modulator takes advantage of the piezoelectric effect present in crystals such as the quartz prism in the J-500A.

The piezoelectric effect arises from the fact that certain crystals are birefringent; that is, they allow light to pass through in two planes which are perpendicular to each other at different rates. If light is travelling along the z axis, this crystal allows light polarized along the x axis to pass, and more slowly, allows light polarized along the y axis to pass. The length of the crystal can be chosen such that light on the y axis is delayed 1/4 wavelength relative to that on the x axis. This 1/4 wavelength delay results in circularly polarized light. In practice the direction of polarization is rapidly alternated by exchanging the positions of the slow and fast axes of the birefringent plate. Crystals exhibiting the piezoelectric effect often have an additional characteristic. The position of the fast and slow axes can be exchanged by reversing the direction of an electric field traveling through them. This electro-optic effect is sometimes called Pockel's effect.

The photomultiplier tube detects the decrease in the intensity of right and left circularly polarized light, and sends a signal to the recorder that is proportional to the ellipticity. An equation which accounts for instrument
sensitivity, solution concentration, and the pathlength of the sample cell yields results in units of ellipticity.

This equation is

\[ [\theta] = n \left[ \frac{(S \times MW)}{(l \times c)} \right] \]  \hspace{1cm} (56)

where \( n \) is the distance of the plotted line from the baseline in cm, \( S \) is the sensitivity setting on the instrument in m°/cm, \( MW \) is the (mean residue) molecular weight of the dissymetric molecule, \( l \) is pathlength of the cell in mm, and \( c \) is concentration in mg/ml of the dissymetric molecule. Units of \([\theta]\) are deg·cm²/mole.
CHAPTER EIGHT

MATERIALS

(+)-Catechin and (-)-epicatechin were purchased from Sigma Chemical Co. Procyanidins B-1, B-7, and A-2 were synthesized by Richard Hemingway.

The sodium salts of poly (L-glutamic acid) and poly (D-glutamic acid) were purchased from Sigma. Their molecular weights (and degree of polymerization) as reported by Sigma are 51000 (340) and 66000 (380) respectively. Measurements were also made on another sample of poly (L-glutamic acid) purchased from Miles-Yeda Ltd. This sample has a reported molecular weight of 106000 and the degree of polymerization is 822.

Poly (S-carboxymethyl-L-cysteine) (PCMS), was provided by Hiroshi Maeda of Nagoya University, Nagoya, Japan. The first step of this synthesis is the preparation of poly-S-carbobenzoxy methyl-L-cysteine by the polymerization of S-carbobenzoxy methyl-N-carboxy-L-cysteine anhydride. Poly-S-carbobenzoxy methyl-L-cysteine is suspended in acetic acid saturated with hydrogen bromide. After overnight stirring at 0°C, the solution was pumped out to make it free from excess hydrogen bromide. The solution was evaporated in vacuo to a small volume and the residue treated with ethyl ether. The white precipitate
was suspended in ethyl ether, washed and separated by centrifugation. The procedure was repeated with a great amount of ethyl ether, acetone, and ethyl ether, successively, until the supernatant became colorless (Ikeda 1967). Samples are of molecular weight (degree of polymerization) 53,000 (330) and 90,000 (560).

Poly (N\(^\text{5}\)-hydroxypropyl-glutamine), (PHPG), was purchased from Miles-Yeda Ltd. Molecular weight of the sample used was 45000 and degree of polymerization was 245.

Poly \(\gamma\)-benzyl-L-glutamate (PBLG), a necessary precursor for the synthesis of poly N-hydroxyalkyl glutamines, was purchased from Sigma. Poly (N\(^\text{5}\)-hydroxybutyl glutamine) (PHBG) was synthesized by Robert McCord via the reaction of PBLG in dioxane at 60°C with 4-amino-1-butanol. PBLG is dissolved in dioxane at 60°C with mechanical stirring. The amino alcohol is added slowly. Stirring is continued at 60°C in a closed system for 20 to 40 hours. The reaction mixture is then poured into 5 volumes of chloroform. The resulting precipitate is washed with chloroform and ether (Lotan 1965). Absorbance at 257 nm is due to remaining benzyl groups. This was very small in all samples measured (\(< 1\%\) benzyl groups remaining). Samples used have a molecular weight (degree of polymerization) of 200,000 to 400,000 (1000 – 2000).

The copolymer poly (N\(^\text{5}\)-hydroxy(ethyl:butyl) glutamine) (PH(E:B)G) was prepared by Erin Hawkins. The synthesis involves the aminolysis of PBLG using a mixture
of hydroxyethanolamine and hydroxybutanolamine. In the fraction used, 63% of the sidechains were butyl (Hawkins 1975).

The copolymer poly \((N^5\text{-hydroxybutyl glutamine}:L-arginine)\) \((P(HBG:Arg))\) was prepared by Leed in the lab of Scheraga. The first step of this synthesis involves the copolymerization of the \(N\)-carboxy anhydrides of \(N^5\text{-tert}-\)butylcarbonyl-L-ornithine and \(\gamma\text{-benzyl-L-glutamate}\) in dioxane using sodium methoxide as an initiator. The resulting copolymer was converted to the 4-hydroxybutyl glutamine derivatives by treatment with 4-amino-1-butanol. The 5-amino protecting group of L-ornithine was removed using 3N HCl. The resulting water-soluble copolymers were treated with O-methylisourea in water at a pH 10.0 at 0-4°C (Konishi 1977). The samples were fractionated in our lab by T. H. Lin. The procedure used for fractionation includes dissolution in 0.9 M NaCl followed by the addition of nine parts of methanol. This mixture is titrated with ether until a turbidity is visible. The mixture is centrifuged, the pellet is redissolved in water, dialyzed against water until the conductivity of the copolymer solution no longer decreases, and lyophilized. The supernatant from the centrifugation is treated with repeated ether titrations, each titration resulting in another fraction to be dialyzed and lyophilized.

A number of synthetic polymers were purchased from Scientific Polymer Products, Inc. These include poly
ethylene glycol (PEG), poly vinyl pyrrolidone (PVP), poly acrylamide (PA), and hydroxypropyl cellulose (HPC).
CHAPTER NINE

METHODS

All solutions are prepared with deionized distilled water. Polymers and tannins are stored in a jar containing dessicant in the freezer and allowed to come to room temperature before weighing the sample. Masses between 2.5 mg and 50 mg are weighed on a Cahn Electrobalance Model DTL. Masses between 50 mg and 1000 mg are weighed on a Mettler PE 160. In all cases, stock solutions are prepared and dilutions made from these because small volume dilutions are more accurate than small mass weighings (final concentrations are necessarily low). This also assures greater precision between two solutions of the same fluorophore whose emissions are compared.

Solutions prepared for fluorescence measurements contain polymer and fluorophore. Polymer concentrations are taken to the upper limit of solubility but not greater than 10 mg/ml. Table 16 lists the polymers and the maximum concentrations of solutions on which measurements were made. Solutions of fluorophores are made so that the absorbance at the exciting wavelength does not exceed 0.1. This is to avoid self-absorption. Since 280 nm is the absorbance maximum, it is the exciting wavelength. In the case of (+)-catechin, (-)-epicatechin, and the procyanidins the
Table 16.—Maximum polymer concentrations

<table>
<thead>
<tr>
<th>Polymer</th>
<th>Concentration mg/ml</th>
<th>Concentration mole res/l</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHBG</td>
<td>10</td>
<td>0.050</td>
</tr>
<tr>
<td>PHPG</td>
<td>10</td>
<td>0.054</td>
</tr>
<tr>
<td>PH(E:B)G</td>
<td>3.0</td>
<td>0.016</td>
</tr>
<tr>
<td>P(HBG:Arg)</td>
<td>0.4</td>
<td>0.002</td>
</tr>
<tr>
<td>pGlu</td>
<td>10</td>
<td>0.078</td>
</tr>
<tr>
<td>PCMS</td>
<td>0.015</td>
<td>0.000093</td>
</tr>
<tr>
<td>PEG</td>
<td>8.0</td>
<td>0.24</td>
</tr>
<tr>
<td>PVP</td>
<td>8.0</td>
<td>0.072</td>
</tr>
<tr>
<td>PA</td>
<td>1.0</td>
<td>0.014</td>
</tr>
<tr>
<td>HPC</td>
<td>1.0</td>
<td>0.0043</td>
</tr>
<tr>
<td>pLys</td>
<td>10</td>
<td>0.068</td>
</tr>
<tr>
<td>pArg</td>
<td>10</td>
<td>0.057</td>
</tr>
</tbody>
</table>
final concentration of solutions is near 0.008 mg/ml.

For circular dichroism measurements the concentration of polymer solutions is between 0.1 mM (for PCMS) and 1.0 mM. Catechin and procyanidin concentrations are 0.03 mM, 0.15 mM and 0.30 mM.

Absorbance is measured on the Hewlett Packard 8451A Diode Array, a single light path instrument. For this reason, care must be taken when determining the baseline to use the same cell with the same face toward the light source for sample as for baseline measurement. A 1 cm quartz cell was used. When the absorbance of polymers is reported, the baseline is the solvent, i.e., water. When the absorbance of catechins or procyanidins alone is reported, the baseline is water. When the absorbance of catechins or procyanidins in a solution containing polymer is recorded, the baseline is a solution with the same concentration of polymer. There is no temperature regulation for absorbance measurements. Temperature is assumed to be room temperature, here maintained at 20 ± 2°C.

Fluorescence measurements are made on an SLM 8000C spectrofluorometer. This instrument is described in a previous chapter. Baselines are measured for each sample according to the same pattern described for absorbance measurements. The instrument is set such that slit widths on the excitation and emission sides are 4 nm. The voltage to the reference channel is set such that at the maximum of
the emission band, the counts approach but do not exceed 10,000 and the ratio of channel A to channel B is less than but close to 1. Depending on the fluorophore, this is found to be between 380 and 405 volts. Polarizers are set to the 'magic angle' conditions, that is, 0° on excitation side and 54.7° on the emission side. Emission is measured perpendicular to excitation. The temperature in the sample chamber is regulated with an Allied Fischer water bath, and temperature is recorded. A quartz cell with a 1 cm pathlength is used. The fluorescence of the baselines is counted for one second. The fluorescence of samples is counted for 5 seconds and averaged to be reported as counts per second. Measurement is made in a ratio mode; that is, values are recorded as the ratio of counts emitted from the sample to light intensity that reaches the reference channel.

The degree of intermolecular interaction between catechin and the polymers is recorded as the ratio of the integrated intensity of emission from a solution of catechin in polymer, I, to the integrated intensity of emission from a solution of catechin alone, I₀. This ratio is corrected for any differences in catechin concentration by making use of the fact that intensity is inversely proportional to the absorbance at the exciting wavelength. The equation used is

\[ \frac{Q}{Q₀} = \frac{I}{I₀} \times \frac{A₀}{A} \]

(57)

where I and I₀ are as defined in the previous paragraph, A₀
is the absorbance at 280 nm of a solution of catechin without polymer and \( A \) is the absorbance at 280 nm of a solution of catechin with polymer. All emissions that are compared have been measured on the same day under as close to the same conditions as possible.

Circular Dichroism spectra are measured on a JASCO J-500A spectropolarimeter. Measurement is done with temperature carefully regulated by a Lauda water bath. Cells with 1 mm, 2 mm, 5 mm, and 10 mm pathlength were used. Since circular dichroism is a measure of the conformation of the polymers, the baselines used are as follows. For a solution of catechin and polymer, the ellipticity of a solution of the same concentration of catechin alone is measured. Then the solution of catechin and polymer is measured in the same cell. Measured baselines are subtracted manually when analyzing results. To determine ellipticity, the following equation is used.

\[
\theta = \frac{(n \cdot S \cdot MW)}{(c \cdot l)}.
\]  

(58)

\( S \), which is in m\(^2\)/cm, was set to 1 or 2, depending on solution concentration. \( MW \) is the molecular weight. The denominator is the product of the solution concentration in mg/ml (or g/dl) and the cell's path length in mm. The distance in cm between the baseline and the samples' circular dichroism on the chart paper defines \( n \) (in cm). The time constant of the instrument is set to 16 seconds. The product of the chart speed and wavelength expansion is (0.5 cm/s) x (5 nm/cm) or 2.5 nm/s. The positions of the
peaks at this speed were not shifted from their positions at the lower speed of 1 nm/s.
CHAPTER TEN

RESULTS

Absorbance

The absorbance spectra of (+)-catechin, (-)-epicatechin, and procyanidin B-1 are shown in Figure 41. The three solutions measured were 0.008 mg/ml in the three small proanthocyanidins. The absorbance at 280 is near 0.7 for all three. In solutions containing polymer and catechin or procyanidin, a solution of the same concentration polymer is used as the baseline for absorbance measurements.

Absorbance measurements are reported for all the polymers of this study from 190 nm to 450 nm. The spectra are shown in Figure 42. They are taken on very concentrated solutions of polymer. There are two overlapping reasons for this: 1) The high concentrations are those to which (+)-catechin and procyanidin B-1 are added (see Table 16), so the absorbance behavior of solutions of these concentrations at 280 nm must be characterized. 2) Interest is in behavior of the tail of the peak at 200 nm; specifically interest is in the behavior at 280 nm. The tail is emphasized at high concentrations. Polypeptides show the expected peak at 200 nm due to the $\pi$ to $\pi^*$ transition. If the tail of this peak
Figure 41. Absorbance spectra of (a) (+)-catechin, (b) (-)-epicatechin, and (c) procyanidin B-1.
Figure 42. Absorbance spectra of the polymers used in this study. (a) poly vinyl pyrrolidone; (b) polyacrylamide; (c) poly ethylene glycol; (d) hydroxypropyl cellulose.
Figure 42 (cont.) Poly Glutamines: (e) poly [N-hydroxybutyl glutamine], molecular weight = 60,000; (f) poly [N-hydroxybutyl glutamine], molecular weight = 400,000; (g) copoly [N-hydroxyethyl:hydroxybutyl glutamine]; (h) poly [N-hydroxypropyl glutamine]; (i) copoly [(N-hydroxybutyl glutamine):arginine]
Figure 42 (cont.) Polypeptides: (j) poly (L-glutamic acid); (k) poly (D-glutamic acid); (l) poly [S-carboxymethyl cysteine]; (m) poly (L-arginine); (n) poly (L-lysine).
is too long, there may be an effect on the absorbance of the catechin and procyanidin chromophores at 280 nm. Since the extinction coefficient at 280 nm for (+)-catechin, and (-)-epicatechin, is so much greater than the extinction coefficient of the polymers (Table 17) it is not expected that absorbance due to polymers will affect the apparent absorbance of catechins.

Table 17.—Absorption Coefficients of Polymers, (+)-Catechin, and (-)-Epicatechin

<table>
<thead>
<tr>
<th>Polymer</th>
<th>( e \ (1 \ mol^{-1}cm^{-1}) ) at 280 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHBG</td>
<td>1.7</td>
</tr>
<tr>
<td>PHPG</td>
<td>2.1</td>
</tr>
<tr>
<td>PH(E:B)G</td>
<td>91</td>
</tr>
<tr>
<td>P(HBG:Arg)</td>
<td>0.41</td>
</tr>
<tr>
<td>p-L-Glu</td>
<td>0.72</td>
</tr>
<tr>
<td>p-L-Arg</td>
<td>1.7</td>
</tr>
<tr>
<td>p-L-Lys</td>
<td>1.5</td>
</tr>
<tr>
<td>PVP</td>
<td>2.9</td>
</tr>
<tr>
<td>PEG</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>PA</td>
<td>1.8</td>
</tr>
<tr>
<td>HPC</td>
<td>4.1</td>
</tr>
<tr>
<td>monomer</td>
<td></td>
</tr>
<tr>
<td>(+)-catechin</td>
<td>3800</td>
</tr>
<tr>
<td>(-)-epicatechin</td>
<td>4600</td>
</tr>
</tbody>
</table>

Poly glutamines are synthesized from poly-L-benzyl glutamate. One of the steps in this synthesis is debenzylination via amide-ester interchange. If this step is not allowed to go to completion, hydroxyalkyl groups are interspersed with benzyl groups. This would affect conformation transitions and tannin interactions. Benzyl
groups absorb at 257 nm. The absence of a band at 257 nm upholds the notion that debenzylation is complete.

At high concentrations, polymer aggregates may form. Aggregates result in broad featureless bands due to the scattering of light at all wavelengths. The bands in the spectra shown in Figure 42 do not provide evidence for aggregation.

To study the interaction of (+)-catechin, (-)-epicatechin, and procyanidin B1 with the synthetic polymers listed in the Materials section, two spectroscopic techniques are employed. First, circular dichroism demonstrates the effect of the procyanidins on conformational transition undergone by the polypeptides. Second, if the presence of polymer in a solution of the fluorescent procyanidins affects the emission of the procyanidins, that is certainly indicative of some interaction.

**Catechins and Procyanidin B-1**

Before discussing results for solutions of polymer and procyanidin, it is necessary to determine the spectroscopic characteristics of the procyanidins alone. As each polymer is discussed, its circular dichroic characteristics will be described. None of the polymers studied exhibit fluorescence in the range studied.

**Circular Dichroism**

A plot of the ellipticity of (+)-catechin, (-)-epicatechin, and procyanidin B1 is shown in Figure 43.
Figure 43. Ellipticity of (a) (+)-catechin, (b) (-)-epicatechin, and (c) procyanidin B-1.
Figure 43. (cont.)
The spectral range does not extend below 190 nm even though the instrument's range is to 180 nm. This is because the catechins and procyanidin absorb a significant amount of light at this wavelength. While the instrument can measure ellipticity at lower wavelengths, there must be a sufficient amount of light getting through the sample to the PMT to be measured. Another characteristic in cd spectra is apparent in Figure 43. At lower wavelengths, the intensity of light is less, and the noise in the signal increases. For this reason, the error in the spectra varies with wavelength, being the greatest at lower wavelengths and decreasing as the wavelength is increased.

Circular dichroism for these compounds has not been reported in water, but has been determined in methanol (Barrett, et al. 1979). Using the equation

\[
[\theta] = 3300 \Delta \varepsilon
\]  

(59)
to convert \([\theta]\) to \(\Delta \varepsilon\), one can compare these results. Such an analysis is tabulated in Table 18. Agreement is rather good despite the different solvents. There is more of a difference for procyanidin B1 than for the monomers.

In Chapter 6 a brief discussion of the source of the cd bands for 4-arylflavan-3-ols was given. Smaller bands in the range of 210 to 240 nm and bands at 280 nm are said to be due to the stereochemistry of the C-4 and C-2 carbons respectively. A large band at 210 nm is due to the transition moment vector of the A and A' rings (Gaffield, et al. 1985).
Table 18.— Circular Dichroism of Catechins and Procyanidin B-1

<table>
<thead>
<tr>
<th></th>
<th>in methanol*</th>
<th>in water</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$, nm</td>
<td>$\Delta\varepsilon$</td>
</tr>
<tr>
<td>(+)-catechin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>21.5</td>
<td>200</td>
</tr>
<tr>
<td>205</td>
<td>-4.7</td>
<td>225</td>
</tr>
<tr>
<td>227</td>
<td>-2.23</td>
<td>[275]</td>
</tr>
<tr>
<td>275</td>
<td>-0.52s</td>
<td>280</td>
</tr>
<tr>
<td>282</td>
<td>-0.87</td>
<td>[290]</td>
</tr>
<tr>
<td>(-)-epicatechin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>8.9</td>
<td>200</td>
</tr>
<tr>
<td>206</td>
<td>-13.8</td>
<td>210</td>
</tr>
<tr>
<td>224</td>
<td>-3.0s</td>
<td>239</td>
</tr>
<tr>
<td>239</td>
<td>1.57</td>
<td>240</td>
</tr>
<tr>
<td>275</td>
<td>-0.65</td>
<td>278</td>
</tr>
<tr>
<td>281</td>
<td>-0.71</td>
<td></td>
</tr>
<tr>
<td>308</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>procyanidin B-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>0</td>
<td>195</td>
</tr>
<tr>
<td>200</td>
<td>67.5</td>
<td>203</td>
</tr>
<tr>
<td>236</td>
<td>10.9s</td>
<td>210</td>
</tr>
<tr>
<td>267</td>
<td>1.27</td>
<td>238</td>
</tr>
<tr>
<td>276</td>
<td>1.31</td>
<td>[280]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>284</td>
</tr>
</tbody>
</table>

* Source: Barrett, et al. 1979
s = signifies a shoulder to a peak
values in brackets indicate measurements made after solutions had turned yellow.
The difference between the structures of (+)-catechin and (-)-epicatechin occurs at C-2. There is, however, no difference between the two spectra in the bands at 280 nm. The most obvious difference in the cd spectra occurs at 210 nm. In the case of monomers it appears that the $\pi$ to $\pi^*$ transition occurs at 200 nm. This is very similar between the two isomers, although more strongly positive for (-)-epicatechin.

Note that the circular dichroism of dimer is not simply the sum of the spectra of the monomers. The dimer spectrum fits the general pattern proposed by Gaffield (1985), but the shoulder at 200 nm is unexpected. It may be due to the $\pi$ to $\pi^*$ transition in the A ring of (-)-epicatechin which is not covalently bound as is the A' ring of (+)-catechin in this dimer. The shoulder height is slightly more than half the height of the peak at the same wavelength for (-)-epicatechin. In the dimer there are half as many unbound A rings at any concentration. The positive peak at 240 nm is much stronger in the dimer where the C-4 carbons of (-)-epicatechin are all involved in the covalent bond. This clearly has some effect on the circular dichroism.

Circular dichroism was measured on solutions of (+)-catechin and procyanidin B-1 that had turned yellow due to the oxidation of the phenol. In general, the spectra were unaltered, but new small peaks appear at 275 and 290 nm for (+)-catechin and at 280 nm in the case of
procyanidin B-1. These are indicated on Table 18 with brackets. It may be that some of the peaks reported for procyanidins in methanol are from solutions that were not fresh, i.e., slightly yellow. Greatest confidence can be placed in values at low wavelengths which are least affected by the molecular rearrangement that results in the color change of the solutions.

When the ellipticity of polymers is measured in a solution containing polymer and procyanidin, a solution of the same concentration procyanidin is used as the baseline. The procyanidins have significant ellipticity at 190 nm and 222 nm, two positions of descriptive bands for polypeptides. Therefore care must be taken to keep the concentration of procyanidin low enough to limit its contribution to the circular dichroism of a solution of polymer and procyanidin. To accomplish this, the stipulation was placed on catechin concentration that it contribute less than 10% to the signal at the peaks.

Fluorescence

The fluorescent emission of (+)-catechin and procyanidin B1 in water has a maximum near 322 nm. These spectra are shown in Figure 44. The two solutions measured have the same absorbance at 280 nm. The spectra are corrected for instrument response by multiplying them onto the correction factors for the PMT of the instrument used.

Bergmann, et al. (1986) has indicated that the quantum yield of procyanidins is greater in hydrophobic
Figure 44. Corrected fluorescent emission of (+)-catechin and procyanidin B-1. Samples measured under identical conditions with nearly the same absorbance. The same emission scale applies to both spectra.
solvents. To determine the effect of additional polymers on the fluorescence of procyanidins, the emission and absorbance of solutions of procyanidin alone and solutions of procyanidin with polymer are measured under conditions as close to identical as possible. The results are compared according to the equation

\[
\frac{Q}{Q_0} = \frac{I}{I_0} \cdot \frac{A_0}{A}
\]  

(60)

where the zero subscript indicates a solution of procyanidin alone, absorbance is measured at 280 nm and I refers to the intensity integrated from 285 nm to 410 nm. These limits completely encompass the emission band without picking up scattering from the excitation light at 280 nm. If \( Q/Q_0 \) is less than one, the fluorescence of the procyanidin is decreased by the presence of polymer. If \( Q/Q_0 \) is greater than one, the fluorescence of procyanidin is enhanced by the presence of polymer. If \( Q/Q_0 \) is equal to one, the polymer has no effect on the procyanidin's fluorescence.

Although solutions are made so as to keep the concentration of procyanidin very similar in two solutions being compared (one with and the other without polymer), comparing absorbance accounts for any errors in solution preparation as well as for any effect of polymer on absorbance. In most cases

\[
\frac{Q}{Q_0} = \frac{I}{I_0};
\]  

(61)

that is, \( A_0/A \) is very close to one.
Synthetic polymers

Four synthetic polymers are examined for their effects on catechin and procyanidin fluorescence. The polymers were chosen by the nature of their side chains and backbones.

Poly vinyl pyrrolidone, PVP, is flexible with an amide group that is part of the pyrrolidone ring. It can be a hydrogen bond acceptor and has the added feature that three methylenes may contribute to a hydrophobic attraction. Figure 45a shows that PVP enhances the fluorescence of procyanidin B1 much more than the fluorescence of (-)-epicatechin. The solid filled circles in Figure 45a are results obtained by Bergmann (1986). Assuming similar experimental error in Bergmann's work, there is reasonable agreement.

Poly ethylene glycol, PEG, has no side chains, but its backbone contains oxygen, a hydrogen-acceptor. (-)-Epicatechin fluorescence is enhanced to the degree of \( \frac{Q}{Q_0} = 1.42 \) at 4 mg/ml. Procyanidin B1 fluorescence is enhanced significantly more. At 8 mg/ml PEG enhances fluorescence to \( \frac{Q}{Q_0} = 2.3 \). The curve does not level off in the range studied (Figure 45b).

Polyacrylamide is a flexible polymer with an amide side chain. It forms a gel at low concentrations. Measurements were made at concentrations below gel-formation. There is a slight enhancement of fluorescence of (-)-epicatechin and procyanidin B1 due to
Figure 45. Relative fluorescence of (+)-catechin (□) and procyanidin B-1 (+) in solutions of polymer as a function of polymer concentration. (a) poly vinyl pyrrolidone, 0 to 8 mg/ml. Filled circles are results of Bergmann (1986) (b) polyethylene glycol, 0 to 8 mg/ml (c) polyacrylamide, 0 to 2 mg/ml (d) hydroxypropyl cellulose, 0 to 2 mg/ml. Error bars are shown in the upper left corner of each plot.
Figure 45. (cont.).
polyacrylamide. As a function of concentration (Figure 45c) this effect levels at $Q/Q_0 = 1.32$ for concentrations greater than 1 mg/ml.

Hydroxypropyl cellulose is a stiff polymer with a number of hydrophobic regions as well as a capacity for hydrogen bonding. The highest concentration at which it permits a clear solution in water is 1 mg/ml. At these low concentrations hydroxypropyl cellulose enhances the fluorescence of (+)-catechin and procyanidin B1. It appears in Figure 45d that if solubility permitted an increase in hydroxypropyl cellulose concentration, no further increase in enhancement would occur.

It is interesting to note that for all four of these synthetic polymers the relative fluorescence; that is, the value of $Q/Q_0$ is greater than one. This is in agreement with effect of a hydrophobic solvent on the fluorescence of catechins. That is, all of the model polymers provide the tannin monomers and dimer with a slightly more hydrophobic environment. Clearly, the effect of poly vinyl pyrrolidone is greater than the effect of the other polymers. It is also apparent that the dimer responds more to the presence of polymer than does the monomer.

**Expected Behaviour of Polypeptides**

Considering these results, it is interesting to propose some expectations for the behavior of the polypeptides studied. Earlier work in this area suggests
that hydrogen bonding is necessary for tannin/protein interaction as is hydrophobic interaction. All of the synthetic polymers studied here are capable of hydrogen bond formation. If they collapse around the tannin monomer or dimer a more hydrophobic environment may be provided. Hydroxypropyl cellulose is not capable of such a collapse, but the hydroxypropyl side chains may be sufficiently hydrophobic to account for the degree of enhancement of fluorescence due to this polymer's presence.

Among the polypeptides studied, the poly glutamines fit into one category. This includes PHBG, PHPG, PH(E:B)G, and P(HBG:ARG). One may predict some interaction because the carbonyl in the backbone of polypeptides is a hydrogen bond acceptor. The interaction may depend on the accessibility of the backbone due to the conformation of the polymer as well as to flexibility; i.e., when the polypeptide is not in an α-helix. There may also be some dependence on the degree of hydrophobicity in the side chains.

Interaction of tannins with the polypeptides with a pH-induced transition depends not only on the accessibility of the backbone for hydrogen bonding, but also on the acidity of the side chain. If the side chains are acidic or positively charged, it is not likely that any interaction will occur. Flavan-3-ols are not charged, and contribute to hydrogen bonds by donating a hydrogen. It is not likely they would be attracted to positively charged
polypeptides. Hoon, et al. (1980) observed that cationic detergents were the least effective at dissociating tannin/protein complexes. If the side chains are basic or negatively charged they may attract the catechins, but may also lead to removal of hydroxyl hydrogens and the subsequent degradation of the catechin.

When poly (L-glutamic acid) is uncharged, it drops out of solution. When in a random coil with its backbone the most exposed, this polypeptide is negatively charged and is not expected to attract catechin, a hydrophobic molecule. When sufficiently uncharged to allow the catechins to approach, the backbone of poly (L-glutamic acid) is in an α-helix and not very exposed. One may, however, consider interaction with the carbonyl of the side chain.

Poly (L-arginine) and poly (L-lysine) have a different effect. At low pH, when these polymers are randomly coiled, the backbone is exposed but the side chains, because of their positive charge, inhibit interaction. When the side chains are uncharged, they are basic and may attract the hydrogen donating catechins. However, it is not possible to study catechins at high pH because of their degradation.

Poly [N-hydroxybutyl glutamine]

Circular Dichroism

The circular dichroism spectrum of poly [N-hydroxybutyl glutamine] indicates that this polymer
Figure 46. The temperature-induced transition of PHBG as measured by circular dichroism. The inset shows the [θ] at 222 nm as a function of temperature (□) and also with (-)-epicatechin concentrations of 0.01 mg/ml (†), 0.05 mg/ml (♦), and 0.1 mg/ml (△). Error bars on the middle line apply to all three spectra.
undergoes a temperature-induced helix-to-coil transition in water (Lotan 1966). Ellipticity of this polymer at three temperatures as a function of wavelength is shown in Figure 46. At low temperature a significant amount of the polymer is helical as indicated by the large positive peak at 190 nm and the negative peak at 222 nm. There is a trend toward a helix-to-coil transition as indicated by the decreasing intensity in both of these bands.

The four sets of points in the inset of Figure 46 describe the temperature-induced transition of PHBG and PHBG with (-)-epicatechin at three concentrations. The range of concentrations of (-)-epicatechin covers one order of magnitude without extending the limits described earlier; i.e., the signal of the procyanidins must not exceed 10% of the signal of the polymer at 190 nm and 222 nm. A solution of (-)-epicatechin alone at the corresponding concentration is the baseline for the circular dichroism of each of the three plots. Clearly, the (-)-epicatechin has no effect on the helix-to-coil transition of PHBG.

Fluorescence

The effect of PHBG on the fluorescence of (+)-catechin and procyanidin B1 is illustrated as a function of concentration in Figure 47. The polymer with a molecular weight of 400,000 shows no effect on fluorescence of procyanidins. However, a polymer of molecular weight 60,000 quenches strongly as a function of
Figure 47. The effect of PHBG on the fluorescence of (a) (+)-catechin and (b) procyanidin B-1. The solid line connects points for a polymer of PHBG with a molecular weight of 400,000 while the dotted line connects points for a polymer with molecular weight 60,000. Points shown are for 30°C (□, ◯) and 60°C (▲, △).
concentration. The attempt to determine exactly what the nature of this polymer is leads to the conclusion that this polymer has degraded to a poly glutamate and butanol amine. The circular dichroism of this sample exhibits a strong pH-induced helix-to-coil transition.

To test the postulate that the polymer is not what it seems, the effect of butanolamine on procyanidin fluorescence was determined. It was found that this molecule quenches catechin and procyanidin fluorescence strongly.

When observing the nature of quenching, it is the practice to make a plot of concentration of quencher against the degree of quenching plotted as $\frac{Q_0}{Q}$. Note that this is the reciprocal of other plots presented in this paper. If the Stern-Volmer plot is a straight line, the quenching is noted to be dynamic, and further study of the effect of the quencher on fluorescence lifetime is necessary to completely characterize the nature of the quenching. Dynamic quenching is an excited state event. If the Stern-Volmer plot is an arc with upward curvature, the quenching is said to be static. Static quenching is a ground state event, and occurs before the fluorophore is excited. When the Stern-Volmer plot is an arc with downward curvature, inaccessibility of the fluorophore is suggested.

The Stern-Volmer plot is shown in Figure 48. The points can be connected as a line or as an arc with
Figure 48. A Stern-Volmer plot for butanolamine as a quencher of (+)-catechin (■) and procyanidin B-1 (▲) fluorescence.
downward curvature. In this system, it is not likely that the fluorophore is inaccessible. The occurrence of dynamic quenching can be affirmed with lifetime measurements. Instrumentation for such a measurement was not available at the time these experiments were performed. It is clear, however, that butanolamine quenches catechin fluorescence. The exact nature of the quenching cannot be ascertained.

Quenching of (+)-catechin fluorescence by butanolamine is pH dependent. At a pH above 9.0, quenching is complete and almost immediate. These solutions turn yellow, a behavior also seen in solutions of (+)-catechin in water with a pH above 10.0. At a pH of 7.4 there is no apparent quenching due to butanolamine. However, raising the pH of the butanolamine solution to 8.0 prior to the addition of (+)-catechin results in the quenching of approximately half the fluorescence (\(Q_0/Q = 0.55\)). Subsequent addition of dilute acid to this solution decreases the quenching somewhat, but at a pH of 6.9, the value of \(Q_0/Q\) is still 0.71. There is still considerable quenching even though the pH is below the initial pH in which no quenching is apparent. The shape of the Stern-Volmer plot suggests an interaction between butanolamine and ground state (+)-catechin. Inability to regain the quenched fluorescence suggests the interaction may not be reversible.
Circular Dichroism

The temperature induced transition of PHPG is similar to that of PHBG, but occurs at a lower temperature. Figure 49 shows that at 10°C, the ellipticity at 222 nm is less negative (-8000) than it is for PHBG at 30°C. These results compare well with previous results (Lotan 1965; Lotan 1966; von Dreele 1971). The plot of the temperature-induced transition in the inset to Figure 49 indicates that both (+)-catechin and procyanidin B1 affect the degree of helicity in PHPG. The decrease in helicity at each temperature with the presence of these molecules suggests that the procyanidins may be shifting the transition to a lower temperature.

Fluorescence

It appears, however, in Figure 50 that PHPG has little effect on the fluorescence of catechin or procyanidin B1. Certainly there is no temperature effect. All values fall within experimental error of Q/Q₀ = 1, i.e., no effect.

Copoly [N-hydroxyethyl:hydroxybutyl glutamine]

Circular Dichroism

The circular dichroism of PH(E:B)G shows ellipticity of the same basic shape as for PHBG, but not as strongly helical (Figure 51). A plot of ellipticity versus temperature indicates that the effect of (+)-catechin and procyanidin B1 is to increase the degree of polymer in
Figure 49. The temperature-induced transition of PHPG as measured by circular dichroism. The inset shows [θ] at 222 nm as a function of temperature for PHPG (□), PHPG + (+)-catechin (+), and PHPG + procyanidin B-1 (◇).
Figure 50. The effect of polymer on the fluorescence of (+)-catechin and procyanidin B-1. PHBG [MW = 60000] (□); PH(E:B)G (♦); PHPG (◇); P(HBG:Arg) (△).
Figure 51. The temperature-induced transition of PH(E:B)G as measured by circular dichroism. Plots are for 5°C (□), 25°C (+), and 40°C (◇).
helical conformation at a given temperature (Figure 52). This indicates that the procyanidins shift the helix-to-coil transition of PH(E:B)G to a higher temperature.

**Fluorescence**

A plot of the relative fluorescence of (+)-catechin and procyanidin Bl due to the presence of PH(E:B)G (Figure 50) suggests that at higher temperatures where the conformation is more random; i.e., above 20°C, PH(E:B)G is more likely to quench fluorescence.

*Poly [(N-hydroxybutyl glutamine):arginine]*

**Circular Dichroism**

Another copolymer of PHBG is copoly [(N-hydroxybutyl glutamine):arginine]. The plot of ellipticity versus wavelength is very similar to that of PHBG (Figure 53). The degree of helix present at any given temperature is slightly less than for PHBG, an effect of the presence of arginine. The inset to Figure 53 compares the temperature-induced transition of these two polymers. The two lines appear to be converging. This suggests that the effect of arginine on ellipticity decreases as the amount of polymer in random coil increases. At a sufficiently high temperature, both polymers are random. The effect of (+)-catechin and procyanidin Bl on the ellipticity of P(HBG:Arg) at 222 nm as illustrated in Figure 54 is minimal.

**Fluorescence**

The effect of P(HBG:Arg) on the fluorescence of
Figure 52. The temperature-induced transition of PH(E:B)G (□) and the effect of (a) (+)-catechin and (b) procyanidin B-1 at proanthocyanidin concentrations of .05 mg/ml (+) and 0.1 mg/ml (△).
Figure 53. The temperature-induced transition of P(HBG:Arg). The cd spectra are shown for 30°C (■) and 60°C (♦). The inset is a plot of [θ] vs temperature for P(HBG:Arg) and the 60,000 PHBG.
Figure 54. The effect of (+)-catechin (a) and procyanidin B-1 (b) on the temperature-induced transition of P(HBG:Arg). Shown for the absence of (□) and three concentrations of (+)-catechin and procyanidin B-1; 0.01 mg/ml (+), 0.05 mg/ml (◇), and 0.1 mg/ml (▲). Error bars are shown in the lower left corner of each plot.
(+)-catechin and procyanidin B1 is within experimental error of the same value for the 60000 sample of PHBG (Figure 50). The quenching is independent of temperature. Concentration dependence was not examined due to a shortage of polymer.

**Poly (L-Arginine)**

It may be supposed that the quenching seen in the copolymer of arginine with hydroxybutyl glutamine is due to the arginine rather than to displaced butanolamine. Figure 55a is the result of the experiment to measure the effect of poly (L-arginine) at pH = 8.2 on the fluorescence of (+)-catechin and procyanidin B1. In the upper portion of the range of concentrations studied, some quenching began to occur. The solutions were allowed to stand for one week. After this time the experiment was repeated. The effect of poly (L-arginine) on the fluorescence of (+)-catechin and procyanidin B1 does not alter with time. It is not clear from the results obtained if the quenching is static or dynamic. A Stern-Volmer plot shows a slightly arced line which can be drawn as a straight line within experimental error.

**Poly (L-Lysine)**

The above experiment was also done on poly (L-lysine) at a pH of 7.8. Results shown in Figure 55b suggest less effect on fluorescence than is seen for poly (L-arginine). This may be due to the slightly lower pH or to the fact that this amino acid is not as basic as is
Figure 55. The effect of (a) poly (L-arginine) (pH = 8.2) and (b) poly (L-lysine) (pH = 7.8) on the fluorescence of (+)-catechin (■) and procyanidin B-1 (＋).
poly (L-arginine). At this pH, poly (L-lysine) has no effect on the fluorescence of the procyanidins.

**Poly glutamic acid (D and L)**

**Circular Dichroism**

Poly glutamic acid goes through a pH-induced helix-to-coil transition as pH is increased. The midpoint of this one step transition is near 5.7. The behavior of poly glutamic acid has been well studied (Myer 1969; Krimm 1974). The ellipticity at three values of pH is plotted as a function of wavelength in Figure 56a. The effect of (+)-catechin on the pH-induced transition is negligible as Figure 56b illustrates.

**Fluorescence**

The effect of poly (L-glutamic) acid and poly (D-glutamic acid) on the fluorescence of (+)-catechin and procyanidin B1 is measured on a series of solutions that include and go beyond (above and below) the transition pH. Results are plotted in Figure 57. Although there may be a sudden increase in the ability of poly glutamic acid to enhance fluorescence at the pH just above the transition, the jump is almost within the range allowed for experimental error. It is safe to say that both poly (L-glutamic acid) and poly (D-glutamic acid) slightly enhance the fluorescence of (+)-catechin and procyanidin.

**Poly (S-carboxymethyl cysteine)**

**Circular Dichroism**

Circular dichroism measurements were performed on
Figure 56. The pH-induced transition of poly (L-glutamic acid). (a) Ellipticity at 3 pH as a function of wavelength. (b) Ellipticity at 220 nm as a function of pH and the effect of (□) 0.01 mg/ml (+)-catechin, (+) 0.05 mg/ml (+)-catechin, and (Φ) 0.1 mg/ml (+)-catechin.
Figure 57. The effect of (a) poly (L-glutamic acid) and (b) poly (D-glutamic acid) on the fluorescence of (+)-catechin (□) and procyanidin B-1 (+) as poly (Glutamic acid) goes through a pH-induced transition.
two samples of poly (S-carboxymethyl cysteine), PCMS; one with a degree of polymerization equal to 330, the other equal to 560. Figure 58 is the cd spectrum for the smaller polymer at two different values of pH. These results agree well with results on the same samples published by the group that provided the sample (Maeda, et al. 1982), although the peaks obtained in this experiment are less intense than the published results of Maeda. The solid line in Figure 59a is of the pH-induced sheet to coil transition for the smaller polymer, a best fit line from the open squares. The dashed line is Maeda's result (Maeda, et al. 1984) on the same sample of PCMS. Maeda's results indicate a sharper transition, but pH of transition is the same. Also shown on this figure are four points that indicate the ellipticity of PCMS when (+)-catechin is in solution with the polymer. The points suggest a loss of cooperativity of the transition due to the presence of (+)-catechin.

Figure 59b illustrates the pH-induced transition for the polymer with a degree of polymerization equal to 560 as represented by the ellipticity at 220 nm and 200 nm. The five additional points on the plot are measured ellipticities of PCMS with (+)-catechin in solution. In all cases (+)-catechin was added before pH was adjusted. These ellipticities do not fall on the line. They may support a shift to lower pH of the transition.

(+)-Catechin does affect the transition. The nature of that effect cannot be conclusively found from
Figure 58. Circular dichroism spectra of poly (S-carboxymethyl cysteine) (DP = 330) at pH 4.3 (□) and pH 5.9 (+). Also shown are results obtained by Maeda, et al. (1982) (dashed line).
Figure 59. pH-induced transition of poly (S-carboxymethyl cysteine) as measured by circular dichroism. (a) DP = 330 (□), with (+)-catechin added (○). Dotted line indicates results of Maeda, et al. (1984). (b) DP = 560 at 200 nm (□) and at 220 nm (+); at 200 nm with (+)-catechin added after equilibration (○).
the evidence presented here. In the sample of PCMS with a lower degree of polymerization there appears to be a loss of cooperativity. In the polymer with a higher degree of polymerization, the pH of transition is shifted. In both cases there are not really enough points to draw a sound conclusion. These experiments were repeated until the samples sent from Japan were exhausted without obtaining more conclusive results.

Fluorescence

PCMS aggregates at very low concentrations. The effect of this polymer on the fluorescence of (+)-catechin was measured at a very low concentration of PCMS, much lower than observations on any other polymer. The (+)-catechin concentration, however, was similar to that used with other polymers. It is not surprising, therefore, that no effect on (+)-catechin fluorescence was observed.

Discussion

There is some effect of (+)-catechin and procyanidin B1 on the temperature-induced transitions of PHPG and PH(E:B)G. The β-sheet coil transition of PCMS is broadened by the presence of (+)-catechin. The effect of (+)-catechin on the transition as shown in Figure 59 is very similar to the effect of decreasing the edge effect penalty of r as seen in Figure 25. It is possible that (+)-catechin alters the sheet-coil transition by hydrogen bonding with the initial strand of a sheet, in effect decreasing the penalty ascribed to residues in that initial
strand; i.e., \( r \) increases. The tannin monomers and dimer do not affect the transition of any other polypeptide in this work at the concentrations studied.

PHBG, P(HBG:Arg), and PHPG do not significantly affect the fluorescence, and PH(E:B)G goes from a slightly enhancing effect to a slightly quenching effect around 20°C. In the last example, within experimental error, it could be said there is no effect of polymer on the fluorescence of procyanidins. The quenching seen in the plot for PHBG is due to the appearance of butanolamine which coincides with the degradation of the side chains in the polymer. In the case of P(HBG:ARG) quenching may be due to butanolamine or to arginine since poly (L-arginine) does quench at high concentrations. However, quenching by poly (L-arginine) is not really strong enough to account for quenching by the copolymer. It seems more likely that the effect seen in Figure 50 is due almost entirely to the presence of butanolamine in solution.

The quenching of (+)-catechin and procyanidin B1 fluorescence can be explained as being due to amine groups that are not positively charged. The lack of fluorescence enhancement by the polypeptides suggests either that there is no interaction or, possibly, that interaction cannot be studied in dilute solution. Some interaction is apparent in the circular dichroism of PHPG and PH(E:B)G where the concentration ratio of (+)-catechin and procyanidin B1 to polymer is much higher. It seems possible that at higher
concentrations of tannin monomer and dimer, some effect on the fluorescence could be observed.

The results of this study can be considered in relationship to previous work. Bergmann (1986) notes that poly (L-proline) quenches fluorescence of (+)-catechin. There are two things about poly (L-proline) that make it unique from other polypeptides. One is the presence of a five member ring. The other is the effect of that ring on the conformation of a chain, i.e., a restriction of rotation about the N - Cα bond. It has been noted that tannins have a high affinity for proteins with high levels of poly (L-proline) (Asquith 1986), an observation that supports the importance of any effect of the sidechain. There is also the fact that PVP has a unique and strong interaction with catechins. Although the effect on fluorescence is quite different than the effect of poly (L-proline) (enhancement versus quenching), the fact that there is an effect suggests that both are capable of attracting catechins to their vicinities. What happens after that is unique to each case.

Bergmann (1986) has reported that the interaction of poly (L-proline) with (+)-catechin depends on the conformation of poly (L-proline). Poly (L-proline is capable of forming two helices of different handedness: Form I being a right-handed helix and Form II a left-handed helix. In a solvent system shown by cd to support Form I, the fluorescence of (+)-catechin is quenched much more
strongly than in a solvent system supportive of Form II. This suggests that the backbone structure of proline residues in proteins must be in a conformation similar to that for Form I in poly (L-proline).

In the study of tannin/protein interactions, the strong interaction of catechins with proline remains unique. It is interesting to note that one can sense the presence of tannins in such things as tea, wine, and cranberry juice by their characteristic tartness. They tend to leave one's tongue dry. This is because the salivary proteins are precipitated by catechins and their effectiveness as lubricants is impaired. Proteins in the juices secreted by salivary glands are remarkably high in proline. If, as has been proposed (Bate-Smith 1954), the evolutionary purpose of tannins is to prevent predation of plants, precipitation of salivary proteins is an effective and remarkably specific means.

Still unresolved is the nature of interaction with other amino acids. Tannins are capable of precipitating proteins with lower levels of proline (to a much lower degree than those high in proteins) (Asquith 1986), but only when catechin and protein concentrations are much higher than the concentrations used in this study. Since the concentration of fluorophores is limited by the simple law that the absorbance at the exciting wavelength must be less than but close to 0.1, increasing tannin concentration would require changing to front-face
illumination. In this technique light does not pass through the solution, but hits the surface instead. A high concentration of fluorophore will not quench its own fluorescence if that fluorescence never passes through the solution. Any further studies on tannin/protein interaction using fluorescence should take advantage of this technique.
APPENDIX I
SYMBOLS USED

Polypeptide Chain

\( \phi \) -- Rotation about the N-C\( ^\alpha \) bond in a polypeptide chain. The dihedral angle is 0° when C\( ^\alpha \)-C is trans to N-H.

\( \psi \) -- Rotation about the C\( ^\alpha \)-C bond in a polypeptide chain. The dihedral angle is 0° when C\( ^\alpha \)-C is trans to C-O.

\( \Omega \) -- All the dihedral angles for all the residues of a polypeptide chain.

P(\( \Omega \)) -- Probability of a given set of dihedral angles for all the residues of a polypeptide chain.

sw(\( \Omega \)) -- Statistical weight for a given set of dihedral angles for all the residues of a polypeptide chain. P(\( \Omega \)) = sw(\( \Omega \))/Z

Transformation Matrix

\( \theta \) -- Supplement of the angle between consecutive virtual bonds; used in the transformation matrix.

\( \phi \) -- Dihedral angle for a bond. The trans conformation is 0°, and \( \phi \) increases with right hand motion. Used in the transformation matrix.
Statistical Weights

$\delta$ — Statistical weight assigned to each bend in a sheet.

$f_1$ — Statistical weight assigned to residues in an interstrand loop.

$\sigma$ — Modifies the statistical weight assigned to a helix to account for the penalty of initiation in the helix coil transition.

$s$ — Statistical weight assigned to a residue in an $\alpha$ helix in the helix coil transition.

$r$ — Statistical weight assigned to residues in a $\beta$ sheet with no corresponding residues in a previous sheet.

$t$ — Statistical weight assigned to each residue in a sheet.

Other

$[\theta]$ — Ellipticity

$b_i$ — The $i$th residue in a strand with no corresponding residues in the preceding strand.

$B_{ij}$ — The $i$th residue in a strand with a corresponding residue in the previous strand. The previous strand is $j$ residues long. $i \leq j$.

$c$ — A residue in a random coil region.

$cd$ — Circular dichroism

$h$ — A residue in a helical region.

$I$ — Maximum number of residues allowed in a strand.

$J$ — Dimension of the statistical weight matrix.
\[ J = I^*(I + 3)/2. \]

- \( n \) -- number of residues in a polypeptide chain.
- \( N \) -- number of virtual bonds in a polypeptide chain.
- \( Q \) -- quantum yield
- \( r \) -- end to end distance of a polymer.
- \( r^2 \) -- square end to end distance of a polymer.
- \( \langle r^2 \rangle \) -- mean square end to end distance of a polymer.
- \( Z \) -- conformation partition function.
APPENDIX II

ABBREVIATIONS

Polymers

PHBG — poly [N-hydroxybutyl glutamine]
PHPG — poly [N-hydroxypropyl glutamine]
PH(E:B)G — copoly [N-hydroxy(ethyl:butyl) glutamine]
P(HBG:ARG) — copoly [(N-hydroxybutyl):(arginine)]
p-L-arg — poly (L-arginine)
p-L-lys — poly (L-lysine)
p-L-glu — poly (L-glutamic acid)
p-D-glu — poly (D-glutamic acid)
PCMS — poly (S-carboxymethyl cysteine)
PVP — poly vinyl pyrrolidone
PEG — poly ethylene glycol
PA — polyacrylamide
HPC — hydroxypropyl cellulose
APPENDIX III
MATRIX DEFINITIONS

F — This matrix includes the statistical weight matrix, U, and the generator matrix, G.

Fb1 — Generator matrix for the bond of a bend.

Fb2 — Generator matrix for the first bond in a new strand after a bend.

Fs — Generator matrix for a bond in a strand of a sheet.

G — The generator matrix; unique for every conformation of a residue.

Gc — Generator matrix for a bond in a random coil.

ISL — Generator matrix for a bond in an interstrand loop.

li — Virtual bond i; the bond that connects α carbon of residue i - 1 to the α carbon of residue i.

li^T — the transform of li.

r — end to end vector.

Ti — Transformation matrix that transforms the coordinate system of bond i + 1 into the coordinate system of bond i.

Tbl — Transformation matrix for the first corner of a bend.
$T_b^2$ --Transformation matrix for the completion of a bend.

$T_s$ --Transformation matrix for a strand.

$U$ --The statistical weight matrix which accounts for nearest neighbor interaction.
APPENDIX IV

STATISTICAL WEIGHT MATRIX FORMULAE

Following is a listing of the mathematical formulae that define the nonzero elements in the statistical weight matrix for a β sheet coil transition. The 'C' programming language is one in which the initial element of an array is element 0. Therefore, in these definitions the initial element of each row and column is the zeroth elements. For an element in the ith row and jth column, the position is named $u_{ij}$

1) Column 0

$$u_{i0} = 1; \quad i = 0$$

$$2 \leq i \leq I + 1$$

$$2I \leq i \leq I(I + 3)/2$$

2) Columns in which 'rt' appears; Column j; $u_{ij} = rt$:

   a) For $1 \leq j < I + 1$, $i = j - 1$.

   b) For $3 \leq j < I + 1$, $i = (j - 1)I - (j - 2)(j - 3)/2$

3) Columns in which '6t' appears; Column j; $u_{ij} = \delta t$:

   a) For $I + 1 \leq j < I + I$, $i = j + 1 - I$.

   b) Let $d = j + 1$ then $f(j) = ((d - i)[3(I + 1) - d]/2) - 1$

      for $b = 0, 1, 2, \ldots, 2I - j$

      $i = f(j) + m$.

4) Columns in which 't' appears; Column j; $u_{ij} = t$
Define a function

\[ f(g) = (g - I)[3(I + 1) - g]/2 \]

\[ g = I + 2, I + 3, \ldots, I + I \]

For \( f(g) - 1 \leq j < [f(g) + 2I - g], \) \( i = j - 2I + g - 1. \)
APPENDIX FIVE

A PROGRAM TO CALCULATE MEAN SQUARE END-TO-END DISTANCE

Following is a listing of the program used to do the calculations in this study. The program is written in the 'C' programming language (Kernigham and Ritchie 1979). Comments enclosed in /* */ are explanatory and not part of the program.
This program calculates the mean square end to end distance of a homopolymer. It makes use of the generator matrix theory of Paul J. Flory and serial matrix multiplication methods for the determination of the conformation partition function as developed for the sheet coil transition by Wayne L. Mattice. Generator matrices defined in the program are those for poly-(L-alanine). Statistical weights are read from an external file named "par_file". Since it is necessary to declare the size of arrays according to the value of I, I is defined in the program. Changes in the value of I require recompilation.

#include <stdio.h>
#include <math.h>
define I 15

/* The generator matrices */

/* interstrand loop */

float ISL[5][5]=
{\(1.0,0.0,0.0,0.0,0.0\),
 {0.0,1.0,0.0,0.0,0.0},
 {0.0,0.0,1.0,0.0,0.0},
 {0.0,0.0,0.0,1.0,0.0},
 {0.0,0.0,0.0,0.0,1.0}};

/* for p(L-ala) random coil */

float Gc[5][5]=
{\(1.0,3.88,1.15,4.48,14.44\),
 {0.0,0.51,0.2,0.59,3.8},
 {0.0,-0.046,-0.61,0.21,0.0},
 {0.0,0.65,-0.23,-0.3,0.0},
 {0.0,0.0,0.0,0.0,1.0}};

/* first corner of a bend */

float Fb1[5][5]=
{\(1.0,0.0,9.4,0.0,22.09\),
 {0.0,0.0,1.0,0.0,4.7},
 {0.0,1.0,0.0,0.0,0.0},
 {0.0,0.0,0.0,-1.0,0.0},
 {0.0,0.0,0.0,0.0,1.0}};

/* second corner of a bend */

float Fb2[5][5]=
{\(1.0,0.0,6.89,0.0,11.868\),
 {0.0,0.0,1.0,0.0,3.445},
 {0.0,-1.0,0.0,0.0,0.0},
 {0.0,0.0,0.0,1.0,0.0},
 {0.0,0.0,0.0,0.0,1.0}};

/* in the sheet */

float Fs[5][5]=
{\(1.0,6.89,0.0,0.0,11.868\),
 {0.0,1.0,0.0,0.0,3.445},
 {0.0,0.0,1.0,0.0,0.0},
 {0.0,0.0,0.0,1.0,0.0},
 {0.0,0.0,0.0,0.0,1.0}};
/* External declaration of the generator matrices */
/* multiplied by the appropriate statistical weights */

float Gcu[5][5], ttFs[5][5], dtFb1[5][5], tFb2[5][5];
float tFs[5][5], fISL[5][5], ttGcu[5][5];

/* External declaration of the row matrix for the */
/* F matrix and for the determination of the */
/* conformation partition function, Z. */

float row[5*(I*(I + 3)/2)] = {1.0,0.0,0.0,0.0,0.0};
float zrow[I*(I + 3)/2] = {1.0,0.0,0.0};

/* Some external values required for passing values */
/* between functions. */
float zt = 0.0;
float sz = 0.0;
float val, s[5];

/* The main program */
main()
{
    float TAU, DELTA, T, WTRAN, F, inc;
    float nrow[5*I*(I + 3)/2];
    float rsqr, Zrsqr, Z;
    int a, b, c, j, r, N, J, loop, tloop;
    long int column, loc;

    extern float ISL[5][5], Gc[5][5], Fb1[5][5], Fb2[5][5];
    extern float Fs[5][5], Gcu[5][5];
    extern float tFb1[5][5], dtFb1[5][5], tFb2[5][5];
    extern float fISL[5][5], ttGcu[5][5], tFs[5][5];
    extern float row[5*(I+3)/2], zrow[I*(I + 3)/2];
    extern float val, sz;

    FILE *pars;
    FILE *frp;
    FILE *res;

    /* The following lines of text read statistical weights */
    /* and the chain length from an external file, then */
    /* write the numbers to the file which will later */
    /* receive the results. */

    pars = fopen("par_file", "r");
    fscanf(pars, "\d\t", &N);
    fscanf(pars, "\d\t", &TAU);
    fscanf(pars, "\d\t", &DELTA);
    fscanf(pars, "\d\t", &T);
    fscanf(pars, "\d\t", &inc);
}
fscanf(pars, "%f\t", &WTRAN); /* A weighting factor */
/* multiplied onto all */
/* statistical weights */
/* to keep the value of */
/* Z within the */
/* computer's limits */

fscanf(pars, "%f\t", &F); /* Interstrand loops */

J = I*(I + 3)/2;
fclose(pars);
res = fopen("results", "w");
fprintf(res, "I is %d\n", I);
fprintf(res, "N is %d\n", N);
fprintf(res, "TAU is %f\n", TAU);
fprintf(res, "T is %f\n", T);

fprintf(res, "DELTA is %f\n", DELTA);
fprintf(res, "WTRAN is %f\n", WTRAN);
fprintf(res, "F is %f\n", F);

/* The function "conxma(a, B, C)" multiplies */
/* every element in matrix B by a to result */
/* in matrix C */

conxma(WTRAN, &Gc[0][0], &Gcu[0][0]);
conxma(WTRAN*TAU*T, &Gc[0][0], &tGcu[0][0]);
conxma(WTRAN*TAU*T, &Fs[0][0], &ttFs[0][0]);
conxma(WTRAN*DELTA*T, &Fb1[0][0], &tFb1[0][0]);
conxma(WTRAN*T, &Fb2[0][0], &tFb2[0][0]);
conxma(WTRAN*F, &ISL[0][0], &tISL[0][0]);

/* In the following nested loops, the indeces are: */
/* b -- keeps track of residue number */
/* column -- to create one element of a new */
/* row, the entire old row is */
/* multiplied onto the column of */
/* the square matrix that is the */
/* column for the row element that */
/* is to be determined. This index */
/* is the column position in the */
/* statistical weight matrix, U. */
/* a -- the value of 'column' signifies a */
/* position in the statistical weight */
/* matrix. In the F matrix, each */
/* statistical weight is multiplied onto */
/* a 5 X 5 generator matrix. The value */
/* of 'a' indicates the column of the */
/* generator matrix. The column number */
/* in the F matrix is 'column' + 'a' */
for (b = 1; b < N; b++){
    for (column = 0; column < J; column++){
        a = 0;
        loc = 5*(column);
        for (a = 0; a < 5; a++){
            val = 0.0;
            matmul((column), a, &row[0]);
            /* This function does the multiplication */
            /* and summation necessary for the */
            /* multiplication of a row onto a matrix.*/
            /* The column number of the F matrix is */
            /* 'loc + a'.
            nrow[loc + a] = val;
        }
        /* This loop yields five row elements */
        /* for every one value of 'column' */
    }
}

frp = fopen("rows", "w");

fprintf(frp, "N is %d\n", N);
fprintf(frp, "For b = 1f the new row is\n " , b);

for (c = 0; c < 5*J; c++){
    row[c] = nrow[c];
    /* reassigns row with its new values */
    fprintf(frp, "%e\n", row[c]);
}

fclose(frp);

/* The last time the row is multiplied onto the F */
/* matrix, the virtual bond lengths must be set to zero */
/* to account for the fact that the generator matrices */
/* count number of bonds while the statistical weight */
/* matrices count number of alpha carbons. To do this, */
/* all of the generator matrices are rebuilt with the */
/* elements that are '1' set equal to zero. The */
/* 'matmul' function is called one last time with the */
/* modified generator matrices.
*/

lastmat(WTRAN, &Gc[0][0], &Gcu[0][0]);
lastmat(WTRAN*TAU*T, &Gc[0][0], &ttGcu[0][0]);
lastmat(WTRAN*TAU*T, &Fs[0][0], &ttFs[0][0]);
lastmat(WTRAN*DELTA*T, &Fb1[0][0], &ttFb1[0][0]);
lastmat(WTRAN*T, &Fb2[0][0], &ttFb2[0][0]);
lastmat(WTRAN*T, &Fs[0][0], &ttFs[0][0]);
lastmat(WTRAN*F, &ISL[0][0], &ttISL[0][0]);
for (column = 0; column < J; column++) {
    a = 0;
    loc = 5*(column);
    for (a = 0; a < 5; a++) {
        val = 0.0;
        matmul((column), a, &row[0]);
        nrow[loc + a] = val;
    }
}
for (c = 0; c < 5*J; c++) {
    row[c] = nrow[c];
}
/* The final column is multiplied onto the final row by */
/* summing the row elements that correspond to the */
/* nonzero elements in the final column. */
Zrsqr = row[4];
for (j = 2; j <= I; j++) {
    Zrsqr = Zrsqr + row[5*j + 4];
}
for (j = 2*I; j <= J - 1; j++) {
    Zrsqr = Zrsqr + row[5*j + 4];
}
cpfunc(N, &T, &IAU, &DELT, &WTRAN, &F); /* Calling this*/
/* function */
/* results */
/* in the */
/* calculation */
/* of Z, here */
/* defined as sz*/
rsqr = Zrsqr/sz;
fprintf(res, "For T = %f:\n", T);
fprintf(res, "WTRAN = %f:\n", WTRAN);
fprintf(res, "The value of Z is %e:\n", rsqr);
fprintf(res, "The mean square radius is %f:\n", rsqr);
fprintf(res, "\n"); /* Here I send results */
/* to a file named 'results' */
}
matmul(cc, c, prow) /* This is the function that */
int cc, c; /* multiplies the ith elements */
float *prow; /* of column [5*cc]+c] by the */
/* ith element of the old row to*/
/* determine the [(5*cc) c] */
/* element of the new row. */
extern float Gcu[5][5], fISL[5][5], ttGcu[5][5];
extern float ttFs[5][5], dtFbl[5][5], tFb2[5][5];
extern float tFs[5][5];

extern float s[5], val;
int g, i, d, a, m, cdelta, Fc, Fg;

if (cc — 0){ /* This condition describes */
    /* the first row of */
    for(d = 0; d < 5; d++){ /* the statistical */
        s[a] = *(prow + d); /* weight matrix and*/
        Gcmul(c); /* the first five */
        /* rows of the F matrix. */
        /* The value of c tells the */
        /* function 'Gcmul' which */
        /* column of the 5 x 5 */
        /* matrix to multiply onto */
        /* the row, s[]. This row */
        /* is five elements of the */
        /* old row read to a new */
        /* location. */
        for(i = 2; i <= I; i++){
            a = 0;
            for(d = 5*i; d < 5*i + 5; d++){
                s[a] = *(prow + d);
                a++;
            }
            Gcmul(c);
        }
    }
    for (i = 2*I; i < I*(I + 3)/2; i++){
        a = 0;
        for(d = 5*i; d < 5*i + 5; d++){
            s[a] = *(prow + d);
            a++;
        }
        Gcmul(c);
    }
    return;
}
if(cc — 1){
    i = 0;
    a = 0;
    for(d = 5*i; d < 5*i + 5; d++){
        s[a] = *(prow + d);
        a++;
    }
    ttGcmul(c);
}
if (cc > 1 && cc < I + 1) /* TAU*T times a matrix Fs */
    i = cc - 1;
    a = 0;
    for (d = 5*i; d < 5*i + 5; d++) {
        s[a] = *(prow + d);
        a++;
    }
    ttFsmul(c);
}

if (cc >= 3 && cc <= I) {
    i = (cc - 1)*I - (cc - 2)*(cc - 3)/2;
    a = 0;
    for (d = 5*i; d < 5*i + 5; d++) {
        s[a] = *(prow + d);
        a++;
    }
    ttFsmul(c);
}

if (cc >= I + 1 && cc < I + I) /* DELTA*T times Fbl */
    i = cc + I - I;
    a = 0;
    for (d = 5*i; d < 5*i + 5; d++) {
        s[a] = *(prow + d);
        a++;
    }
    dtFblmul(c);
    cdelta = cc + 1;
    Fc = (cdelta - I)*3*(I + 1) - cdelta)/2 - 1;
    for (m = 0; m < (2*I - cc); m++) {
        i = Fc + m;
        a = 0;
        for (d = 5*i; d < 5*i + 5; d++) {
            s[a] = *(prow + d);
            a++;
        }
    }
    dtFblmul(c);
if (cc >= I + 1 && cc < 2*I){ /* F times ISL */
i = cc;
a = 0;
for (d = 5*i; d < 5*i + 5; d++){
s[a] = row[d];
a++;
}
fISLmul(c);
}

if (cc >= 2*I && cc < 3*I - 1){ /* T times matrix Fb2 */
i = cc + 1 - I;
a = 0;
for (d = 5*i; d < 5*i + 5; d++){
s[a] = row[d];
a++;
}
tFb2mul(c);
}

for(g = I + 3; g <= (2*I); g++){ /* T times matrix Fs */
    Fg = (g - I)*(3*(I + 1) - g)/2;
    if(cc >= (Fg - 1) && cc < (Fg + 2*I - g)){
        i = cc - 2*I + g - 1;
a = 0;
        for(d = 5*i; d < 5*i + 5; d++){
s[a] = row[d];
a++;
    }
tFsmul(c);
}
}
return;

conxma(const, pmat, pcm) /* This function */
float const; /* multiplies statistical */
float *pmat, *pcm; /* weight times */
{ /* geometry matrices. */
    int a, b;
    float hold;
    for(a = 0; a < 5; a++){
        for(b = 0; b < 5; b++){
            hold = const*(*pmat);
            *pcm = hold;
            pmat++;
            pcm++;
        }
    }
}
return;
}

lastmat(const, pmat, pcm)  /* This function redefines */
float const;                /* the values in each of */
float *pmat, *pcm;          /* the 'C' matrices so that*/
{                          /* the length of the last */
  int a, b;                /* bond is zero. */
  float hold;

  *pcm = const*(pmat);      
    pcm++;  pmat++;        
  for(a = 1; a < 5; a++)(
    *pcm = 0.0;             
    pcm++;  pmat++;        
  }
  for(b = 0; b < 3; b++)(
    for(a = 1; a < 5; a++)(
      *pcm = const*(pmat);  
        pcm++;  pmat++;    
    )
    *pcm = 0.0;             
      pcm++;  pmat++;      
  }
  for(a = 0; a < 5; a++)(
    *pcm = const*(pmat);  
      pcm++;  pmat++;    
  )
}

return;

}

Gcmul(gc)  /* The next series of functions */
int gc;    /* multiplies the row, 5 elements*/
{          /* at a time, by the matrices */

  extern float s[5], val, Gcu[5][5];

  int a, b, j;

  float k;

  for(j = 0; j < 5; j++){
    k = s[j] * Gcu[j][gc];
    val = val + k;
  }

  return;
}

ttGcumul(gc)  int gc;
extern float s[5], val, ttGcu[5][5];
int a, b, j;
float k;

for(j = 0; j < 5; j++)
    k = s[j] * ttGcu[j][gc];
    val = val + k;
return;
}

ttFsmul(gc)

int gc;
{

extern float s[5], val, ttFs[5][5];
int j;
float k;

for(j = 0; j < 5; j++)
    k = s[j] * ttFs[j][gc];
    val = val + k;
}
return;

}

tFsmul(gc)

int gc;
{

extern float s[5], val, tFs[5][5];
int j;
float k;

for(j = 0; j < 5; j++)
    k = s[j] * tFs[j][gc];
    val = val + k;
}
return;

}

dtFblmul(gc)

int gc;
{

}
extern float s[5], val, dtFb1[5][5];
int j;
float k;

for (j = 0; j < 5; j++){
    k = s[j] * dtFb1[j][gc];
    val = val + k;
}
return;
}

fISLmul(gc)
int gc;
{
extern float s[5], val, fISL[5][5];
int j;
float k;

for (j = 0; j < 5; j++){
    k = s[j] * fISL[j][gc];
    val = val + k;
}
return;
}

tFb2mul(gc)
int gc;
{
extern float s[5], val, tFb2[5][5];
int j;
float k;

for(j = 0; j < 5; j++){
    k = s[j] * tFb2[j][gc];
    val = val + k;
}
return;
}

/* This is the central program from zcalc.c */
/* Its function is to calculate the value of Z */


```
cpfunc(NN, pt, ptau, pdelta, pwtran, pf)
int NN;
{
    extern float zrow[];
    extern float zt;
    int a, b, zcol;
    float t, tau, delta, wtran, f;
    float newrow[I*(I + 3)/2];
    int J = I*(I + 3)/2;
    
t = *pt;
tau = *ptau;
delta = *pdf;
wtran = *pwtran;
f = *pf;
a = b = 0;
    /* NN is the number of square matrices */
    /* This loop creates a new row */
    for (b = 0; b < NN; b++){
        zcol = 0;
        while (zcol < J){
            zt = 0.0;
            zmamul(zcol, &t, &tau, &delta, &wtran, &f);
            /* zmamul returns one value of */
            /* of a new row. */
            newrow[zcol] = zt;
            zcol = zcol + 1;
        }
        for (a = 0; a < J; a++){
            zrow[a] = newrow[a];
        }
        a = 0;
    }
    zscalar();  /* returns a scalar value of Z */
    return;
}

zmamul(zc, ptt, pttau, ptdelta, pwtran, ptf)
    /* In this program, zrow[i] is */
    int zc;
    /* multiplied by defined nonzero*/
    /* elements of the statistical */
    /* weight matrix one column at */
    /* a time. */
    
    
    float T = *ptt;
    float TAU = *pttau;
```
float DELTA = *ptdelta;
float WTRAN = *ptwtran;
float F = *ptf;
extern float zrow[], zt;
int i, zg, m, Fzc, zFzg, zdelta;
float zk = 0.0;

/* In the following sequences, rows with nonzero*/
/* elements are defined. */
if (zc == 0) {
    /* The value of */
    /* all nonzero */
    zt = zrow[0] * WTRAN;
    /* elements in */
    for (i = 2; i < I + 1; i++) {
        /* column one is */
        zk = zrow[i] * WTRAN;
        zt = zt + zk;
    }
    /* "1". */
    for (i = 2*I; i < I*(I + 3)/2; i++) {
        zk = zrow[i] * WTRAN;
        zt = zt + zk;
    }
    return;
}

/* Columns defined by the value of zc */
/* the following two loops contain the */
/* value TAU*T. */
if (zc >- 1 && zc < I + 1) {
    i = zc - 1;
    zk = zrow[i] * WTRAN * TAU * T;
    zt = zt + zk;
}
if (zc >- 3 && zc < I + 1) {
    i = (zc - 1) * I - (zc - 2) * (zc - 3)/2;
    zk = zrow[i] * WTRAN * TAU * T;
    zt = zt + zk;
}
/* elements with the value F */
if (zc >- I + 1 && zc < 2*I) {
    i = zc;
    zk = zrow[i] * WTRAN * F;
    zt = zt + zk;
}
/* elements with the value DELTA*T */
if (zc >- I + 1 && zc < I + I) {
    i = zc + 1 - I;
    zk = zrow[i] * WTRAN * DELTA * T;
    zt = zt + zk;
}
zdelt = zc + 1;
Fzc = ((zdelt - I) * (3*(I + 1) - zdelta)/2) - 1;
for (m = 0; m < 2*I - zc; m++)(
    i = Fzc + m;
    zk = zrow[i] * WTRAN*DELTA*T;
    zt = zt + zk;
}

for (zg = I + 2; zg <= I + I; zg++)(
    zFzg = (zg - I)*(3*(I + 1) - zg)/2;
    if (zc >= (zFzg - 1) && zc < (zFzg + 2*I - zg))(
        i = zc - (2*I) + zg - 1;
        zk = zrow[i] * WTRAN*T;
        zt = zt + zk;
    }
)
return;
}

zscalar() /* Function to multiply the final row by */
    /* the appropriate column matrix to give */
    /* the scalar value of Z */
{
    extern float zrow[], sz;
    int i;
    sz = zrow[0];
    for (i = 2; i <= I; i++){
        sz = sz + zrow[i];
    }
    for (i = 2*I; i < I*(I + 3)/2; i++){
        sz = sz + zrow[i];
    }
    return;
}
APPENDIX VI

MAGIC ANGLE CONDITIONS

The total intensity of light emitted by a fluorophore has components polarized in three directions. If the exciting light is polarized along the z axis, the total intensity of emitted light is $S$,

$$ S = I_z + I_x + I_y \quad \text{(VI-1)} $$

$$ I_z = I \quad \text{(VI-2)} $$

$$ I_x = I_y = I_\perp \quad \text{(VI-3)} $$

$$ S = I + 2I_\perp \quad \text{(VI-4)} $$

However, the detector cannot sense light polarized in the line perpendicular in the direction coming directly toward it. Therefore, if $I_{\text{obs}}$ is detected light,

$$ I_{\text{obs}} = I + I_\perp \quad \text{(VI-5)} $$

(flavrophore)
The intensity of light the detector can sense with an emission polarizer present is defined by

\[ I_\alpha = I_\parallel \cos^2 \alpha + I_\perp \sin^2 \alpha \]  

(VI-6)

The definition of anisotropy, \( r \), is

\[ r = (I_\parallel - I_\perp)/S \]  

(VI-7)

then substituting (VI-4) into (VI-7)

\[ I = [S(1 + 2r)]/3 \]  

(VI-8)

\[ I_\perp = [S(1 - r)]/3 \]  

(VI-9)

Putting these two equations into (V-6) results in

\[ I_\alpha = (1/3)S(1 + 2r)\cos^2 \alpha + (1/3)S(1 - r)\sin^2 \alpha \]  

(VI-10)

\[ = (1/3)S(\cos^2 \alpha + \sin^2 \alpha) + (1/3)Sr(2\cos^2 \alpha - \sin^2 \alpha) \]

\[ = (1/3)S + (1/3)Sr(2\cos^2 \alpha - \sin^2 \alpha) \]

If

\[ 2\cos^2 \alpha - \sin^2 \alpha = 0 \]

then

\[ I_\alpha = (1/3)S \]

that is, \( I \) is directly proportional to \( S \).

\[ 2\cos^2 \alpha = \sin^2 \alpha \]

when

\[ \alpha = 54.7^\circ \]
REFERENCES


Bate-Smith, E. C. Food 1954, 23, 124.

Bergmann, W. R., Barkley, M., Mattice, W. Polymer Preprints 1986, 27, 320.


Flory, P. J. Macromolecules 1974, 7, 381.


Gray, J. C. Phytochemistry 1978, 17, 495.


Ikeda, S., Fukutoma, A., Imae, T., Yoshida, T. Biopolymers 1979, 18, 335.


Lader, H. J., Mandelkern, L. *Biopolymers* 1979, 18, 2607.
Mattice, W. L., Scheraga, H. A. *Biopolymers* 1984a, 23, 1701.


Thompson, R. S., Jacques, D., Haslam, D., Tanner, R. J. N. *JCS Perkin I* 1972, 1387.

Tiffany, M. L., Krimm, S. *Biopolymers* 1968, 6, 1379.


vanSumere, C., Albrecht, J., Dedonder, A., DePooter, H., Pe, I. in "The Chemistry and Biochemistry of Plant


VITA

EDUCATION: Louisiana State University
            Baton Rouge, Louisiana

            Degree: Ph.D., August 1987
            Major: Physical Chemistry
            Minor: Biochemistry

Central College
            Pella, Iowa

            Degree: B.A., May 1983
            Major: Chemistry

HONORS: Alumni Federation Fellow, LSU
        Member of Phi Lambda Epsilon, Alpha Mu Chapter
        Rollscreen Scholar, Central College
        Member of Alpha Zeta Mu, Central College

PUBLICATIONS: L. Tilstra, T. C. Strickland, D. S. Clark,
               W. L. Mattice, Colloid & Polymer Science,

               W. L. Mattice, L. Tilstra, Biopolymers,

MEETINGS: ACS National Meeting, New Orleans, Louisiana;
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Luanne F. Tilstra

Major Field: Chemistry

Title of Dissertation: Polypeptides: Conformational Transition and Complex Formation with Catechins and Procyanidins

Approved:

[Signatures]

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination: July 14, 1987