THE EFFECT OF SIDE CHAIN COMPLEXITY ON THE
DIMENSIONS OF POLYPEPTIDE CHAINS

A thesis presented for the degree of
DOCTOR OF PHILOSOPHY

by

YVONNE PATERSON

The Russell Grimwade School of Biochemistry
University of Melbourne

1978
To the memory of my father, Leon Du Bois;
my opinions evolved from his principles and my personality
was moulded by his love;

and to John Paterson who gave me a model of
courage and self-confidence to emulate.
ADDENDA TO THESIS

"THE EFFECT OF SIDE CHAIN COMPLEXITY ON
THE DIMENSIONS OF POLYPEPTIDE CHAINS"

1. It is difficult to find a reasonable physical explanation for the use of certain parameters employed in empirical energy functions (pp.50-54). As discussed in the review of conformational energy calculations (pp.15-18) the calculation of conformations of minimum energy with the aid of potential energy functions based on empirical observations of the energetic behaviour of small molecules is expedient but far from rigorous. Obviously the calculations are only as good as the functions and parameters used and this is why it is necessary to continually update the functions as new data become available. Some terms in the partitioned energy function are there simply to balance the energy equation and reproduce the energetic behaviour of model compounds. The torsional energy term (included in some form by all workers in this field) could be considered to be an example of making the results fit the facts as discussed on p.54. However, it must be emphasised that empirical energy calculations are truly empirical and must be made to reflect the actual behaviour of molecules even if we are not always certain of the physical reasons for it.

Other workers use different parameters and potential functions but most of these have not been revised in the past few years. The Cornell group's methods have been used here because they have been refined recently. Experimental data on the rotational barriers around the N-Cα and Cα-C' bonds published since Flory's pioneering work in this field (cited on p.61) suggest that these torsional barriers are small or non-existent. For this reason, and in keeping with the tenet that the functions used should be based on empirical evidence, no torsional energy terms were included for φ or ψ, or for χ3 and χ5 of the hydroxyethylglutamine residue.

Another point worthy of emphasis is that the value of the conformational energy calculated may vary, depending on the potential functions used. It is not valid to infer conformational stability of one point on the (φ,ψ) map over another from isolated energy values. The librational entropy of the molecule must be included in the assessment either by integrating over the conformational energy well containing the minima or by using analytical procedures (Lewis et al., 1973a). The purpose of the work presented in this thesis was to describe the entire conformational space of the residues in order to compute statistically averaged chain dimensions; thus, no attempt was made to locate minimum energy conformations nor should any conclusions of the conformational stability of specific conformations be made from this work. However, the relative stability of various classes of side chain and backbone conformations were deduced by including entropic considerations by the integration method, and these have been compared in Tables III.3; III.4; III.5; III.7 and III.8.
2. It was not stated explicitly (p.162) that elution behaviour is related to Stoke's radius rather than molecular weight. The shape of the protein standards used to calibrate the molecular weight elution profile is quite different from that of a disordered polypeptide chain and it would therefore have been more precise to run the calibration and test samples under denaturing conditions. However, the method was used (with reservations) to crudely assess the changes in the degree of polymerisation of the polymer being investigated. Independent measurements of the molecular weights of the polymers by end-group titrations always agreed with the estimations using gel-filtration, within the experimental accuracy of the methods (assessed as ±1,000 Daltons, see p.163).

3. Flory's matrix method (p.67) may only be applied when interactions between neighbouring units do not influence the rotations of \( \phi \) and \( \psi \) (see discussions on pp.42-43 and pp.117-118). The \( \beta \)-methyl valine residue is very sterically restricted but model building of di- and tri-peptides suggests that, provided the peptide bond is maintained in the planar transconfiguration, interactions between the t-butyl side chains and between side chains and adjacent backbones would not interfere with the independent rotational freedom of neighbouring residues.

4. Conformational characteristics of a random coil polypeptide would be determined solely by short range interaction if it were in a \( \theta \) solvent (p.39). In practice, however, random coil polypeptides are almost never examined in \( \theta \) solvents. Therefore, under conditions usually used, long-range interactions and polymer-solvent interactions exert their effects. In practice, however, the dimensions of randomly coiled polypeptides are usually measured in good solvents and then corrected by independent measurements of the 2nd osmotic virial coefficient in the solvent used.

5. Although the asymptotic limit can be more easily defined if \( C_n \) is plotted against \( 1/n \) instead of against \( n \) (as in Fig.IV.3) the asymptotic limits in Figs. IV.2, IV.3 and IV.1 were derived mathematically and not by graphical methods. Nevertheless, they are marked on the Figures for the convenience of the reader. The \( C_n \) v. \( n \) plot was chosen to describe the rate at which the characteristic ratio converges since it is the one most frequently used by other workers and should be more familiar to the reader.

6. p.58 C-H = 1.09 Å should read C-H = 1.00 Å

7. Eq.14 should read:

\[
C_\infty = \left( \frac{<r^2>_o}{n_p} \right)_{p=\infty} = \lim_{n_p \to \infty} \left[ (E_+<T>_o)(E_-<T>_o)^{-1} - (2/n_p)<T>_o(E_+<T>_o)^n_p (E_-<T>_o)^{n_p} \right],
\]

\[
{\text{p}}_{-\infty}
\]
ACKNOWLEDGEMENTS

My deepest gratitude is due to my friend, mentor and supervisor Professor S. J. Leach whose patience, knowledge and wisdom have enriched my professional life during my time as his student.

I would also like to acknowledge the help afforded me by Dr. A. W. Burgess on the theoretical studies in this thesis and for his supervision during a six month study leave period taken by Professor Leach.

I am indebted to many people for helpful discussions on various aspects of the work presented in this thesis, notably Mr. D. J. Evans and Drs. G. Némethy and P.E.E. Todd. Dr. I. D. Rae gave me a great deal of invaluable advice on the NMR experiments for which I am grateful. I must also thank Professor H. A. Scheraga for making available to me the computer program library of the Department of Chemistry, Cornell University.

There are others who have made a contribution to this thesis by generously giving their time and friendship to myself and my family during the last four years. They include: Dorothy Gillin, Winifred Lewis, Elise Minasian and Nina Vassalo.

Finally I must thank Suzanne Porritt and John Morley for their care and patience in, respectively, typing the manuscript and drawing the figures of this thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>(ii)</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>(iii)</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>(xi)</td>
</tr>
<tr>
<td><strong>CHAPTER I  WHAT MAKES PROTEINS FOLD?</strong></td>
<td></td>
</tr>
<tr>
<td>1. The Kinetics of Protein Folding</td>
<td>1</td>
</tr>
<tr>
<td>2. Structure Prediction of Proteins</td>
<td>7</td>
</tr>
<tr>
<td><strong>A. Empirical Methods of Structure Prediction</strong></td>
<td></td>
</tr>
<tr>
<td>(i) Statistical Correlations</td>
<td>8</td>
</tr>
<tr>
<td>(ii) Bayesian Methods</td>
<td>10</td>
</tr>
<tr>
<td>(iii) Stereochemical Rules for Prediction</td>
<td>12</td>
</tr>
<tr>
<td>(iv) Statistical Mechanical Approach</td>
<td>13</td>
</tr>
<tr>
<td><strong>B. Conformational Energy Calculations</strong></td>
<td></td>
</tr>
<tr>
<td>(i) Molecular Orbital Calculations</td>
<td>15</td>
</tr>
<tr>
<td>(ii) Empirical Energy Calculations</td>
<td>18</td>
</tr>
<tr>
<td>(a) The Incorporation of Peptide-Solvent Interactions</td>
<td>18</td>
</tr>
<tr>
<td>(b) The Selection of Starting Conformations</td>
<td>23</td>
</tr>
<tr>
<td>1. Using External Data</td>
<td>24</td>
</tr>
<tr>
<td>2. Using the Dominance of Short Range Interactions</td>
<td>26</td>
</tr>
<tr>
<td>(c) The Incorporation of Medium and Long Range Interactions</td>
<td>28</td>
</tr>
</tbody>
</table>
3. The Disordered Protein Chain
   A. The Unfolding of Proteins
   B. Helix ↔ Coil Transitions in Polypeptides
   C. The Statistical Mechanics of Randomly Coiled Polypeptides
      (i) Branched Polypeptides
      (ii) Monte Carlo Methods

CHAPTER II  THEORETICAL METHODS

1. Nomenclature and Abbreviations
2. Conformational Energy Calculations
   A. Generation of Conformations Using the Program ECEPP
   B. Empirical Potential Energy Functions
      (i) Electrostatic Energy
      (ii) Nonbonded Energy
      (iii) Hydrogen Bonded Energy
      (iv) Torsional Energy
   C. Data Input to ECEPP
   D. Selection of Rotameric States for Starting Conformations
   E. Energy Minimization
   F. Conformational Probability Calculations
   G. Representation of the (ϕ,ψ) Conformational Space
3. Calculation of the Characteristic Ratios (C₀) of Homopolypeptides
   A. Generation of Random Polypeptide Chains
CHAPTER III THE EFFECT OF SIDE CHAIN COMPLEXITY ON THE ENERGETICALLY PREFERRED CONFORMATIONS OF AMINO ACID RESIDUES

1. Introduction 75

2. Results 77
   A. Minimum Energy Conformations for the N-acetyl-N'-methylamides of β-methylalanine and β-methylvaline 77
   B. Conformational Energy Contour Diagrams 77
   C. Rotamer Populations for the N-acetyl-N'-methylamides of β-methylalanine and valine 82
   D. Conformational Probability Contour Diagrams 82
   E. Conformational Energy Calculations for the N-acetyl-N'-methylamide of Hydroxyethylglutamine 90

3. Discussion 98
   A. Conformational Energy Minima and the Total Conformational Space 98
   B. Probability of Occurrence of Backbone Conformations 104
   C. Probability of Occurrence of Side Chain Conformations 109

4. Conclusions 114
CHAPTER IV  THE AVERAGE RANDOM COIL DIMENSIONS
OF HOMO AND COPOLYPEPTIDES

1. Introduction

2. Results
   A. Unperturbed Chain Dimensions of Polypeptides
      with Increasingly Branched Side Chains
   B. Unperturbed Dimensions of Poly-L-hydroxy-
      ethylglutamine and its Copolymers

3. Discussion
   A. The Effect on Calculated Characteristic
      Ratios of Changes in the Energy Parameters
      and Covalent Geometry
   B. The Choice of Integration Interval in
      Characteristic Ratio Calculations
   C. The Effect of Incorporating Energy
      Minimization in the Energetic Statistical
      Weighting
   D. Treatment of Side Chains and Comparison
      with Experimental Results
   E. The Unperturbed Dimensions of Copoly-
      peptides Containing Hydroxyethylglutamine

4. Conclusions
CHAPTER V  STUDIES ON PROCEDURES FOR CONVERTING
POLYGLUTAMIC ACID TO POLYHYDROXYETHYL-
GLUTAMINE

1. Abbreviations 144
2. Introduction 145
3. Methods 151
   A. Materials 151
   B. Syntheses 153
      (i) Thin Layer Chromatography 153
      (ii) Preparation of the N-hydroxysuccinimide
           Ester of Propionic Acid 153
      (iii) Preparation of the N-hydroxysuccinimide
           Ester of Poly-L-glutamic acid 155
      (iv) Preparation of Polyhydroxyethyl-
           glutamine Using a Water Soluble
           Carbodiimide 155
      (v) Preparation of the Urea Derivative
           of N-ethyl,N'- (3-dimethylamino-
           propyl) Carbodiimide 160
      (vi) Cleavage of Polypeptides Containing
           Glutamic Acid with a Water
           Soluble Carbodiimide 160
   C. Quantitative Estimations 161
      (i) Estimation of Molecular Weights
          of Polymers 161
          (a) Ninhydrin Determination of
              Amino Terminal Groups 161
(b) Gel Filtration

(ii) Estimation of Polypeptide Concentration

(iii) Estimation of Pyroglutamyl N-Terminal Groups

4. Results and Discussion

A. Conversion of Poly-L-glutamic acid Side Chains to their N-hydroxysuccinimide Esters

(i) Estimation of the Unreacted Side Chain Carboxylic Groups of Poly-L-glutamic acid

(a) Stability of N-hydroxysuccinimide Esters to Hydrolysis

(b) Quantitative Assay of N-hydroxysuccinimide Esters by $^1$H-NMR

(ii) Effect of Reaction Conditions on Yields of N-hydroxysuccinimide Esters of Poly-L-glutamic acid

B. Preparation of Polyhydroxyethylglutamine Using a Water Soluble Carbodimide

C. Investigations into the Mode of Peptide Chain Fission by a Water Soluble Carbodiimide

(i) Changes in the Molecular Weight Distribution of Poly-L-glutamic on Treatment with N-ethyl,N'-(3-dimethylaminopropyl) Carbodiimide
(ii) The Effect of N-ethyl,N'- (3-dimethylaminopropyl) Carbodiimide on Polypeptides and Proteins Containing Single (Glu) and Double (Glu-Glu) Sequences

(iii) Changes in the $^1$H-NMR Spectrum of Poly-L-glutamic acid on Reaction with N-ethyl,N'- (3-dimethylamino- propyl) Carbodiimide

(iv) Possible Reaction Mechanisms for the Fission of Poly-L-glutamic acid by N-ethyl,N'- (3-dimethylamino- propyl) Carbodiimide

BIBLIOGRAPHY

APPENDIX I

(a) Standard Residue Data Used in Chapter II, Section 2.C

(b) Input Specifications for the Standard Residue Data

1. Atomic Numbering System
2. Amino Acid Residue Example: $\beta$-methylalanine
3. Amino End Group Example: Amino-COCH$_3$
4. Carboxyl End Group Example: Carboxyl-NHCH$_3$
| APPENDIX II                              | Mathematical Derivation of Statistical Weighting Procedure Used in Chapter II, Section 3.B | 235 |
| APPENDIX III                             | Listing of Program Flory CR Referred to in Chapter II, Section 3.D                         | 236 |
SUMMARY

The aim of the work presented in this thesis was to examine the restrictions that the size of the side chain of an amino acid residue imposes on the total conformational (energy) space available to the adjacent peptide backbone.

The conformational space of a series of N-acetyl-N'-methylamide derivatives of glycine, L-alanine, β-methyl-L-alanine, L-valine, β-methyl-L-valine and hydroxyethyl-L-glutamine was explored using empirical energy calculations with the most recently revised conformational energy parameters. Of the various regions of the total conformational space of the peptide backbone, the area surrounding the $C_7$ conformation is the most favoured for all the residues studied, except the ones with the bulkiest (β-methyl-L-valine) and longest (hydroxyethyl-L-glutamine) side chain. The steric constraints imposed by these two side chains favoured the extended backbone conformation. The preferred side chain conformations were predicted for those residues with multiple side chain rotameric states. L-Valine and β-methyl-L-alanine have only three side chain conformations, defined by the value of the variable side chain torsion angle $\chi^1$; it was found that the trans rotameric state ($\chi^1 \approx 180^\circ$) predominates for these residues. Hydroxyethyl-L-glutamine has seven side chain torsion angles which, using the rotational isomeric state model chosen for this study, generate 2,916 possible side chain conformations; however, one of these is overwhelmingly favoured (32%) and reasons for this preference are discussed.
In the randomly coiled state the conformational energy space of single amino acid residues can be used to compute the unperturbed dimensions of a polypeptide chain since under these conditions short range interactions predominate. This provides a convenient method of assessing the accuracy of conformational energy calculations since their dimensions can also be derived experimentally. The unperturbed dimensions of homopolyamino acids composed solely of each residue featured in this study and also of copolypeptides of the type (hydroxyethyl-L-glutamine-X-X)\textsubscript{n}, where X = glycine, L-alanine or \(\beta\)-methyl-L-alanine were derived. In the course of the investigation the conditions under which the approximations inherent in this type of calculation are valid were rigorously established. The most accurate estimates of the characteristic ratios of the homopolymers of glycine, L-alanine, \(\beta\)-methyl-L-alanine, L-valine and hydroxyethyl-L-glutamine were 2.12, 8.06, 7.02, 9.21 and 9.75. The side chain of \(\beta\)-methyl-L-valine imposed profound steric restrictions on the backbone which resulted in a slowly converging characteristic ratio (94.8 at a chain length of 500 residues).

The characteristic ratios of the copolypeptides (hydroxyethyl-L-glutamine-X-X)\textsubscript{n}, where X = glycine, L-alanine and \(\beta\)-methyl-L-alanine were 2.27, 8.56 and 7.73, respectively. Since hydroxyethyl-L-glutamine is a water solubilizing residue it should be possible to derive the characteristic ratios of copolypeptides of this type experimentally from their hydrodynamic behaviour in aqueous solution. In order to do this it is necessary to synthesize these copolypeptides which involves
establishing a method for converting the side chain of glutamic acid to that of hydroxyethyl-L-glutamine. The conversion of poly-L-glutamic acid to polyhydroxyethyl-L-glutamine via its active N-hydroxysuccinimide ester and also by coupling the side chains to ethanolamine using a water soluble carbodiimide was explored. It was discovered that the carbodiimide could cleave peptide bonds in poly-L-glutamic acid and some of the features of this reaction and possible mechanisms for it were investigated.
CHAPTER I

WHAT MAKES PROTEINS FOLD?
1.

The studies described in Chapters II-IV of this thesis are concerned mainly with the effects of side chain length and branching on the energetically preferred conformations of polypeptide chains. The context within which the studies are set is that of protein folding and the factors which may govern the pattern of folding in polypeptide chains of complex amino acid sequence. It therefore seems appropriate to set the research studies in context by surveying some of the current ideas on the dynamics, energetics and statistics of protein folding and unfolding. In general, the emphasis in this review chapter will be on current advances during the last two or three years since a number of comprehensive reviews survey the earlier work.

1. THE KINETICS OF PROTEIN FOLDING

The estimated time required for a random search for the conformation of lowest free energy of a protein is impossibly large, whereas the time required for the actual refolding of a denatured protein is very short, usually only a few minutes (Ristow and Wetlaufer, 1973; Anfinson, 1964, 1972, 1973; Schechter et al., 1970). There is evidence that proteins are composed of domains (each of 40-150 amino acid residues) which could fold independently since they are continuous regions which can be completely surrounded by a closed surface through which the polypeptide backbone enters and exits only once (Wetlaufer, 1973). Such essentially independent centres for folding would facilitate the rapidity with which the protein attains its overall conformation, but
2.

even a chain length of 100 residues would take about $10^{85}$ s
to randomly sample all conformations available to it
(Wetlauffer, 1973).

Thus it has been recognised that the folding of a protein after biosynthesis into its native conformation is a cooperatively directed phenomenon. Reviews which cover the kinetics of protein folding have appeared recently (Anfinson and Scheraga, 1975; Schulz, 1977; Némethy and Scheraga, 1977). This discussion will be restricted to the currently popular theories of how the folding of a protein proceeds; that is, either by a diffusion collision mechanism or via the formation of nucleation sites which direct the process. Quite how this cooperativity is expressed is not known, but two structural features of proteins are frequently proposed as possible nucleation sites - the β-bend or chain reversal and the α-helix. These two local structures play an important role in the two most favoured nucleation theories of protein folding, the β-bend in that originally proposed by a group of workers at Cornell University, New York, (see e.g. Lewis et al., 1971; Scheraga, 1973a; Anfinson and Scheraga, 1975) and the α-helix in the theory put forward by V.I. Lim and colleagues at the Institute of Protein Research, Moscow (Gudkov et al., 1977; Lim, 1978).

Both mechanisms suggest the formation of short regions of secondary structure as the first stage in the folding pathway. Lim et al., suggest that short α-helices of 10-20 residues bearing a cluster of hydrophobic side chains on the helix surface would unite into a highly helical globule by hydrophobic bonding. The overall shape of the structure thus formed would be retained whilst the α-helices undergo
rearrangement into the native conformation of the protein. The Cornell group also recognise α-helices as being important in the first stages of protein folding but point out that short segments of α-helix are extremely unstable (Poland and Scheraga, 1970). They have suggested that after the formation of local secondary structure, including α-helices, such as may be found in the completely folded protein, short range interactions in specific dipeptide sequences cause chain reversals to form which direct the local secondary structure into conformations which can be stabilised by long range interactions. The remainder of the polypeptide chain then folds around these foundations. The α-helical regions found in the native protein are clearly correlated with those sequences calculated for denatured proteins by assigning amino acid residues to three categories according to their helix-forming potential (Lewis et al., 1970). This suggests that although helices may form in the folding protein chain before other secondary structures appear these cannot be identified with the small ephemeral nucleation sites proposed by Lim and coworkers, but are sustained during the remainder of the folding process.

Evidence for either theory is scanty. Chain reversals occur widely in proteins in numbers that bear a direct linear relationship with the length of the polypeptide chain (Rose and Wetlauffer, 1977). But whether these are 'passive' structures arising through interactions between remote segments of the chain as suggested by Schulz (1977), Matheson and Scheraga (1978) and by Rose (1978) (the last two groups of workers
identified chain reversals as loops between hydrophobically bound regions) or whether they are 'active' structures determined by the energetic preferences of certain combinations of dipeptides (Zimmerman and Scheraga, 1977) for β-bend backbone conformations, that then direct these regions together is uncertain. There is evidence, however, that short peptide sequences of 3-6 residues can independently form β-bends (e.g. Von Dreele et al., 1978; Kopple et al., 1975; Walter et al., 1972; Howard et al., 1973) indicating the importance of local interactions in promoting this type of conformation.

It is an accepted fact that proteins are formed on a template of messenger RNA by the ribosome polymerising the amino acids in a step-wise process. The polymerisation proceeds from the amino terminus end of the growing peptide chain which is attached to the ribosome only at the point at which the next amino acid in the sequence is to be added (Dintzis, 1961). A polypeptide chain of the size of about 100 residues takes about one minute to be synthesized; does the growing peptide chain start folding during this period and does this process influence the final conformation of the finished protein? If this is the case, there is little evidence for it. In fact, long range interactions are very important in establishing the overall conformation of a protein. There are many examples which demonstrate that most of the protein chain must be present for the molecule to fold into its native conformation; for example the S-peptide and S-protein of ribonuclease do not separately have the conformations they adopt in the native molecule but when added together they interact to give a
3-dimensional structure that resembles ribonuclease despite the fact that they are not covalently linked (Richards, 1958). Isolated fragments of other protein molecules (e.g. myoglobin, Epand and Scheraga, 1968) also often fail to assume the conformation they have in the intact molecule. However, this does not preclude the possibility that short regions of secondary structure do form in the polymerising protein and that having acted as nucleation sites they then rearrange during the final stages of the folding process as proposed by the Russian workers. Some evidence of this is furnished by the fact that amino acid residues known to have a very low propensity for α-helical conformations tend to predominate in the C-termini of helical segments (Kotelchuck et al., 1969), whereas if the helix were growing in both, or either direction randomly, such residues would be expected to congregate at both ends of the helix. In addition, the folding of nascent polypeptide chains may be important for proteins in which long chain interactions are not so important in stabilizing their overall structure; for example, polyribosomes containing collagen chains were recently isolated and shown to adopt a hydrogen bonded structure. These nascent chains were almost fully hydroxylated and were resistant to pepsin digestion suggesting that triple helix formation had occurred between the nascent chains before release from the endoplasmic reticulum surface (Veis and Brownell, 1977).

Proteins are no longer considered to be the inflexible well-defined structures that they appear from X-ray crystallography. Theoretical investigations using molecular dynamics (McCammon et al., 1977) indicate that the protein molecule
can sample a range of conformations that are in the
neighbourhood of its "X-ray structure", but are not identical
with it. An immunological technique that measures the degree
of disorder in antigenic regions of a protein molecule has
also shown that protein molecules are not rigid structures
(Furie et al., 1975; Hurrell et al., 1977; Creighton et al.,
1978). However, they do not appear to be so conformationally
equivocal that major displacements can occur in their secondary
structure without perturbing their overall shape. X-Ray
diffraction studies at near atomic resolution on crystalline
proteins would not have met with such signal success if "site
occupancies" were substantially below unity.

The diffusion-collision model (Karplus and Weaver,
1976) considers the protein molecule to be divided into micro-
domains, each of which is short enough for all conformational
alternatives to be searched through rapidly. Through diffusion,
several microdomains collide and coalesce into a structural
region which has the native conformation. Since about 1/3 of
all residues in globular proteins are involved in chain
reversals (Lewis et al., 1971; Chou and Fasman, 1974b) a protein
consists of 8-10 residue links between turns and Karplus
proposes these as the microdomains; chains of this length would
only take about $10^{-4}$ s to fold. The time for a pair of micro-
domains to diffuse together and coalesce, however, would be in
the range of 0.01 to 10 s, and for several units to come
together instantaneously would take an inordinately long time.

None of the current hypotheses adequately satisfies
the requirements of a mechanism of protein folding, although
they all have merits in terms of explaining various aspects
of it. The true answer probably involves elements of each of
them.
2. **STRUCTURE PREDICTION OF PROTEINS**

If either α-helices or β-bends are nucleation sites for the folding mechanism then the sequences of amino acids found in these structures may be expected to be more closely defined than in structures that have a less critical role in the overall shape of the protein. As the protein samples the conformational space available to it during folding, the nucleation sites would be expected to be short sequences of amino acids with a more limited conformational space available to them, energetically restricted by short range interactions so that these sequences in the chain are statistically more likely to 'linger' in any energy wells found during a conformational search, whether random or directed. For nearly 20 years (Scheraga, 1961) since the first protein, insulin, was sequenced (Ryle et al., 1955) algorithms have been sought that will determine the shape of a protein in a given environment from a knowledge of its amino acid sequence. The approaches used have fallen into two broad categories. First are empirical methods which attempt to predict the conformation of a protein by correlating the increasing body of X-ray diffraction data of the structure of proteins, of which over 50 have now been elucidated (Matthews, 1977), with their amino acid sequence. Various empirical approaches have employed the powerful tools of classical statistical analysis, information theory, stereochemical theory and statistical mechanics and have centred largely on predicting the positions of β-bends, β-pleated sheets and α-helices in proteins from their amino acid sequences. The various predictive models have been
assessed by several workers (Schulz et al., 1974; Burgess et al., 1974; Matthews, 1975; Tanaka and Scheraga, 1976a,c; Maxfield and Scheraga, 1976; Argos et al., 1976; Lenstra et al., 1977; Lenstra, 1977) and reviewed very recently by Chou and Fasman (1978).

The second approach attempts an a priori prediction of the structure of a protein from a consideration of the physico-chemical interactions of the atoms within each residue in the sequence. It relies on the hypothesis that the stability of a unique structure under native conditions implies that the Gibbs free energy of the system (in this case the protein and its solvent) is at a minimum for this conformation, and thus searches for a folding of the protein chain that minimizes its conformational energy.

The empirical approach has been reviewed extensively by Chou and Fasman (1978) and only a brief description of the methods available will be given here. The a priori approach of conformational energy calculations will be covered more extensively since it is to this method that the work in this thesis applies.

A. EMPIRICAL METHODS OF STRUCTURE PREDICTION

(i) Statistical Correlations

Many predictions of secondary structure are based on parameters obtained from correlations between amino acid sequence and conformation of proteins using classical statistical analytical techniques. These attempts include the work by Nagano (1973, 1974), Chou and Fasman (1973, 1974a,b), Kabat and Wu (1973a,b; Wu and Kabat, 1971, 1973) and Krigbaum
and Knutton (1973). Amino acid residues are considered to exist in only 2, 3 or 4 conformational states - helix, $\beta$-sheet, and in some studies also as 'structureless' and $\beta$-bends. The frequency with which each of the 20 naturally occurring amino acids inhabit each of three states in proteins of known structure is then analysed statistically (the precise method used varying in each study). This information is then used to assign a conformational state to each residue in the amino acid sequence of the unknown protein. The disadvantages of the method are three-fold. The first limitation is the paucity of X-ray diffraction data on proteins so that there is a maximum of 50 proteins in the data base available for the predictions and these include many homologous groups of proteins which, if all members were included, would bias the analysis. In some cases not all of the available data has been used for the statistical analyses, e.g. only 11 proteins were the source of data for the study by Kabat and Wu (1973a). Second, this analytical approach assumes that interactions between amino acid residues close together in the primary sequence have the major role in determining the configuration of the backbone in that region, whereas the conformational states of the amino acid residues in a protein crystal are determined also by medium and long-range interactions. Thirdly, there is the problem that a four- or more-state model - let alone a two- or three-state model - simply does not represent the conformational flexibility of amino acid residues in proteins. It has been claimed (Chou and Fasman, 1978) that many empirical algorithms are predicting about 80% of residues correctly as helix or non-helix and as $\beta$-sheet
or non $\beta$-sheet, and with an increase in the size of the data set, as the X-ray diffraction structure of more proteins is elucidated, this level of accuracy should improve. However, the prediction index used to assess the accuracy of a predictive algorithm is often defined differently by each worker, thus making it difficult to compare the accuracy of different methods simply on the claims of the various workers. Also, those methods with high prediction indices have usually only attempted to place each amino acid residue in one of two or three conformational states, but without an improvement in the accuracy of representing the many conformational states of an amino acid residue, it seems unlikely that an approach based on statistical correlations can be used to accurately predict the folding of a polypeptide chain solely from its amino acid sequence. In fact, it has been asserted (Burgess and Scheraga, 1975a) that even a perfect algorithm that correctly assigned every residue in bovine pancreatic trypsin inhibitor to one of five conformational states would not lead to the native structure without the introduction of additional information such as the nature of disulphide bond pairing and finally energy minimization.

(ii) Bayesian Methods

A prediction algorithm that quantifies the information provided by the statistical analytical method has been developed by Robson and Pain (1971, 1974a,b,c; Robson, 1974). Using this information theory approach a statistical analysis of the relation between sequence and conformation in 25 proteins was carried out by Robson and Suzuki (1976). They assigned each residue to one of ten conformational states, thus:
The role of each type of amino acid residue in determining its own conformation and that of its neighbours was quantitatively assessed and the predilection for certain conformational states was correlated with a number of physicochemical properties of the residue using cluster analysis. It was found that the conformational role of a side chain is dependent on its hydrophobicity and its ability to receive or donate a hydrogen bond. They also found that the values of conformational parameters were subject to statistical fluctuations when obtained for a small number of proteins; even using 25 proteins the conformational parameters for a few types of residues still had not converged, demonstrating that the largest data base available at the time of carrying out a statistical investigation should be used. Robson and Suzuki
demonstrated the application of the parameters to structure predictions with a worked example for predicting \( \alpha \)-helical regions in hen egg-white lysozyme and obtained an accuracy of 75-80\%. A similar approach using Bayesian methods of analysis (Maxfield and Scheraga, 1976) assigned 56\% of the residues in 20 proteins to one of five conformational states which together covered the entire conformational space available to the backbone of a residue. [These five states can be roughly described as extended chain, right- and left-handed \( \alpha \)-helix and the two "so called" bridge regions occurring between \(-180^\circ < \phi = 0^\circ \) or \( 0^\circ = \phi < 180^\circ \).] Although the degree of success of this prediction may seem small, a five-state model provides more information about the topography of the chain than does a two-state model that assigns, for example, helix and coil perfectly - if there is a low helix content in the protein little information will have been gained about its structure. The results of these workers also emphasized the size of the data base as a limitation of accuracy in empirical predictive methods and noted that the experimental error in the determination of the X-ray structure of proteins may be sufficient (the errors in \( \phi \) and \( \psi \) can be as much as \( \pm 30^\circ \)) to cause many residues to be assigned to the wrong conformational state in methods that use a multi-state model.

(iii) **Stereochemical Rules for Prediction**

An alternative approach to the use of data from proteins of known structure is that of Lim (1974a-d). With the use of space filling models and the voluminous experimental data on 3-dimensional structures of water soluble globular
proteins Lim has attempted to determine the types of sequences which would be compatible with the formation of \( \alpha \)-helix and \( \beta \)-structure. Particular attention was paid to the shielding of non-polar side chains from water, the solvation of polar groups, and the tight packing of non-polar side chains in the core of the protein. From this set of stereochemical rules Lim derived algorithms (1974d) using which \( \alpha \)-helix and \( \beta \)-structural regions were localized in 25 globular proteins. 70\% of all the residue were correctly assigned to \( \alpha \)-helix, \( \beta \)- and irregular structures. Although this method circumvents the need for a large X-ray crystal data set and the errors inherent in using it, such a heuristic method may well ignore important factors that govern protein folding.

(iv) **Statistical Mechanical Approach**

In a series of papers Tanaka and Scheraga (1976a-d; 1977a,b) developed a statistical mechanical treatment of protein conformations within the context of 1-dimensional short-range interaction models. In the first paper (1976a) a method was presented for evaluating statistical weights of various conformational states of amino acid residues on the basis of conformational information from X-ray crystal structures of native proteins. Then a three-state model for specific sequence polypeptides that included helical (h), extended (ε) and unstructured or 'coil' (C) states was formulated (1976b) and its application to the prediction of protein conformation was described (1976c). This treatment was extended to a four-state model by including chain reversal (R and S) states for residues in the \( i + 1 \) (R) or \( i + 2 \) (S) positions,
respectively, in a β-bend as well as h, ε and C states and the statistical weights for these states were evaluated on the basis of the X-ray coordinates of native proteins (1976d). Subsequently, (1977a) a multi-state model (that theoretically can include any number of conformational states) was formulated for seven conformational states, right- and left-handed α-helical, chain-reversal (R and S), extended, right- and left-handed bridge region and "other" or coil states and applied to the prediction of protein structure. However, while a multi-state model specifies the folding of the backbone of a protein more precisely than do models using a smaller number of states the use of them may be premature as there are not yet sufficient experimental X-ray data available for proteins to determine the statistical weights of the multi-state model accurately. Although the theoretical background to this work is a statistical mechanical one it was found necessary to introduce a simple empirical rule to conserve computer time in assigning α-helical states (1976c). This rule was that the α-helical conformation cannot form if the same state of charge is present at residues i and (i + 3) and at i and (i + 4) without an intervening or neighbouring oppositely charged amino acid. It was formulated on the basis of observations by Maxfield and Scheraga (1975) that such i to (i + 4) electrostatic interactions stabilized α-helices in proteins, and was preserved by Tanaka and Scheraga in subsequent extensions of the conformational states to include chain reversal conformations (1976d) but was eventually eliminated (1977b) by the use of probabilities of higher order than the first or second order probabilities used previously. (In order to predict the conformational states of n residues without recourse to
empirical information it is necessary to use an $n^{th}$ order probability rather than the first order probability for the state of a single residue.) The statistical mechanical method eventually evolved for predicting protein conformation was applied to assigning $h$, $\varepsilon$ and $C$ states to bovine pancreatic trypsin inhibitor and chlostridial flavodoxin (1977b). The accuracy of the predictions was no better, however, than with theoretically less analytical procedures (e.g. Chou and Fasman, 1977). Nevertheless, the authors point out (1977b) that the method has the potential for improvement in reliability since higher order probabilities (i.e. $n > 3$) could be used and that more accurate statistical weights can be calculated as the data bank of protein X-ray crystal structures increases.

B. CONFORMATIONAL ENERGY CALCULATIONS

In this a priori approach to the prediction of protein structure, the most probable peptide conformations are arrived at by calculating the global energy minimum in the multi-dimensional conformational space available to a given amino acid sequence. The two approaches used for this are:

(i) Molecular Orbital Calculations

The molecular orbital approach attempts a quantum mechanical treatment of the problem using methods that deal variously with all valence, $\sigma$ and $\pi$, or even all electrons. The methods are, respectively, the EHT method (Extended Huckel Theory), the CNDO/2 method (Complete Neglect of Differential Overlap) and the PCIL0 method (Perturbation Configuration Interaction over Localised Orbitals) and finally the more
accurate so-called ab initio procedure. These methods, and their applications to the conformation of amino acid residues and peptides, have been recently reviewed (Pullman and Pullman, 1974) and will not, therefore, be described here. This discussion will be restricted to a critique of the molecular orbital approach and its limitations not covered in the Pullman's review.

The application of quantum mechanical methods to the prediction of peptide and polypeptide structures is limited by the large amount of computer time required to carry out ab initio calculations (Pople and Radom, 1973). Thus, the methods used (EHT, CNDO and PCIL0), although starting from a fundamental viewpoint are far from rigorous and the details of the approximate quantum mechanical energy surface can be erroneous (Tonelli, 1971). It is also much more difficult to modify these calculations to give better agreement with experimental observation than it is with empirical energy calculations since the theory depends on a much more limited set of axioms.

The molecular orbital method has therefore been restricted to predicting the conformation of small molecules related to peptides such as amino acid residues (Pullman and Pullman, 1974) and to compounds of pharmacological interest (Pullman and Courrière, 1973; Farnell et al., 1974). Its most useful applications in the structure prediction of proteins, in the opinion of the author, has been in solving specific problems associated with developing potential functions used in empirical potential energy calculations. Thus, CNDO/2 calculations have provided atomic partial charges for estimating the electrostatic energy component of the total empirical
conformational energy (Momany et al., 1975) and were also used to develop an empirical hydrogen bond potential function (McGuire et al., 1972; Momany et al., 1974). Intrinsic rotation potentials in biological molecules have been studied using ab initio calculations on model compounds (Pople and Radom, 1973) and CNDO/2 calculations on N-methylacetamide have indicated that the minimum energy configuration of the peptide bond is non-planar with $\omega = 167.5^\circ$ (Ramachandran et al., 1973), although this is not in agreement with empirical energy calculations which result in a global minimum for N-methylacetamide in the planar trans conformation with $\omega = 180^\circ$ (Zimmerman and Scheraga, 1976) nor with the crystal data of Hagler et al. (1976) of 12 N-methylacetamides, all of which had values of $\omega = 180 \pm 6^\circ$, including N-methylacetamide where $\omega = 180^\circ$.

Discrepancies between the structures predicted by molecular orbital and empirical energy methods have also been reported for the model compound N-acetyl-N-methyl-α-aminoisobutyrylamide (Burgess and Leach, 1973). CIIIO calculations on this highly sterically restricted peptide suggested that the intramolecularly H-bonded C$_7$ ring conformation was energetically favoured. Empirical energy calculations, however, suggested that the right- or left-handed α-helical conformations were dictated by steric constraints. X-Ray diffraction studies have now shown that the right-handed α-helical conformation is taken up in the crystal (Aubry et al., 1978) while infra-red data are equivocal.

Although the molecular orbital approach has the potential to define the conformation of lowest energy of a polypeptide more accurately than the more subjective empirical
calculations, until faster and more economically accessible computers are developed it seems that molecular orbital calculations will not be directly applicable to the protein folding problem.

(ii) **Empirical Energy Calculations**

These are based on the partitioning of the potential energy of the system into several discrete contributions which are then evaluated with the aid of potential energy functions based on empirical observations of the energetic behaviour of small model compounds. Many excellent reviews on the application of empirical energy calculations to the prediction of stable conformations of proteins and on the development of potential energy functions have been published - in this decade by Brant (1972), Ramachandran (1974), Scheraga (1971, 1973a, 1974a, b) and by Némethy and Scheraga (1977). Thus, the refinement of amino acid residue geometry, and the development of energy parameters and potential functions all of which have contributed to the increasing accuracy with which the conformations of small peptides have been predicted will not be reviewed in this thesis which will concentrate on discussing the strategies used in energy calculations, the limitations of the method that still need to be overcome to perfect the technique and advances that have been made in the past year.

(a) **The Incorporation of Peptide-Solvent Interactions**

In almost all attempts to predict the 3-dimensional structure of a peptide using empirical energy calculations, interactions of the molecule with solvent have largely been
ignored - simply because they are not well understood at the molecular level. Solvent is thought to influence conformational stability in several ways (Némethy and Scheraga, 1977):

(i) By acting as a dielectric medium and influencing the electrostatic interactions between charged or partly charged groups;

(ii) By interacting with polar groups and modifying the strengths of hydrogen bonds;

(iii) By changing the interactions between groups (e.g. hydrophobic interactions) due to the effects of such groups on the structure of nearby solvent molecules and

(iv) By the dipole and quadropole moments of the solute molecule polarising the surrounding medium.

Most of the effort of workers in the field has been directed to exploring effect (ii). Krimm and Venkatachalam (1971) considered the effect of introducing water molecules into the vicinity of the carbonyl groups of poly-L-proline and calculated the conformational energies of the cis(I) and trans(II) forms of the polymer. The trans form was shown to be much more sensitive to interactions with water than was the cis form.

Tanaka and Scheraga, (1975a-c) also examined the effect of solvent (n-butylalcohol and benzylalcohol in this case) on peptide carbonyl groups in the I-\longrightarrow-II transition for polyproline. Their theoretical results reproduced the experimental transition curves satisfactorily.
One of the problems to be overcome in incorporating solvent effects into calculations on the conformational stability of globular structures such as proteins, is the variation in exposure of the backbone and side chain atoms of amino acid residues on the outside and inside of proteins. The ease of solvation of the peptide group varies with its accessibility and the latter depends on the conformational states available to the different residues. These factors have been examined for glycine, alanine and serine dipeptides (Ponnuswamy and Manavalan, 1976; Manavalan et al., 1977). It was shown that the only region preferred for solvent accessibility is the left- and right-handed bridge regions for all three dipeptides and the right-handed bridge region for glycine. This observation may explain why many amino acid residues in proteins exist in such conformations (Burgess et al., 1974) in spite of the fact that energy calculations which do not include solvent interactions (Lewis et al., 1973a; Zimmerman et al., 1977a) predict that these conformations are energetically unfavourable. However, solvent accessibility may not be a dominant factor for residues in the interior of proteins. The preferences of amino acid residues for the inside or outside of proteins were analysed by Werz and Scheraga (1978), and although alanine and glycine were shown to have a strong preference for the solvent accessible exterior of the 20 proteins analysed, serine had a below average preference. This study confirmed the qualitative impression that polar groups are generally found on the outside of proteins and non-polar residues on the inside, but seven of the residues
had preferences which were not consistent with the nature of their side chain, indicating that other factors may be more important in determining the conformation of a protein.

The most serious attempt to incorporate solvent interactions into energy calculations is that by Hodes et al. (in press). This study uses a modified hydration shell model originally developed by Gibson and Scheraga (1967) and revised by Hopfinger (1971) and includes the solvent effect (iii) as well as (ii). Free energy terms were introduced to account for specific hydration due to water-solute hydrogen bonding and for nonspecific hydration describing the interaction of the solute with water in the first neighbour shell. The dielectric constant was doubled (over the value of 2 used for calculations in the absence of water) to make a rough allowance for the effect of solvent on the electrostatic interactions between atoms carrying partial charges. The minimum energy conformations for the 20 blocked amino acids in the hydrated state were thus calculated and used in a subsequent study (Hodes et al., in press) on the conformational stability and formation of bends in blocked dipeptides. However, the results of comparing bend probabilities predicted with or without the effects of hydration (Zimmerman et al., 1977b) on the one hand with the bend probabilities of dipeptide sequences observed by X-ray crystal structures of proteins on the other hand are disappointing. There is an improvement in correlation when water is included for about 50% of the 23 dipeptides considered, but with a concomitant worsening of agreement in the other 50%. The peptides studied were variously composed of alanine, glycine, proline, aspartic acid, asparagine,
valine and serine residues. The authors claimed that an improvement occurs for polar dipeptides because these are solvent-accessible, but that non-polar dipeptides that are usually found inside the protein are better served by a model that does not include hydration. However, of the 11 dipeptides that show no improvement in correlation on hydration only three would be classified as "inside bends" on the Werz and Scheraga (1978) criteria. Nevertheless, this ambitious work is an important milestone in the task of accurately predicting the conformation of peptides using empirical energy calculations and can be compared to the early pioneering hard-sphere model used to calculate energetically allowed and disallowed regions of the ($\phi,\psi$) conformation space while ignoring their relative stabilities. It still remains for the solvent effects (i) and (iv) to be incorporated into the model of behaviour of a polypeptide in solution.

The importance of effect (i) is demonstrated by a study by Maxfield and Scheraga (1975) in which it was found that the helix-forming ability of charged amino acids in proteins is profoundly influenced by the effect of neighbouring charged side chains. From a statistical analysis of 14 proteins of known structure these workers found that a positively charged side chain four residues away from a glutamic acid greatly enhances its probability of being helical. Similar results were obtained for aspartic acid, lysine, arginine and histidine. Thus, $\alpha$-helices are obviously stabilized by favourable charge interactions. Since the natural environment of a protein is aqueous the charge effects between ionisable
side chains as well as the N- and C-termini in a medium of such high dielectric constant (~80) must be of paramount importance in determining the conformation of the protein. Incorporation of these effects into empirical energy calculations should further improve their accuracy.

(b) The Selection of Starting Conformations

Given that inevitably it will be possible to accurately calculate the free energy of any protein or peptide conformation of known amino acid sequence the problem then arises as to the practicability of this technique in predicting the structure of lowest energy. There are two closely related problems here. The first is that it is not possible even with small peptides to search the complete multi-dimensional conformational space available to them for structures of low energy. It would take too much computer time to calculate all the possible interactions between pairs of their constituent atoms. Thus, various strategies have been developed to reduce the number of conformations whose energy needs to be calculated, by systematically excluding starting conformations that are unlikely to yield low energy conformations. The second is the existence of many minima in the conformational energy space. Using present energy minimization algorithms it is difficult to "jump" from one low energy region of conformational space to another region of conformational space, as each is explored, from the area selected as potentially containing a low energy minimum. Levitt and Warshel (1975) introduced a 'normal mode thermalization' method to avoid 'trapping' a polypeptide in an energy well surrounded by low
barriers. After each minimization, a new random starting point close to the well is generated by introducing thermal fluctuations into the equations of motion for the molecule and then randomly stopping the thermal vibration. However, as Robson (1975) pointed out, this method does not avoid the possibility that during minimization the simulated folding pathway is confined to a trivially shallow channel which is avoided by more conventional minimization procedures. On the other hand, on using the latter, great care must be exercised in reducing the number of starting conformations so that structures are not eliminated that might have led to low energy minima or even the global minimum on minimization.

The strategies used in reducing the conformational space to be searched for low energy minima fall into two categories. Those that rely on external knowledge of the possible structure under investigation and those that rely on the dogma that short range interactions predominate in determining the conformational stability of a protein. [The dominance of short range interactions in macromolecular configurations was first proposed by Flory (1953) and will be discussed in section 3.C of this chapter. Its application to determining the conformation of a protein has been reviewed by Scheraga (1973b).]

1. Using External Data

Energy calculations have been used for a number of years to refine X-ray crystal data (reviewed by Scheraga, 1974b); with the advent of high resolution NMR techniques the elucidation of the solution conformation of peptides is now
possible. However, this technique can, as yet, only determine peptide conformations inexact ly and often a wide number of structures are consistent with the NMR data available on a peptide. Nevertheless, the information that NMR provides as to likely values of the variable (φ) torsion angles of a peptide can be used as starting points for energy minimization (Gibbons et al., 1970) as, for example, in the calculation of the low energy conformation for met-enkephalin (Iso gai et al., 1977). It has also been shown that NMR can be used to limit the permissible φ and ψ values in peptides where NOE (Nuclear Overhauser Effects) are measurable (Leach et al., 1977; Rae et al., 1977).

Other starting points for possible low energy conformations of a peptide or protein can be obtained by searching for homologous amino acid sequences in related proteins of known structure and then minimizing these in the context of the unknown protein. Thus, a structure for α-lactalbumin was computed using the known crystal structure of lysozyme which has a very similar amino acid sequence (Warme et al., 1974). Conformations for three snake venom inhibitor proteins were computed from homologous sequences in bovine pancreatic trypsin inhibitor (Swenson et al., 1977). The predictive methods described previously in this chapter can also be used to assign probable conformational states to each residue in an amino acid sequence for subsequent minimization as proposed by Burgess et al., (1974) and Levitt and Warshel (1975). Thus, certain residues in the peptide chain were fixed in the α-helical conformation in the simulated folding of bovine pancreatic trypsin inhibitor (Levitt, 1976) and of carp
myogen (Warshel and Levitt, 1976) and Burgess and Scheraga (1975a), using a five-state predictive algorithm determined an initial conformation for bovine pancreatic trypsin inhibitor which on subsequent energy minimization gave a structure with some similarities to the native protein.

Model building has been used to find feasible shapes of peptides and to eliminate unlikely ones, for example, in the study on met-enkephalin cited above. Due to its subjective nature, however, this strategy must be used very conservatively.

2. Using the Dominance of Short Range Interactions

For large peptides and proteins data from the external sources described above must be used to reduce the conformations to be minimized with respect to energy, but for small peptides a more rigorous search of the conformational space is feasible. A total search is still impractical, but by relying on the dominance of short range interactions in the overall conformation of a protein it is possible to assign to each amino acid residue in the sequence only those conformations not excluded by interactions between the side chain and backbone or between backbone atoms in the peptide groups on either side of the residue, i.e. those minima that are within a few kcal/mole of the global minimum in the conformational space of the amino acid residue 'blocked' by N-acetyl and N-methylamide end-groups.

Thus, some minimum energy conformations of the N-acetyl-N'-methylamide derivatives of the constituent amino acids were used to predict the structure of the tripeptide, thyrotrophin releasing factor (Burgess et al., 1973) and four tetrapeptide variants containing lysine, threonine, glycine
and aspartic acid (Howard et al., 1975). However, even with a tetrapeptide, if the permutations of all minimum energy conformations of the constituent amino acids were used the total number of starting conformations for minimization would be of the order of $10^6$. Thus, a selection strategy to reduce this number is necessary but one which will not exclude those likely to lead to low energy conformations upon minimization. In the works cited above only a few conformations, corresponding to the lowest energy minima for each residue, were used as starting points, but with this procedure the cut-off point can be arbitrary - many minima often differ by only a fraction of a kcal/mole and the preference of one minima over another depends not only on its actual free energy value but also its librational entropy (Lewis et al., 1973a). A more satisfactory, i.e. less subjective, system for selection of starting conformations involves finding favourable low energy conformations of blocked dipeptides and tripeptides from the minima of the individual residues and then building up starting conformations for longer peptides from these. At each stage, higher energy minima are eliminated, leading to a reduction of starting points for the complete peptide. This strategy was used recently by Isogai et al. (1977) with met-enkephalin, Simon et al. (1978) with two tetrapeptides, Anderson and Scheraga (1978) with a contraceptive tetrapeptide and Fitzwater et al. (1978) with tuftsin. It is obvious, however, that even with these restrictions, long peptides (say >6 residues) cannot be handled entirely by this method, not only due to the increasing number of permutations to be handled but because long range interactions
become more relevant in their overall structure. Thus, a starting conformation for a particular residue may have been eliminated as not being energetically favoured when present in the single residue or in a dipeptide, but which is tolerated in the larger peptide because, directly or indirectly, it permits the structure as a whole to be stabilized by long range interactions. Nevertheless, with the discovery of an increasing number of biologically active small acyclic and cyclic peptides this technique has an important role in structure prediction studies. Thus, it is of paramount importance to reliably establish the relative importance of low energy minima of amino acid residues. It is to this aspect of structure prediction by empirical energy calculations that the results reported in Chapter III of this thesis are pertinent.

(c) The Incorporation of Medium and Long Range Interactions

To accurately predict the structure of longer peptides medium and long range interactions between atoms must be calculated. Until recently, because of the large amount of computer time required, only a portion of the long range interactions have been computed and those interactions beyond some arbitrary cut-off distance from any given atom have been neglected (Warme and Scheraga, 1974). While these interactions are individually very small, the longer the peptide the more significant they become and since, at such a distance, the attractive term in the non-bonded potential function far outweighs the repulsive term they could serve to make the structure of a globular polypeptide more compact. There have been recent attempts to overcome this problem (Pincus and
Scheraga, 1977) by devising a rapid approximate procedure for computing hitherto neglected long range interactions. Beyond a critical distance of separation (8-17 Å depending on the side chain of the residue) two interacting residues are treated as spheres of given radii instead of as individual atoms - the long range interactions are thus computed in terms of the properties of the residue instead of the individual atoms of each residue. By considering such interactions between two spheres instead of between all pairs of atoms there is a saving of about 40% in computer time for a polypeptide of about 50 residues.

The refinements in energy parameters and potential energy functions, the incorporation of effects which up to now have not been adequately considered - such as hydration and long range interactions, and the improvement in strategies designed to overcome the multiple minima problem have, in the last five years, brought the accurate prediction of the native conformation of a polypeptide by empirical energy calculations into the realm of possibility.

3. **THE DISORDERED PROTEIN CHAIN**

Our knowledge of the forces that govern the stability of the 3-dimensional structure of the folded polypeptide chain has partly been obtained from observations made at various degrees of disordering of the polypeptide chain. Thus the relative importance of short, medium and long range interactions in determining the folding pathway has been derived from three areas of study, the unfolding of native globular proteins, order ↔ disorder transitions in homo and
copolymers and the application of the general principles of polymer random-flight statistics to the completed disordered polypeptide chain.

A. THE UNFOLDING OF PROTEINS

The techniques used to examine the unfolding polypeptide chain rely on detecting the exposure of different parts of the chain on unfolding either by their susceptibility to proteolytic attack by enzymes such as α-chymotrypsin (e.g., Klee, 1967) or in recent years using physical techniques such as NMR, laser Raman spectroscopy and ESR. The proteins most often used as models for unfolding studies have been ribonuclease (which has a chain length of 124 residues with 4 disulphide cross links), staphylococcal nuclease (which consists of 147 residues, with no disulphide bridges) and bovine pancreatic trypsin inhibitor (with 55 residues and 3 disulphide cross links) and one of the features of the unfolding process that has been extensively explored is the detection of possible stable conformational states during the unfolding process (reviewed extensively by Baldwin, 1975).

Two studies using NMR techniques on the unfolding of staphylococcal nuclease induced by alkali (Jardetzky et al., 1971) and by acid (Epstein et al., 1971) indicated the existence of stable conformational intermediates and established the application of NMR techniques in studying unfolding transitions. Local conformational changes during the unfolding of ribonuclease have been monitored using high resolution $^1$H-NMR spectroscopy (Markley, 1975a,b; Markley and Finkenstadt,
1975; Benz and Roberts, 1975a,b). However the application of $^1$H-NMR to unfolding studies is limited due to the overlapping of the resonances from the numerous protons in the protein molecule and its most useful application to conformational studies has been with cyclic and straight chain oligopeptides as described previously in this chapter (section 2.B(ii)). This technique will become more useful as methods are developed for simplifying and interpreting the detailed spectra given by proteins.

A study by Nall and Baldwin (1977) failed to detect any intermediates in folding which had lost the ability to bind specific ligand, but which still had the physical properties of the native enzyme. The system used in this work was the specific binding of the ligand C$^1$-CMP to ribonuclease A in the pH induced unfolding of ribonuclease A; and the unfolding was followed by the change in U.V. absorbance due to three buried tyrosine residues in the protein.

It is often assumed that the pathways of folding and unfolding are the same so that one can study one to understand the other (the principle of microreversibility). On the basis of a set of experimental information gathered from published data by other workers using various techniques (including spectroscopic and optical rotatory changes, susceptibility to proteolytic attack and $^1$H-NMR studies) Burgess and Scheraga (1975b) developed a model for the pathway of thermally induced unfolding of ribonuclease at neutral pH which had six stages. (This scheme of unfolding was considered by Chen and Lord (1976) to be consistent with their laser Raman spectroscopic studies on the thermal folding of ribonuclease.) Three sections
of the ribonuclease chain were quite stable to thermal
denaturation each of which included a chain reversal which
Burgess and Scheraga suggested as the nucleation sites for the
folding of the native molecule. However, this study was on
ribonuclease with its S-S cross links still intact and may not
be relevant to the folding of a newly synthesised polypeptide.

Thus, it has now been established (Creighton, 1977a-d)
with the proteins pancreatic trypsin inhibitor and ribonuclease,
that the refolding reduced proteins do not form sequentially
the three disulphide bonds of the native state instead 'incorrect'
pairings of the cysteine residues take place which then inter­
change to eventually give the native structures. This process
seems to be essential to correct refolding of the molecule since
blocking of cysteine groups to prevent such 'incorrect' pairings
impairs the ability of the unfolded protein to regain its
native conformation. Unfolding of the native protein under
reducing conditions appears to take place by the reverse
pathway.

Observations on the thermal unfolding of a number of
small proteins, e.g. ribonuclease, cytochrome c, made by a
number of workers led Brandts et al., (1975) to propose a
model to explain the two kinetic phases that appeared to be
involved. They suggested that the initial fast phase is actually
due to the unfolding process and that the subsequent slow phase
is due to cis-trans isomerism about the peptide bond between
X-pro sequences. Using model compounds, e.g. L-alanyl-L-proline,
they demonstrated that the expected rate for isomerism is in
satisfactory agreement with the rates actually observed for
protein unfolding. Since a cis peptide bond in X-pro dipeptides leads to a chain reversal (a type VI \( \beta \)-bend, Lewis et al., 1973b) these could be nucleation sites for the reverse process.

Recently Matheson et al. (1977a) demonstrated the use of a non-specific labelling technique in the study of protein unfolding. A reactive aryl nitrene, generated by flash photolysis, attacks the exposed residues of the protein. Since the labelling is completed in 2 msec this technique could be used to detect the order in which residues in the protein chain were exposed during unfolding. Preliminary studies with spin labelling of different sites of the native ribonuclease molecule for subsequent electron spin resonance spectroscopy (Matheson et al., 1977b) also suggest that as the mobilities of the labels vary during the thermally induced unfolding of the protein this technique also offers great potential in the elucidation of the pathways by which proteins unfold.

3. HELIX \( \leftrightarrow \) COIL TRANSITIONS IN POLYPEPTIDES

The study of helix \( \leftrightarrow \) coil transitions of synthetic polyamino acids has provided a quantitative scale to describe the influence of each of the 20 naturally occurring amino acids on the stability of \( \alpha \)-helices in proteins. It has also provided information on the effect of the nature and size of amino acid side chains on helix-promoting ability and on the extent to which medium range interactions between neighbouring side chains are important in stabilizing secondary structure in proteins.
If short range interactions predominate in defining the secondary structure of a polypeptide then the tendency for an amino acid residue to be α-helical in a homopolymer composed only of that residue should reflect its likelihood of appearing in an α-helical conformation in a protein. To quantitatively specify this property Scheraga and coworkers have measured the Zimm-Bragg (1959) parameters $\sigma$ and $S$, which characterize the transition curve of the thermally induced helix-coil transition in polypeptides. Since proteins normally exist in an aqueous environment such transitions should be measured in water. However, the homopolymers of the naturally occurring amino acids are either insoluble in water, are non-helical in water at low temperature or do not randomise with increasing temperature. Thus, a technique for incorporating amino acid residues into random copolymers with a water soluble amino acid residue was developed (Von Dreele et al., 1971a,b). The non-ionic α-helical (in water) poly($N^5$-(3-hydroxypropyl)-L-glutamine) which will be referred to as polyHPG and poly($N^5$-(4-hydroxybutyl-L-glutamine) designated polyHBG were used as the 'hosts' for copolymers with each of the naturally occurring amino acid residues as the 'guest'. In this way, the Zimm-Bragg parameters of 15 of the 20 naturally occurring amino acids have been accurately determined, the last being threonine (Hecht et al., 1978).

The quantities $\sigma$ and $S$ are equilibrium constants for the elementary processes involved in nucleation ($\sigma$) and growth ($S$) of helical sequences (Poland and Scheraga, 1970). Thus, to propagate a helix the dihedral angles of one residue must be restricted to the α-helical conformation and only one hydrogen
bond must be formed; $S$ is the equilibrium constant for this process, thus:

$$\cdots\text{hhh}h\cdots \overset{S}{\longrightarrow} \cdots\text{hhhhhhh}\cdots$$

However, to initiate the formation of a helix in a non-helical region the dihedral angles of three consecutive residues must be restricted to those of an $\alpha$-helix and three hydrogen bonds need to be formed; $\sigma S^3$ is the equilibrium constant for this process:

$$\cdots\text{ccc}\cdots \overset{\sigma S^3}{\longrightarrow} \cdots\text{hhh}\cdots$$

Strictly speaking, the values of $\sigma$ and $S$ determined in aqueous solution indicate the tendency for an amino acid residue to adopt the $\alpha$-helical conformation during the initial folding of a randomly coiled polypeptide chain while the backbone is exposed to water and short range interactions are the only stabilizing forces for forming the $\alpha$-helix. However, helix probability profiles of native proteins calculated using the Zimm-Bragg parameters (plus assignments from other sources for the residues for which $\sigma$ and $S$ were not at that time available) show a close correlation with the actual occurrence of $\alpha$-helical regions in proteins (Lewis and Scheraga, 1971) indicating that during folding long and medium range interactions stabilize $\alpha$-helical regions after they have formed in the chain (Lewis et al., 1970).

Using helix$\leftrightarrow$coil transition theory and values for Zimm-Bragg parameters derived from a statistical analysis of 15 proteins of known amino acid sequence and crystal structure, Froimowitz and Fasman (1974) attempted to predict the helical
regions of several proteins. Overall they correctly predicted 79% of all residues and pointed out that with more accurate values for \( \sigma \) and \( S \) an improved prediction could be made.

Although the complete set of Zimm-Bragg parameters will facilitate the prediction of \( \alpha \)-helical regions in proteins, the importance of medium and long range interactions in stabilizing these regions should not be overlooked. The tendency of non-polar side chains to interact hydrophobically has been suggested as a stabilizing influence in \( \alpha \)-helices. This tendency is apparent from comparative studies on the helix-coil transitions of polymers of alanine and leucine (Ostroy et al., 1970) and of homopolymers with increasingly hydrophobic side chains (polyHEG < polyHPG < polyHBG) (Joubert et al., 1970). In addition, whereas the Zimm-Bragg parameters for charged glutamic acid indicate that it is essentially an indifferent helix promoter (Maxfield et al., 1975), its frequency of occurrence in \( \alpha \)-helices in the crystal structure of proteins is higher than for any other amino acid (Chou and Fasman, 1974a). This discrepancy has been explained on the basis of favourable medium range electrostatic interactions with an oppositely charged side chain four residues away. Such interactions would greatly enhance the probability of a glutamyl residue inducing helicity (Maxfield and Scheraga, 1975).

Valuable insights into the origins of the stabilizing energy of the \( \alpha \)-helix and the nature of the helix\( \leftrightarrow \)coil transition have been obtained by studies on short homopoly-peptides. From theoretical considerations, Schellman (1955) concluded that a polypeptide must reach a critical size before
it can exist in a helical configuration. This has been verified by ORD studies on lysine oligomers in water (Yaron et al., 1971) of chain length from 2 to 22. At pH 11.9 the onset of helicity occurred at a chain length of 12, which slowly increased to 13% for a chain length of 22. A gradual change with increasing chain length in the ORD curves of the oligomers was observed at pH 4.3, but even a chain length of 22 residues, at this pH, did not have an ORD curve identical to helical poly-L-lysine. Similar studies by Schechter et al. (1971) on a polytripeptide (L-Tyr-L-Ala-L-Glu)n when n = 1, 2, 3, 4, 7, 9 and 13 showed that oligomers (n = 1 to 9) had CD spectra resembling those obtained with random coils and that under physiological conditions (0.15 M sodium chloride - 0.02 M sodium phosphate, pH 7.4) only the polypeptide chain with n = 13 showed any evidence of helical content.

Early optical rotatory studies (Goodman et al., 1963; Goodman and Rosen, 1964) and later 1H-NMR studies (Goodman et al., 1969) suggested that in a series of carboxbenzoxo-γ-ethyl-L-glutamate oligomers where n = 2, 3, 4, 5, 6, 7, 8, 9 and 12, the conformation of the heptamer and higher oligomers was α-helical. However, subsequent work on these oligopeptides and on L-isoleucine oligopeptides using infra-red in both the solution and the solid state (Goodman et al., 1971a,b) showed that the preferred conformational state of oligomers of 7 residues or longer was a β-extended structure. This was later confirmed in a number of studies using NMR (Goodman et al., 1975a,b; Schwarz and Fasman, 1976) infra-red (Palumbo et al., 1976 and infra-red and CD (Kubota and Fasman, 1975; Toniolo
and Bonara, 1976) on homo-oligopeptides containing saturated straight chain aliphatic side chains such as alanine or valine and leucine, which demonstrated that pentamers and higher oligomers existed as inter- and intra-molecularly hydrogen-bonded $\beta$-structures and that lower oligomers were largely unstructured. Homo-oligomers of $\beta$-branched side chain amino acid residues also adopt $\beta$-structures and these are more stable than the equivalent straight chain oligomers (Toniolo and Bonara, 1975; Toniolo et al., in press). This finding is consistent with statistical analyses on protein structures which show that valine and isoleucine occur with the greatest frequencies in the $\beta$-regions of proteins (Chou and Fasman, 1974) and led Toniolo (1978) to suggest that valine and isoleucine should stabilize regular parallel $\beta$-pleated sheet structures by permitting inter-strand interactions between the $\beta$-branched side chains. In support of this hypothesis, doublets of the Val...Val, Val...Ile, Ile...Ile type occur frequently in parallel $\beta$-sheets in proteins and infrequently in anti-parallel $\beta$-sheets (Toniolo, 1978).

C. THE STATISTICAL MECHANICS OF RANDOMLY COILED POLYPEPTIDES

The average dimensions of an ideal homopolymer chain have been shown by random flight statistics to vary with the square root of the molecular weight (Flory, 1953). Polymer-solvent interactions and those between remote segments of the molecule tend to increase the average chain dimensions beyond the ideal value. However, choice of an appropriate solvent and temperature, to render the polymer barely soluble, so that contacts between atoms or groups that are near neighbours in
sequence along the chain predominate, reduces the average
dimensions to those of the unperturbed polymer chain (Flory,
1969). Under these conditions the polymer is said to be at the
\( \theta \)-point. Although a native protein in its normal aqueous
environment is unlikely to be at the \( \theta \)-point it has been
demonstrated that short-range interactions do predominate in
determining much of its conformational stability (Kotelchuck
and Scheraga, 1968, 1969; Finkelstein and Ptitsyn, 1971). The
conformational characteristics of the completely disordered
polypeptide chain, however, can be completely defined by the
application of random coil statistical mechanics since only
short range interactions determine its behaviour. The early
treatment of the randomly coiled polypeptide chain by Flory
and coworkers (Brant and Flory, 1965b; Miller et al., 1967;
Schimmel and Flory, 1967; Flory and Schimmel, 1967) has hardly
been improved on for straight chain homopolypeptides with
peptide bonds in the planar trans configuration and the methods
they developed for generating configurational averages in
macromolecules have been recently reviewed (Flory, 1974).
The most versatile method for the evaluation of the contributions
of individual configurations of the monomer units to the
overall spatial configuration of the linear polymer chain has
been the matrix generation technique (Flory, 1969). This
method is used in this thesis and described in Chapter II;
its application to polypeptide chains is discussed in Chapter IV
in the light of the results presented there for homo and
copolypeptides of varying side chain complexity. This review
will be restricted therefore to discussing advances in recent
years on branched polypeptide chains and to the use of Monte Carlo techniques for calculating the average chain dimensions of homopolypeptides where the matrix generation method is not the most appropriate one.

(i) Branched Polypeptides

The unperturbed dimensions of disordered linear polypeptides have been investigated experimentally and shown to conform to the rotational isomeric state representation of the configurational state of a polymer dominated by short range interactions (see Chapter IV of this thesis); this includes disordered native proteins whose chains were rendered linear by reduction with mercaptoethanol (Lapanje and Tanford, 1967) (this work is discussed later in Chapter IV section E). However, the configurational statistics of polypeptide chains that are denatured without breakage of cross linkages have only been investigated since the rotational isomeric state methods (Flory, 1974) of calculating the unperturbed dimensions of polypeptides have been extended so as to apply to macromolecules with branched side chains (Mattice, 1975, 1976; Mattice and Carpenter, 1976). This modified rotational isomeric state theory was used to calculate the unperturbed dimensions of homopolypeptides and sequential copolypeptides, each containing one disulphide bridge (Mattice, 1977a). The polypeptides that were the subject of the study were poly-L-alanine, poly-L-proline, poly(L-alanyl-D-alanine) and the collagen-like sequential copolypeptides poly(L-prolyl-L-prolyl-glycine), poly(L-prolyl-L-alanyl-glycine), poly(glycycl-L-alanyl-L-proline) and poly(L-alanyl-L-alanyl-glycine); the cross link was formed
by replacing the central amino acid residue in each chain by a L-cysteinyl residue and then cross linking two such residues on like chains. A property \( g \), defined as the ratio of the computed mean square unperturbed radius of gyration for branched polypeptides to that for linear polypeptides containing the same number of residues, was calculated for each polypeptide using the new modified rotational isomeric state theory for branched polymers and also using classical random flight statistics (Zimm and Stockmayer, 1949). For short chain lengths (<1,000) the estimates from each procedure varied considerably; for homopolypeptides the variation was a function of molecular weight and the type of monomer unit; for the collagen-like sequential copolypeptides the ratio \( g \) was slightly lower when calculated by the new method than by classical random flight statistics. In partially helical cross linked homopolypeptides (Mattice, 1978) \( g \) was found to vary strongly with helix content when helicity exceeded 20%. Higher degrees of polymerisation were required for \( g \) to achieve a limiting value in partially helical polypeptides than in completely disordered polypeptides.

A further study on completely disordered proteins (Mattice, 1977b) examined the conformational consequences resulting from the presence of a single intact inter-chain disulphide cross link in denatured proteins. Thirty naturally occurring proteins with chain lengths in the range 49-446 residues were cross linked by pairing each combination of cysteinyl residues in like chains (thus, a protein having two cysteinyl residues at positions \( P \) and \( q \) would be linked at positions \( P-q, P-P \) and \( q-q \), giving three branched molecules to consider). The characteristic ratios were calculated from
the mean square unperturbed radii of gyration and compared with those for the constituent linear proteins. There was a wide degree of overlap between the ranges of the two sets of values, but with one exception (thrombin A chain, n = 49) all the cross linked proteins had smaller average chain dimensions as expressed by this ratio. As with the study on polypeptides (Mattice, 1977a) the values for g obtained using the rotational isomeric state method were in general lower than those from classical random flight statistics. Included in the 30 proteins studied were four families of proteins with members whose sequences differed by only a few amino acid residues. The position and frequency of glycyl and prolyl residues was found to determine the variations in characteristic ratios within the members of these families. Thus, a reliable procedure now exists for describing the configurational statistics of branched macromolecules (Mattice, 1975, 1976, 1977a) in a similar way to that used by Flory (1974) in the treatment of linear macromolecules and has been applied to determining the unperturbed dimensions of cross linked proteins (Mattice, 1977b) and homopolypeptides both in the disordered state (Mattice, 1977a) and during helix $\leftrightarrow$ coil transitions (Mattice, 1978).

(ii) Monte Carlo Methods

The matrix generation method for calculating configurational averages of polypeptide chains has been the method of choice for most workers since it is simple to apply and requires a fairly modest amount of computer time. However, these advantages are the result of a simplified representation of the polypeptide backbone by a chain of virtual bonds of
fixed length between each $C_\alpha$ atom. This simplification is made possible by fixing all peptide bonds in the planar trans configuration so that the correlations between sets of rotational $(\phi, \psi)$ angles for neighbouring units can be ignored (Flory, 1969). For most polypeptides this assumption is justified (Zimmerman and Scheraga, 1976), but there is a wealth of evidence (Mandelkern, 1967; Bovey et al., 1968; Torchia and Bovey, 1971; Mattice and Mandelkern, 1971a; Howard et al., 1973; Dorman et al., 1973; Wütrich et al., 1974; Wu et al., 1975) that polypeptides containing imide bonds (i.e. those containing proline, hydroxyproline or N-methylated amino acids) are likely to include peptide bonds in the cis configuration. Although it is not impossible to apply Flory's matrix method to such situations, the system becomes more complex since the residues adjacent to any given residue $i$ can influence the configuration of $i$. One must therefore calculate statistical weight matrices for all permutations of sets of three residues rather than just for a single residue which adopts its configuration independently of neighbouring units when the peptide bond is in the trans configuration. It has been shown (Tanaka and Scheraga, 1975d) that applying the matrix method in this way requires too much computational effort and that the Monte Carlo method is the one choice in such cases. Two polyimino acids that have been treated using these methods are polysarcosine (Sisido et al., 1976) and polyproline (Tanaka and Scheraga, 1975d).

The fractions of cis peptide units occurring in varying chain lengths (up to 40) have been calculated for non-self-intersecting trans/cis polysarcosine (Sisido et al., 1976) and compared with the fractions measured by NMR spectroscopy.
in D$_2$O, CDC$_3$ and ethanol-d$_6$. The theoretically obtained results were far lower than the values obtained for polysarcosine in any of the three solvents. A theoretical estimate of 20% (for n = 30) of cis peptide bonds was found compared to 35-50% of the total peptide bonds in the cis configuration actually being present. Since the calculations are performed in the absence of solvent, presumably the cis form is in practice more stabilized than the trans form in all three solvents. Although characteristic ratios were calculated for trans polysarcosine in this work (Sisido et al., 1976) no estimates were given for polysarcosine containing cis peptide units.

Tanaka and Scheraga (1975) examined the effect on average chain dimensions of incorporating a small number of cis residues into a trans polyproline chain of 100 residues chain length. For the all-trans polypeptide chain they obtained characteristic ratios, averaged over four Monte Carlo simulations, of 208 which reduced to 139 and 43 when 5% and 10% respectively of the peptide bonds were in the cis configuration. (These authors calculated the characteristic ratio by an average bond length scheme; thus, their values are higher than those calculated using the usual (Flory, 1969) virtual bond scheme.) The all-cis chain had a characteristic ratio of 72 which is higher than the trans chain with a small number of cis units present. This is thought to be due to isolated cis peptide bonds introducing chain reversals into a chain of predominantly trans bonds. Their results for all-trans chains were in good agreement with a recent calculation
(Mattice et al., 1973) using the matrix generation technique. Experimental values obtained in a variety of solvents (Mattice and Mandelkern, 1971a) agreed with the dimension obtained for a trans chain including 5% of cis peptide bonds.

In the statistical mechanical treatment of polypeptide chain conformations the chain is considered in the unperturbed state and only short range interactions are considered in the calculations (Flory, 1969). Medium and long range interactions, however, may play an important role in determining the conformational flexibility of disordered polypeptides in an aqueous environment - where they are probably not under θ-conditions. Monte Carlo methods, which are computer simulations of self-avoiding (i.e. non-self-intersecting) lattice flights by a chain molecule are obviously applicable to the effect of excluded volume on polymer chains.

In a series of studies, R.A. Scott and coworkers have utilized Monte Carlo methods to calculate mean square end-to-end distances $<r^2>$ and radii of gyration $<S^2>$ for hard sphere models of polyamino acids (with chain lengths of up to 300 residues) for both unperturbed chains and chains perturbed by long range interactions (Knaell and Scott, 1971a,b; Warvari et al., 1971, 1972a,b,c; Warvari and Scott, 1972; Neves and Scott, 1975). Most of this work has involved refining their techniques and establishing that the hard sphere models used in their studies are appropriate for perturbed chain molecules (Neves and Scott, 1975). They assessed the effect of excluded volume on the chain dimensions of homopolypeptides by comparing the ratio of $<r^2>$ to $<S^2>$ in the unperturbed chain and the non-self-intersecting chain (Warvari et al., 1971, 1972a).
For short chain lengths (N) this ratio was found to pass through a maximum at $N = 15$ to $20$ for polyglycine and polyalanine, 12 to 15 for poly-N-methylalanine and 6 to 9 for polysarcosine, and then decreased slowly to a limiting value for large $N$ (in the range $N = 180$ to $270$) of 6.4-6.6 for the chain with excluded volume and 6.0-6.2 for the unperturbed chain. Their results are in good agreement with those of Tanaka and Nakajima (1972a) who obtained a $<r^2>/<s^2>$ ratio of 6.4 for the non-self-intersecting random flight of poly-L-alanine on a tetrahedral lattice compared to 6.0 for unperturbed chains of infinite chain length. These workers used a more complicated interaction energy potential that included contributions from electrostatic interactions and a Lennard-Jones type, as well as a hard sphere potential.

The model they developed was later employed in calculating the dimensional changes of polypeptide chains in the helix↔coil transition region using poly-L-alanine of chain length 151, as an example (Tanaka and Nakajima, 1972b). The Monte Carlo method in which checking of atomic overlaps was omitted was used in conjunction with the Zimm-Bragg (Zimm and Bragg, 1959) treatment of the helix↔coil transition. The ratio of the mean square end-to-end distance of the partially helical chain to that of the fully random chain (calculated including volume effects) was shown to increase by 50% from fully coiled to a chain containing a mean helical content of about 50% and then to increase rapidly to about 12 times its value for a fully helical chain.
The statistical mechanical behaviour of the randomly coiled polypeptide chain may not seem directly pertinent to the mechanism by which a protein attains the ordered molecular arrangement featured in the native state. Nevertheless, an understanding of the conformational characteristics peculiar to the random coil form are required as the first stage in any ab initio hypothesis which attempts to answer the question: "What makes proteins fold?"
CHAPTER II

THEORETICAL METHODS
1. NOMENCLATURE AND ABBREVIATIONS

The amino acid L-2-aminobutanoic acid has been given the trivial name β-methylalanine in this thesis. It is also known as α-amino-n-butyric acid (Ponnuswamy and Sasisekheran, 1970). L-2-aminoo-(3,3)-dimethylbutanoic acid, which has been referred to as tert-leucine by Pospisek and Blaha (1977) is designated β-methylvaline in this study. N(2-hydroxyethyl)-L-glutamine is frequently abbreviated to HEG. All other nomenclature used is that recommended by the IUPAC-IUB Commission (1970). The covalent structure of these residues together with the other amino acid derivatives that feature in this study are shown in Figure II.1. All amino acids considered are in the L-configuration.

2. CONFORMATIONAL ENERGY CALCULATIONS

Conformational energies of the N-acetyl-N'-methyl amide derivatives of the amino acid residues that are the subjects of this study were calculated using ECEPP (Empirical Conformational Energy Program for Peptides). This program was developed in the Department of Chemistry at Cornell University and is available from the Quantum Chemistry Program Exchange, Chemistry Department, Room 204, Indiana University, Bloomington, Indiana 47401; as program No. QCPE 286.
Fig. II.1 Covalent structure of the N-acetyl-N'-methyl amides of the amino acid residues studied in this thesis.
ECEPP was adapted by the author to be compatible with the Cyber 73 computing system at the University of Melbourne, and also modified to produce conformational energy maps. Sample input data, supplied with ECEPP, were used to verify that the program was functioning correctly after modification. Conformational energy contour maps of the N-acetyl-N'-methyl derivatives of glycine and alanine (see Figures II.2 and III.1) were produced and these were shown to be identical with maps of these molecules produced using ECEPP by Zimmerman et al. (1977a).

A. Generation of Conformations using the Program ECEPP

Contributions to the nonbonded, hydrogen bond, and electrostatic potential energies are computed for each atom pair in the polypeptide whose interatomic distance depends on one or more dihedral angles. This requires knowledge of the coordinates of all atoms in the polypeptide in a single Cartesian coordinate system, for each desired conformation. In this program, the standard residue data set supplies coordinates for each full residue, amino end group, or carboxyl end group. These coordinates correspond to a specific orientation for each of these three types in its individual local Cartesian coordinate system, with all dihedral angles set at 180° (π radians), following the IUPAC convention. A set of sub-programs generates the coordinates in the coordinate system of the whole polypeptide molecule by connecting the individual residues and end groups, while

1. i.e. Constructed by plotting energy values on a 2-dimensional map in which the torsion angles φ and ψ are the ordinate and abscissa, respectively (with values from -180° to +180°).
Fig. II.2  Conformational energy contour map of N-acetyl-N'-methylglycylamide. The contour lines are labelled with values for energies (in kcal/mole) above those at the minima indicated by filled circles.
adjusting each dihedral angle (if necessary) to the value specified by the user. When the relative energies of two or more conformations of a given peptide are computed, each conformation after the first is defined by the assignment of new values to one or more dihedral angles. The entire generation process is repeated for each such new conformation.

B. Empirical Potential Energy Functions

In the calculation of the conformational energy of a peptide by ECEPP all bond lengths and bond angles are assumed to remain constant and that the conformation is a function only of the variable dihedral angles. The program computes energy terms for only that part of the total molecular potential energy which varies with the torsion angles. This variable or conformational energy is treated as the sum of electrostatic, nonbonded, hydrogen-bonded and torsional contributions. An additional loop-closing potential is included if the peptide contains one or more intramolecular disulphide bonds, but this term was not applicable in this study.

The first three energy components are computed as the sum of terms for all atom pairs $i,j$ (including hydrogen atoms) separated by at least one degree of rotational freedom, i.e. atom pair whose interatomic distance is a function of one or more dihedral angles. The relations used to compute each type of energy are described below.
(i) **Electrostatic Energy (EES)**

The potential energy of each pairwise interaction is computed as a Coulomb potential between atom-centered monopole partial charges $q$:

$$U_{el} = 332.0 \frac{q_i q_j}{D r_{ij}}$$  

where $q_i$ and $q_j$ are in electronic charge units, $D$ is the dielectric constant, taken as 2.0 (McGuire et al., 1972), $r_{ij}$ is the internuclear distance in Angstrom units, and 332.0 is the factor needed to convert to energy units of kcal/mole. The standard residue data supplied with the program provide values of $q(332.0/2.0)^{1/2}$ for each atom. The values in the data set supplied with this program were estimated from the overlap normalized CNDO/2 partial charges (Momany et al., 1975). (Note that contributions to EES are computed for all atom pairs separated by at least one degree of rotational freedom, including the hydrogen atoms (H) and acceptor atoms (X) involved in hydrogen bonds. See Section 2.B(ii) for the special treatment of nonbonded interactions between such atoms.)

(ii) **Nonbonded Energy (ENB)**

In this program, a modified Lennard-Jones 6-12 potential (eq.2) is used to compute the nonbonded interaction energy of an atom pair $i,j$ (Momany et al., 1975).

$$U_{NB} = F_A \frac{k_l}{r_{ij}^{12}} - C_A \frac{k_l}{r_{ij}^6}$$  

The first term in eq.2 represents the repulsive energy between the atoms, while the second term represents the
attractive energies, such as dispersion. The coefficients $A^{kl}$ and $C^{kl}$ are assigned specific values for each combination of atom types $k$ and $l$. $F$ is assigned the value 0.5 for 1-4 interactions or 1.0 for 1-5 interactions, as defined below. An atom pair interaction is considered to be 1-4 when there is only one degree of rotational freedom, i.e. when the distance between the interacting atoms is a function of only one intervening dihedral angle. An example is the interaction between atoms $C'$ and $C^\beta$, which is dependent only on the dihedral angle $\phi$:

$$
\begin{array}{c}
i \\
C' \quad \quad \quad \quad C^\beta \\
N \quad \quad \quad C^\alpha \\
\hline \hline \\
i + 1 \quad \quad \quad i + 2 \\
i + 3 \\
\end{array}
$$

Interactions of this type have been termed "1-4", since the majority of them are between atoms separated by three bonds, i.e. atoms $i$ and $i + 3$ in a chain. When one or more intervening bonds are part of a ring, as in proline, the atoms involved in a "1-4" interaction may be separated by more than three bonds. All variable interactions which are not "1-4" are treated in the program as "1-5", even though in the molecule they may correspond to 1-6 or higher interactions.

The parameters $A^{kl}$ and $C^{kl}$ are related directly (Momany et al., 1974) to the depth of the energy minimum ($\varepsilon$) and the corresponding internuclear distance ($r_0$) as follows:

$$
A^{kl} = \varepsilon r_0^{12} \quad \quad C^{kl} = 2.0\varepsilon r_0^6
$$
Substitution of these relations in eq. 2 yields the form which is the basis of the actual computation (eq. 4).

\[ U_{\text{NB}} = \varepsilon (r_o/r_{ij})^{12} - 2.0\varepsilon (r_o/r_{ij})^{6} \] (4)

(iii) Hydrogen Bonded Energy (included in ENB)

The program treats as a hydrogen bond any interaction between designated donor and acceptor atoms. The donors (H) are amine, amide, hydroxyl or carboxylic acid hydrogens, and the acceptors (X) are uncharged ring nitrogens, amide nitrogens, or hydroxyl, ester, carbonyl or carboxylic acid oxygens. In these cases, the Lennard-Jones approximation (eq. 2) is replaced by a hydrogen bond potential of the form:

\[ U_{\text{HB}} = \alpha'_{\text{HX}/r_{\text{HX}}}^{12} - \beta_{\text{HX}/r_{\text{HX}}}^{10} \] (5)

where \( \alpha'_{\text{HX}} \) and \( \beta_{\text{HX}} \) are specific coefficients for the different combinations of donors and acceptors. (Note that eq. 5 is used in place of eq. 2 only for the interaction between H and X, and not for the interaction between X and the atom attached to H.) In terms of the depth of the energy minimum (\( E_{\text{min}} \)) and the corresponding internuclear distance (\( R_{\text{min}} \)), the coefficients are expressed as follows:

\[ \begin{align*} 
\alpha'_{\text{HX}} &= 5.0E_{\text{min}}R_{\text{min}}^{12} \\
\beta_{\text{HX}} &= 6.0E_{\text{min}}R_{\text{min}}^{10} 
\end{align*} \] (6)

For convenience in computation, the following substitutions are made:

\[ \begin{align*} 
E_{\text{min}} &= 4.29\varepsilon \\
R_{\text{min}} &= 0.775r_o 
\end{align*} \] (6a)
in order to reduce eq.5 to the form:

\[ U_{\text{HB}} = \varepsilon \left( \frac{r_0}{r_{\text{HX}}} \right)^{12} - 2.0\varepsilon \left( \frac{r_0}{r_{\text{HX}}} \right)^{10} \]  

(7)

Eq.7, thus, has coefficients formally similar to those in eq.4, but \( \varepsilon \) and \( r_0 \) are here defined differently (eq.6a).

(iv) Torsional Energy (ETOR)

Although the preceding terms are the major contributors to the conformational energy of a polypeptide, the energies obtained with these terms for small molecules do not always reproduce experimental rotational barriers. For this reason, additional terms are added to the conformational energy, in order to bring the computed and experimental rotational barriers into agreement. These terms are computed for all peptide bond (\( \omega \)) torsion angles, and for designated side-chain (\( \chi \)) and end group torsion angles, but not for \( \phi, \psi \) or analogs thereof (Momany et al., 1975). The form of the potential function used is:

\[ U_{\text{TOR}} = \left( \frac{U_0}{2} \right) (1 \pm \cos n\theta) \]  

(8)

where \( U_0 \) is the difference in kcal/mole between the experimental barrier height and that calculated from the electrostatic, nonbonded and hydrogen bond potential functions, \( \theta \) is the value of the dihedral angle, and \( n \) gives the symmetry of the barrier. ETOR is the sum of such contributions for all \( \omega \) angles and all designated \( \chi \) and end group angles, except those for cystine C-S and S-S bonds. The standard residue data designates which \( \chi \) and end group angles have torsional terms, and supplies specific values for \( U_0, n \) and the sign of the cosine term.
C. Data Input to ECEPP

As initial input ECEPP reads a data set containing standard residue information to be used in the subsequent generation of atomic coordinates and computation of relative potential energy. The Standard Residue Data supplied with the program consists of a basic data set for 26 amino acid residues and 20 end groups and a supplementary set of 10 amino acid residues. This includes all of the amino acids found in proteins and many end groups found in synthetic peptides. For the purpose of this study an abbreviated set of six residues and four end groups were used and these are shown in Appendix I. Since the amino acid residues β-methylalanine, β-methylvaline and hydroxyethylglutamine were not included in the Standard Residue Data supplied with ECEPP, properly formatted data sets for these residues were drawn up and are shown in Appendix I together with notes on the atomic numbering system used by ECEPP and a card by card description of this input data.

The geometry of these molecules was not available from direct crystallographic data and was adapted, together with partial electronic charges from the data available on the most closely related amino acids and other model compounds as follows.

\[
\text{N'-acetyl-N'}'-\text{methyl-β-methyl-alanyl-amide}
\]

The geometry of the lysine residue (Momany et al., 1975) was adopted for the coordinates of the β-methyl-alanyl residue to all atoms up to and including the C\(^\gamma\) atom, which left only the C\(^\gamma\) methyl hydrogen atoms to place. These were
placed, using analytical geometrical methods at the three corners at the base of a tetrahedron with the C$^\gamma$ atom at the apex and with one of the hydrogen atoms in the same plane as the C$^\alpha$-C$^\beta$-C$^\gamma$ atoms. When CART (a program designed to calculate the Cartesian coordinates of an atom relative to a frame of reference defined by atoms whose coordinates are known) became available from Cornell University, Chemistry Department, the coordinates of the terminal methyl group were checked.

Appropriate partial charges on the side chain atoms were selected from Table III of Momany et al. (1975). In this Table most of the aliphatic side chains had a charge of -0.075 on the terminal carbon atom and +0.025 on each of the three terminal hydrogen atoms. These values were adopted for the C$^\gamma$ atom and adjoined H atoms for $\alpha$-amino-n-butyric acid. Where the C$^\beta$ methylene group was situated in a straight section of hydrocarbon chain (e.g. in leucine or lysine, but not in valine or isoleucine) with no electro-negative or positive end groups close to it (e.g. as occur in methionine, cysteine, threonine and serine) the charge on the C$^\beta$ atom was taken as -0.03 and that on the adjoined hydrogen was taken as 0.015. This C$^\beta$ value was used for the $\beta$-methyl-alanyl residue, but values of 0.03 were taken for the adjoined H atoms to maintain a net charge of zero on the molecule. The values for partial charges for the backbone atoms were taken from the lysine data set.
N-acetyl-N'-methyl-β-methyl-valinyl-amide

The geometry of the valine residue (Momany et al., 1975) was adopted for β-methylvaline as far as the Cβ carbon atom.

The structure of the tert-butyl group of neopentane (CH₃)₃CCN determined by electron diffraction was reported by Livingston et al. (1960). The parameters of the molecule from this study were:

\[ C-C = 1.54 \pm 0.01 \text{ Å} \]
\[ \angle CCC = 109.5 \pm 1 \]

In addition, the x-ray crystal structure of N-t-butyl-5,6-dihydrodibenz (c,f) azocine (Hardy and Ahmed, 1974) gives the following dimensions for the tert-butyl group.

The tert-butyl group for β-methylvaline was therefore positioned with each Cγ carbon atom at the corner of a regular tetrahedron with the Cβ atom as the apex with:

\[ C-C = 1.53 \text{ Å} \]
\[ \angle CCC = 109.471° \]
The C\textsuperscript{Y} hydrogen atoms for each C\textsuperscript{Y} methyl group were also positioned at the corners of a regular tetrahedron with:

\[ \text{C-H} = 1.09 \text{Å} \]
\[ \angle \text{HCH} = 109.471° \]

The partial electronic charges for the atoms in this data set were taken from valine (Momany et al., 1974) except for the three C\textsuperscript{Y} atoms. These were adjusted to keep the molecule electrically neutral and so that each of these atoms had an equal charge.

\text{N-acetyl-N'-methyl-hydroxyethyl-glutamyl-amide}

The geometry and partial charges of glutamine were used for HEG up to the C\textsuperscript{Y} atom and the geometry and partial charges of serine were used from the C\textsuperscript{S} atom to the end of the side chain (Momany et al., 1975). For the side chain amide group (C\textsuperscript{S}-N\textsuperscript{ε} and adjoined substituents), the backbone peptide bond geometry (Momany et al., 1975) and the partial charges of N-methylacetamide (Momany et al., 1974) were used. Slight adjustments in the partial charges of substituents of the side chain were made to keep the molecule electrically neutral. The values of the partial charges, bond lengths and bond angles of the side chain of hydroxyethyl glutamine are listed in Table II.1. The coordinates of the side chain atoms were positioned using CART so that the molecule was in the correct starting conformation for use in ECEPP (i.e. all of the dihedral angles set at 180°). A molecular drawing of this conformation is shown in Figure II.3. It will be observed that the
### TABLE II.1

**PARTIAL CHARGES, BOND LENGTHS AND BOND ANGLES OF THE SIDE CHAIN OF HYDROXYETHYL-L-GLUTAMINE**

<table>
<thead>
<tr>
<th>Side-chain Atom</th>
<th>Partial Charge (e.s.u.)</th>
<th>Bond Length (Å)</th>
<th>Bond Angle (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$^\beta$</td>
<td>-0.030</td>
<td>C$^\alpha$-C$^\beta$ 1.53</td>
<td>C$^\gamma$C$^\alpha$C$^\beta$ 111.0</td>
</tr>
<tr>
<td>H ($\beta$)</td>
<td>0.020</td>
<td>C$^\beta$-H 1.0</td>
<td>N C$^\alpha$C$^\beta$ 111.0</td>
</tr>
<tr>
<td>C$^\gamma$</td>
<td>-0.110</td>
<td>C$^\beta$-C$^\gamma$ 1.53</td>
<td>C$^\alpha$C$^\beta$C$^\gamma$ 111.0</td>
</tr>
<tr>
<td>H ($\gamma$)</td>
<td>0.053</td>
<td>C$^\gamma$-H 1.0</td>
<td>C$^\beta$C$^\gamma$C$^\delta$ 111.0</td>
</tr>
<tr>
<td>C$^\delta$</td>
<td>0.465</td>
<td>C$^\gamma$-C$^\delta$ 1.53</td>
<td>C$^\gamma$C$^\delta$O 120.5</td>
</tr>
<tr>
<td>O ($\delta$)</td>
<td>-0.387</td>
<td>C$^\delta$=O 1.23</td>
<td>O C$^\delta$N$^\varepsilon$ 124.5</td>
</tr>
<tr>
<td>N$^\varepsilon$</td>
<td>-0.344</td>
<td>C$^\delta$-N$^\varepsilon$ 1.325</td>
<td>C$^\gamma$C$^\delta$N$^\varepsilon$ 115.0</td>
</tr>
<tr>
<td>H ($\varepsilon$)</td>
<td>0.164</td>
<td>N$^\varepsilon$-H 1.0</td>
<td>C$^\delta$N$^\varepsilon$H 124.0</td>
</tr>
<tr>
<td>C$^\zeta$</td>
<td>0.045</td>
<td>N$^\varepsilon$-C$^\zeta$ 1.453</td>
<td>C$^\delta$N$^\varepsilon$C$^\zeta$ 121.0</td>
</tr>
<tr>
<td>H ($\zeta$)</td>
<td>0.025</td>
<td>C$^\zeta$-H 1.0</td>
<td>H N$^\varepsilon$C$^\zeta$ 115.0</td>
</tr>
<tr>
<td>C$^\eta$</td>
<td>0.130</td>
<td>C$^\zeta$-C$^\eta$ 1.53</td>
<td>N$^\varepsilon$C$^\zeta$C$^\eta$ 111.1</td>
</tr>
<tr>
<td>H ($\eta$)</td>
<td>0.02</td>
<td>C$^\eta$-H 1.0</td>
<td>C$^\varepsilon$C$^\eta$O$^\theta$ 112.0</td>
</tr>
<tr>
<td>O$^\theta$</td>
<td>-0.310</td>
<td>C$^\eta$-O$^\theta$ 1.425</td>
<td>C$^\eta$O$^\theta$H 110.0</td>
</tr>
<tr>
<td>H ($\theta$)</td>
<td>0.170</td>
<td>O$^\theta$-H 1.0</td>
<td>N C$^\alpha$C$^\varepsilon$ 110.3</td>
</tr>
</tbody>
</table>
Fig. II.3  
$N'$-acetyl-$N'$-methyl amide of hydroxyethylglutamine shown with all dihedral angles $= 180^\circ$. 
side chain is not in the fully extended position in this conformation since the Sequence Rule (IUPAC-IUB Commission on Biochemical Nomenclature, 1970) defines the starting position of the principal torsion angle $\chi^3$ about the bond $\text{C}^\gamma$-$\text{C}^\delta$ to be when $N^c$ (rather than $\text{O}$) is trans to the $\text{C}^\beta$ in the side chain.

D. Selection of Rotameric States for Starting Conformations

The barrier to rotation about most aliphatic ($\text{SP}^3$ hybridised) C-C bonds has three-fold symmetry; this is also the case for C-O (hydroxyl) and C-N (amine) bonds (Momany et al., 1975). The values for the side chain torsion angles about these types of bonds at which this intrinsic torsional potential is minimal are $180^\circ$ and $\pm 60^\circ$. The configurations defined by the torsion angle adopting these values ($\pm 30^\circ$) are designated trans, gauche$^+$ and gauche$^-$ respectively. The Newman Projections of these rotameric states for valine are shown in Figure II.4. For side chain rotations involving carbon atoms with symmetric substitutions these values are equivalent. Thus, only one value ($60^\circ$) need be considered for $\chi^1$ for the alanyl and $\beta$-methyl-valinyl residue and $\chi^2$ for the valinyl, $\beta$-methyl-alanyl and $\beta$-methyl-valinyl residues. Starting values of $180^\circ$ or $\pm 60^\circ$ were used for all other side chain C-C and C-O rotations (i.e. $\chi^1$ for $\beta$-methylalanine and valine and $\chi^1$, $\chi^2$, $\chi^6$ and $\chi^7$ for HEG).

The angles of rotation around the peptide bonds ($\omega$) and the torsion angle of the side chain amide bond ($\chi^4$) of HEG, were kept at $180^\circ$ so that the substituent atoms to
Fig. II.4 Newman projections along the $C^\alpha$-$C^\beta$ bond for valine. $\chi'$ is the first variable side chain torsion angle, $g$ refers to the gauche conformations and $t$ to the trans conformations.
these bonds were kept in a planar trans configuration. This is energetically favored over the cis state by several kcal/mole (Momany et al., 1975).

The available experimental data (Yan et al., 1970; Warshel et al., 1970; Scott and Scheraga, 1966) and molecular orbital calculations (Yan et al., 1970) suggest that torsional barriers to rotation around $\text{C}'\text{N}-\text{CC}'$ and $\text{NC}-\text{C}'\text{N}$ bonds are small or non-existent and that energy fluctuations during rotations around these bonds are adequately represented by interatomic interactions. Thus, a six-state rotational isomeric model was used for the side chain torsion angles $\chi^3$ and $\chi^5$ of HEG which was chosen to maximize the distance between the adjoined substituents to these bonds, values for $\chi^3$ and $\chi^5$ being $\pm 30^\circ$, $\pm 90^\circ$ and $\pm 150^\circ$. The total number of side chain conformations thus generated for each backbone conformation by the permutations of the rotameric states used for each side chain bond was one for alanine and $\beta$-methylvaline; three for $\beta$-methylalanine and valine; and 2,916 for hydroxyethylglutamine.

The conformational space of the backbone of all of the residues in this study, except hydroxyethylglutamine, was described by using $10^\circ$ increments in $\phi$ and $\psi$, i.e. 1296 points on the $(\phi, \psi)$ plane. However, the large number of side chain conformations that needed to be considered for each $(\phi, \psi)$ pair precluded such a rigorous search for HEG and the conformational space for this molecule was covered using $30^\circ$ increments in $\phi$ and $\psi$ within those regions energetically allowed for the alanine residue, a total of 31 points on the $(\phi, \psi)$ plane.
E. Energy Minimization

Energy minimizations were performed using ECEPP in conjunction with a function-minimizing subroutine POWELL (M.J.D. Powell, 1964). ECEPP was modified to do this by replacing the call in the main program to the subroutine that usually controls the calculation of the total energy of the molecule (ENERGY) by a call to POWELL. This then calls ENERGY and successively increments the dihedral angles specified by the user until a minimum value of the total conformational energy (ETOT) is found. Minimization was terminated when the conformational energy changed by less than 0.01 kcal/mole between successive calculations.

The energy distribution over the total conformational space available to the backbone of each residue, except HEG, was calculated in 10° increments of φ and ψ, i.e. at 1296 points in the (φ,ψ) plane. The side chain torsion angles were first held fixed in the 'staggered' positions of their torsional minima. From the energy map thus produced, conformations in regions of the surface with energies within 6 kcal/mole of the global minimum were re-calculated allowing the side chain angles torsional freedom around these positions to minimize the overall conformational energy. For β-methylvaline only a small region of the (φ,ψ) plane had energies which were even within 10 kcal/mole of the global energy minimum; in this case, therefore, all of this region was re-calculated.
Because of the prodigious number of variable dihedral angles in the HEG residue energy minimizations were performed only on certain low energy conformations selected on the basis of criteria discussed in Chapter III, Section 2.E of this thesis.

F. **Conformational Probability Calculations**

The partition function, $Z$, for the entire conformational space of the molecule is the sum of the Boltzmann distributions of the conformational energy value for each degree of freedom in the molecule (Flory, 1974).

$$Z = \sum \sum \cdots \sum \exp\left(-\frac{E(\phi, \psi, \chi_1 \cdots \chi_n)}{RT}\right)$$

(9)

where $R$ is the gas constant and $T$, the temperature, was taken to be $293^\circ K$ throughout this study. The statistical weight $w(\phi, \psi)$ for each backbone conformation is:

$$w(\phi, \psi) = \sum \sum \cdots \sum \exp\left(-\frac{E(\chi_1 \cdots \chi_n)}{RT}\right)$$

(10)

Thus, the conformational probability $P(\phi, \psi) = w(\phi, \psi)/Z$ is the probability for the occurrence of each backbone conformation of the molecule.

The statistical weight $w(\chi_1 \cdots \chi_n)$ for each side chain conformation is given by:

$$w(\chi_1 \cdots \chi_n) = \sum \sum \exp\left(-\frac{E(\phi, \psi)}{RT}\right)$$

(11)

and the probability for the occurrence of each side chain conformation of the molecule is:

$$P(\chi_1 \cdots \chi_n) = \frac{w(\chi_1 \cdots \chi_n)}{Z}$$

(12)
The probability for the occurrence of each rotational isomeric state for each variable side chain torsion angle for the residue $P(x^i_k) = w(x^i_k)/Z$, where $x^i_k$ refers to the $k^{th}$ rotational isomeric state of the $i^{th}$ side chain torsion angles:

$$w(x^i_k) = \sum_{\phi \psi} \sum_{j \neq i} \frac{\exp(-E(\phi, \psi, x^j_n)/RT)}{n(n \neq k)}$$

G. Representations of the $(\phi, \psi)$ Conformational Space

Energy contour diagrams were plotted to represent the conformational space of the N-acetyl-N'-methylamides of glycine, alanine, valine, $\beta$-methylalanine and $\beta$-methylvaline. For molecules with non-symmetric substitution on the $\beta$ carbon atoms, namely, valine and $\beta$-methylalanine, three conformational energy diagrams were produced with $\chi^1$ corresponding to either the trans, gauche$^+$ or gauche$^-$ rotameric states. The most probable side chain conformation for the residue at each point on the $(\phi, \psi)$ plane is that which gives the molecule the lowest total conformational energy. Thus, the total conformational space available to the backbone can be represented by the lowest conformational energy value of the three rotamers at each point on the $(\phi, \psi)$ plane. Composite maps for valine and $\beta$-methylalanine were constructed in this way.

To represent the relative probabilities of various backbone conformations for the amino acid residue the values of $P(\phi, \psi)$ may be expressed as percentages of all backbone conformations and plotted on a $(\phi, \psi)$ diagram.
(Pullman and Pullman, 1974). This type of diagram is particularly useful for residues with side chain rotamer populations and is a more accurate expression of the contribution of each type of side chain population to the total probability of occurrence of a \((\phi, \psi)\) pair than the composite energy maps described above. Contour diagrams of this type were drawn for alanine, \(\beta\)-methylalanine, valine and \(\beta\)-methylvaline. However, for hydroxyethylglutamine the values were plotted but not contoured since the 30° grid size was too broad to allow good interpolation.

All contour diagrams were plotted using CONTOR - an automatic contouring program that takes randomly spaced data triplets and draws a contour map through them on a CALCOMP936 plotter. The interpolation phase of the program uses a weighted least-squares polynomial surface fitting procedure (McLain, 1974) and the author of the program is Ian J. Lilley (University of Melbourne). The program resides on public file in the program library of the University of Melbourne Computer Centre. This library program was modified slightly to plot data of this type by increasing its storage dimensions to take a larger number of data points than it normally handles and by reducing its requirements for computer time by by-passing the grid construction routines since these data are already spaced on a regular grid.
3. **CALCULATION OF CHARACTERISTIC RATIO \((C_{\infty})\) OF HOMOPOLYPEPTIDES**

Since P.J. Flory (1969) established the theoretical methods for calculating the dimensions of statistically coiling non-cooperative polypeptide chains most workers have accepted the approximations inherent in the methods. For this study it was thought necessary to establish the conditions under which these approximations are valid and to decide when more refined techniques are justified. The peptide geometry of Momany et al. (1975) was used to compare the effect of side chain length and branching on the unperturbed dimensions of homopolypeptides using the residues glycine, alanine, β-methylalanine, valine and β-methylvaline as examples. The effect of including all side chain atoms (including hydrogen atoms) when constructing the backbone conformational energy surfaces used to weight the unperturbed dimensions was explored. The effect of using 10° versus 30° grids on the \((\phi, \psi)\) maps for weighting the energy surfaces was also compared. Different methods of weighting with multiple side chain conformations were used and the effect of improving the treatment of flexibility in side chain torsion angles by including minimized energy values in the weighting procedure was examined. The results obtained are compared in Chapter IV with those obtained in earlier work using less refined techniques, and are used to compute the chain dimensions of polyHEG and copolymers containing this residue described in that chapter.
A. Generation of Random Polypeptide Chains

The unperturbed dimensions of homopolypeptides were calculated as the mean square end-to-end distance \( \langle r^2 \rangle_0 \) and expressed as the characteristic ratio \( (C_\infty) \) as the chain length \( (n) \) tends to infinity, by the method of Brant and Flory (1965b):

\[
C_\infty = \frac{\langle r^2 \rangle_0 / n}{\langle r_p^2 \rangle_0} = \left[ (E + \langle T \rangle)(E - \langle T \rangle)^{-1} - \frac{2}{n} \langle T \rangle \right]_{1,1} \quad (14)
\]

where \( E \) is the identity matrix of order 3 and \( \langle T \rangle \) is given by:

\[
\langle T \rangle = z^{-1} \sum_{\phi, \psi} \Gamma(\phi, \psi) \exp\left(-E(\phi, \psi)/RT\right) \quad (15)
\]

where \( z \) is the partition function for the conformational space available to the peptide residue:

\[
z = \sum_{\phi, \psi} \exp\left(-E(\phi, \psi)/RT\right) \quad (16)
\]

This expression for \( z \) does not take into account side chain rotational isomeric states.

\( \Gamma(\phi, \psi) \) is the matrix that transforms the coordinate systems of the \( (i + 1) \)th residue to that of the \( i \)th residue and is the product of a series of matrices for five separate rotations, three of which are fixed by the fixed geometry of the peptide residue and two which are variable and dependent on \( \phi \) and \( \psi \), thus:

\[
\Gamma(\phi, \psi) = \Gamma_\phi \cdot \Gamma_{\psi i+1} \cdot \Gamma_\theta \cdot \Gamma_{\psi i+1} \cdot \Gamma_\eta \quad (17)
\]
The angles that define the peptide geometry of these transformations were calculated from the geometric parameters of Momany et al. (1975) and have the values $\zeta = 14.9^\circ$, $\theta = 70.7^\circ$ and $\eta = 20.9^\circ$ (see Figure II.5).

$E(\phi,\psi)$ is the conformational energy of the residue for each backbone conformation defined by the $(\phi,\psi)$ pair and $R$ is the gas constant. $T$, the temperature, was taken to be 293$^\circ$K in this study.

B. Energetic Statistical Weighting of $C_\infty$

Evaluations of the partition function $Z$ and the average transformation matrix $<T>$ were carried out at equal intervals of $\phi$ and $\psi$ throughout their ranges. Two intervals were tested, namely $30^\circ$ and $10^\circ$ for each residue and two sets of $(\phi,\psi)$ conformational energies, one in which $\chi^1$ (the first side chain torsion angle) was held fixed at the torsional minimum and one in which it was allowed to vary to adopt the value that minimized the overall conformational energy.

For those molecules with three side chain conformations determined by $\chi^1$, $<T>$ and $C_\infty$ were calculated for each rotamer and for the overall molecule. Two methods were used to weight $<T>$ to take into account the contributions from all three side chain conformations:

(a) The most probable side chain conformation for the residue at each point on the $(\phi,\psi)$ plane is that which gives the molecule the lowest total conformational energy. Thus, at each point on the $(\phi,\psi)$ surface $<T>$ was weighted using the Boltzmann factor for the lowest of the three energy values of the three rotameric states of the side chain. This will be
Fig. II.5 Peptide backbone geometry, angles are shown in degrees and distances are in Angstroms.
designated as composite weighting, since conformational energy maps constructed by this procedure have been referred to as composite maps by Paterson and Leach (1978) and by Miller and Goebel (1967). This method of weighting neglects entropic considerations of the side chains, since each \((\phi, \psi)\) pair is counted only once although there are values of \(\phi\) and \(\psi\) for which more than one side chain rotamer gives energetically allowed backbone conformations. Such backbone conformations are statistically more probable than those \((\phi, \psi)\) pairs where only one rotamer is allowed. To overcome this problem a method of statistically weighting \(T(\phi, \psi)\) was used as follows:

(b) This approach considers that the polyamino acid is a random co-polymer of the residue in three rotational isomeric states. The average transformation matrix for a co-polymer is given by P.J. Flory (1969) as:

\[
<T> = \frac{1}{Z} \sum_{k=1}^{n} P(x_k^1) T_k^z
\]  

where \(P(x_k^1)\) is the probability of a residue in the \(k^{th}\) rotameric state occurring in the polymer and \(T_k^z\) is the transformation matrix of the residue in that state.

The probability of the occurrence of each rotameric state is dependent on the statistical weight \(w(x_k^1)\) of each rotamer. Thus:

\[
P(x_k^1) = \frac{w(x_k^1)}{Z}
\]

where

\[
w(x_k^1) = \sum_{\phi, \psi} \exp(-E(\phi, \psi, x_k^1)/RT)
\]
Eq. 18 is mathematically equivalent (see Appendix II for proof) to statistically weighting the average transformation matrix over all side chain rotational isomeric states for each \((\phi, \psi)\) pair using the Boltzmann factors for all three rotamers:

\[
<T> = Z^{-1} \sum_{\phi \psi} \sum_{k=1}^{3} \exp(-E(\phi, \psi, \chi_k)/RT)
\]

where \(\chi_k\) refers to the \(k^{th}\) rotameric state (i.e. trans, gauche\(^+\) or gauche\(^-\)) of the first side chain torsion angle, and \(Z\) in this case includes the Boltzmann factors for all rotational isomeric states.

\[
Z = \sum_{i=1}^{36} \sum_{j=1}^{36} \sum_{k=1}^{3} \exp(-E(\phi_i, \psi_j, \chi_k)/RT)
\]

For polyhydroxyethylglutamine, with multiple side chain rotational isomeric states, eq. 21 is extended to include these to give eq. 23.

\[
<T> = Z^{-1} \sum_{\phi \psi} \sum_{i=1}^{7} \sum_{k=1}^{n} \exp(-E(\phi, \psi, \chi_i)/RT)
\]

Where \(n\) is the number of rotational isomeric states chosen for the side chain torsion angle \(\chi_i\).

The numerical integration interval used for the evaluation of the \((\phi, \psi)\) space available to this molecule was 30° within the space allowed energetically for alanine. Although a more refined interval would have been more accurate, the comparison of different methods of energetic statistical weighting described in this thesis suggests that this interval is adequate for homopolypeptides with "alanine-like" characteristic ratios of 8 ± 2. Since the
characteristic ratio of polyhydroxyethylglutamine using this integration interval fell into this range, further refinements in the description of the backbone conformational space were not thought to be necessary.

C. Unperturbed Dimensions of Sequential Copolypeptides

The characteristic ratio of a sequential copolypeptide $[M_a^a M_b^b M_c^c]^n$ where $n = n_p/3$, is given by eq. 24 (Miller et al., 1967):

$$<r^2>/n_p l_p^2 = 1 + (2/n_p l_p^2)[1,0,0,0,0]$$

$$X[G_{Mb}G_{Mc}G_{Ma}]^{1/3}(n_p+1)-1XG_{Mb}$$

Where the vector $l_p = (l_p, 0, 0)$ and $G_{Mx}$ is a composite matrix of order 5 x 5 constructed from the average transformation matrix $<_{x}T>$ for each type of residue in the polymer, thus

$$G_{Mx} = \begin{bmatrix}
1 & l_p^{T}<_{x}T & 0 \\
0 & <_{x}T & l_p \\
0 & 0 & 1
\end{bmatrix}$$

where $O$ is the zero matrix of order 3 and $l_p^{T}$ is the transpose of $l_p$.

The average transformation matrix $<_{x}T>$ for hydroxyethylglutamine was evaluated from eq. 23. The matrices for glycine, alanine and $\beta$-methylalanine were evaluated at $10^\circ$ intervals over $\phi$ and $\psi$. In the case of alanine and $\beta$-methylalanine, side chain angles were allowed some freedom to move from the intrinsic torsional minima in order to
minimize the overall conformational energy and the transformation matrix for β-methylalanine was averaged over all rotational isomeric states as described above.

The characteristic ratio for the copolypeptides \([\text{HEG-Gly-Gly}]_n\), \([\text{HEG-Ala-Ala}]_n\) and \([\text{HEG-β-methylAla-β-methylAla}]_n\) were calculated from eq. 24 and eq. 25 for increasing values of \(n\) until a limiting value was reached. To save computer time, the matrix product \((G_{mb}G_{mc}G_{ma})\) was computed and then successively squared until the desired power for \(n\) was reached. Thus if \(S = \text{number of times the matrix is squared}:\)

\[
2^S = 1/3(n_p + 1) - 1
\]

and \(n_p = 3(2^S + 1) - 1\)

Hence, if e.g. \(n = 1000\), then \(1/3(n_p + 1) - 1\) requires 333 operations, but \(S = \log_2 333 = 8\) operations.

D. Computer Programs for Calculating \(C_\infty\)

Programs to calculate the characteristic ratios of homo- and co-polypeptides were written in Fortran Extended Language (version 4.0) for the Control Data, Cyber 73 computer. The programs were written so that the characteristic ratios were weighted either with conformational energy spaces or with previously constructed conformational probability spaces (e.g. for polyhydroxyethyl glutamine). A flow chart of the logic of the program that weights the characteristic ratio for each monomer type with one energy space only
(e.g. for alanine, glycine or \(\beta\)-methylvaline) and which will calculate the characteristic ratio for both homopolypeptides and copolypeptides composed of these monomers is shown in Figure II.6, and a listing of the program is given in Appendix III.

This program was tested by calculating the characteristic ratio of a freely rotating peptide chain (i.e. with no energetic weighting) using the peptide geometry of V. Sasisekheran (1962) and which was used also by P.J. Flory (1969) (i.e. \(\zeta = 13.2^\circ\), \(\theta = 70.0^\circ\), \(\eta = 22.2^\circ\)). The result obtained was the same as that of Flory, namely 1.93. The values for the characteristic ratios of polyglycine and polyalanine using the peptide geometry of V. Sasisekheran were also sufficiently close to those of Flory to assume that the program was functioning correctly (they were not, of course, identical since the conformational energy surfaces used to weight the characteristic ratios differed from those used by Flory).

The sub-routine for calculating the characteristic ratios of copolytripeptides (CRCOPOL) was tested using the published values of Miller et al. (1967) for the elements of \(<\bar{T}>\) matrices for polyglycine and polyalanine:

\[
<T>_{\text{gly}} = \begin{bmatrix}
0.33 & -0.11 & 0 \\
-0.16 & -0.34 & 0 \\
0 & 0 & -0.11
\end{bmatrix}
\]

\[
<T>_{\text{ala}} = \begin{bmatrix}
0.49 & 0.19 & 0.60 \\
-0.10 & -0.58 & 0.22 \\
0.66 & -0.27 & -0.27
\end{bmatrix}
\]
Subroutine Coord.
Calculates rotation matrices Tζ, Tθ and Tη from peptide bond angles ζ, θ and η.

Do main loop I = 1, NJ

Read descriptive title for each calculation of <T>
Read energy values for (φ,ψ) space ETOT(φ,ψ) used to weight the <T> being calculated
Truncate energy values >20 kcal/mole to avoid overflow

<T> = 0.0
T(φ,ψ) = Tζ . Tθ . Tη

<T> = <T> + T(φ,ψ) exp(-ETOT(φ,ψ)/RT)

Has <T> been weighted for all (φ,ψ) pairs?

NO

Calculate partition function Z
Z = Z + exp(-ETOT(φ,ψ)/RT)

<T> = <T>/Z

Write Title, <T> and Z

Do you wish to calculate the CR for the homopolypeptides of each monomer type?

NO

CALL CRMONO
This calculates the CR from <T> and the identity matrix E

YES

Have all <T> been calculated for each monomer type, i.e. I = NJ?

NO

FIG.II.6
Do you wish to calculate the CR for the copolytripeptides?

END ← NO → YES

CALL CRCOPOL
This calculates the CR of the copolymer from composite matrices GM derived from the <T> of each monomer unit.

Read NR = \log_2 \frac{1}{3} (n + 1) - 1
NR is the number of times the matrix product is squared
Read NS the increments of NR for which you want to know the values of the CR

Construct GM_1 from <T>_1
GM_2 from <T>_2
GM_3 from <T>_3

<z>_1 = GM_1 \cdot \langle \rangle
<z>_2 = GM_1 \cdot GM_2
<z>_3 = T_3 \cdot GM_3

Square the matrix product

I = 1, NR

<z>_3 = <z>_2 \cdot <z>_3

NO

Is I = NS

NO

Is I = NR

<z>_4 = 1 \cdot <z>_3 \cdot <z>_1

n_p = 3(I + 1) - 1

\text{CR} = 1 + \frac{(2/n_p)^2}{T_4} \rightarrow \text{WRITE } n_p \text{ and CR } \rightarrow \text{RETURN}

\begin{align*}
T_1 &= <T>(NSTEP-1) \\
T_2 &= E - <T> \\
T_3 &= T_2^{-2} \\
T_4 &= E - T_1 \\
T_5 &= \left(\frac{2}{NSTEP}\right) \cdot <T> \cdot T_4 \cdot T_3 \\
T_6 &= T_2^{-1} \\
T_7 &= E + <T> \\
T_{\text{FIN}} &= T_7 \cdot T_6 + T_5 \\
\text{CR} &= \frac{T_{\text{FIN}}(1,1)}{<z>_4} \\
\end{align*}

Is NSTEP > NRES, NO

Write $T_{\text{FIN}}$ and $T_{\text{FIN}}(1,1)$

\text{Fig. II.6}  Flow Diagram of Program for Calculating Characteristic Ratios
\( <G>_{\text{ala}} \) and \( <G>_{\text{gly}} \) matrices were constructed by CRCOPOL and used to calculate the \( C_\infty \) of polyglycine and polyalanine considering them to be poly(\( M_aM_aM_a \)). The values obtained, namely, 2.0 for polyglycine and 9.1 for polyalanine were the same as those of Miller et al. (1967). Unfortunately, there are no published results on the computed chain dimensions of sequential copolymers that could be used to verify the method used in this work. However, Miller et al. (1967) published values of the characteristic ratios of random copolymers containing glycine and alanine. The \( C_\infty \) of poly(AlaAlaGly) and poly(GlyGlyAla) were calculated with CRCOPOL using the \( <T> \) matrices of Miller et al. and were 3.3 and 2.3, respectively. These values are consistent with the results for random copolymers, namely, 3.7 for 33% glycine, 67% alanine and 2.4 for 67% glycine, 33% alanine, published in the study of Miller et al.
CHAPTER III

THE EFFECT OF SIDE CHAIN COMPLEXITY ON THE ENERGETICALLY PREFERRED CONFORMATIONS OF AMINO ACID RESIDUES
1. **INTRODUCTION**

A central tenet in the prediction of the conformations of larger peptides is the belief that the conformational space available to the peptide backbone immediately adjacent to any residue is largely governed by the nature and conformations of the latter's side chain. Thus, there has been increasing interest in the experimental conformational analysis of the side chains of biologically active peptides which has been facilitated by the application of high resolution techniques such as NMR (Jones et al., 1976; Gibbons et al., 1972). Conformational energy calculations have been used to predict the most probable conformation of a peptide where there is more than one conformation consistent with the experimental data (Isogai et al., 1977; Gibbons et al., 1970).

The number of possible side chain conformations for most naturally occurring amino acids is sufficiently large to preclude a total search of the conformational space available to the backbone, exceptions being glycine, alanine and proline. Attention has therefore been focused on establishing their minimum energy conformations (Lewis et al., 1973a; Zimmerman et al., 1977a). However, amino acids with short side chains, branched at the C₆ atom, are small enough to allow an exhaustive description of the total space available to such amino acid residues. The side chain conformations of these amino acids are uniquely defined by the value of the variable side chain torsion angle $\chi^1$. Valine and threonine are the only amino acids that occur in
proteins for which the above simplifications apply. However, β-methylalanine, an amino acid that is frequently used in synthetic analogs of biologically-active peptides (Jorgensen et al., 1971; Wieland et al., 1971) can also be treated in this way. Substituting a third methyl group onto the Cβ carbon atom of valine, to give β-methylvaline, gives a three-fold symmetric side chain for which only one value of χ1 need be considered, namely that in which the three Cγ methyl groups are staggered with respect to the backbone bonds.

β-Methylvaline could be an interesting model for steric restrictions imposed by bulky side chains on the backbone conformations of the polypeptide chain. Residues which impose well defined local backbone conformations would be very useful to engineer the conformations of synthetic peptides. One such residue is α-amino-iso-butyric acid (Burgess and Leach, 1973) that occurs in cyclic peptide antibiotics and is branched at the α carbon atom. A tert-butyl group at the Cβ atom might also be supposed to restrict the values of φ and ψ adjacent to it, thus an exploration was undertaken of the conformational space available to this residue.

Hydroxyethylglutamine was included in this study because of its properties as a water solubilizing residue in randomly coiled polypeptides. These are important for the work reported in Chapter IV. It is also interesting as a model for comparing the effects of lengthy side chains on the (φ,ψ) space and minimal energy conformations available to amino acid residues.
2. RESULTS

A. Minimum Energy Conformations for the N-acetyl-N'-methyl amides of \( \beta \)-methylalanine and \( \beta \)-methylvaline

The conformational energy minima of the \( \beta \)-methylalanine and \( \beta \)-methylvaline residues are listed in Tables III.1 and III.2, respectively. Since the object of this study was to explore the effect on the flexibility of the backbone of increasingly 'bulky' and 'lengthy' side chains, the entire \((\phi, \psi)\) conformational space was searched as discrete 10° increments with variations only in side chain torsion angles \(\chi^1\) and \(\chi^2\) to achieve energy minimization. These 'minima', therefore, are minima only to the nearest 10° in \(\phi\) and \(\psi\). Conformational energy minima for alanine and valine have been reported by other workers (Lewis et al., 1973a; Zimmerman et al., 1977a; Miller and Goebel, 1968) and the values obtained in the present study are compared with those in section 3.A.

In addition to the energy and dihedral angle values, the Tables show the position of each minimum in the \((\phi, \psi)\) plane according to the conformational letter code devised by Zimmerman et al. (1977a).

B. Conformational Energy Contour Diagrams

The energy contour maps of the alanine and \( \beta \)-methylalanine residues are shown in Figure III.1(a) and (b), respectively, and of valine and \( \beta \)-methylvaline in Figure III.2(a) and (b), respectively. The maps for \( \beta \)-methylalanine and valine are composite maps constructed as described in Chapter II. The maps for the individual
<table>
<thead>
<tr>
<th>Confl. Letter Code</th>
<th>( \phi )</th>
<th>( \psi )</th>
<th>( \chi^1 )</th>
<th>( \chi^2 )</th>
<th>( \Delta E ) (kcal/mole) ( (E_\circ = -2.476) )</th>
<th>Torsion Angles</th>
<th>( \Delta E ) (kcal/mole) ( (E_\circ = -3.116) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-80</td>
<td>90</td>
<td>-173</td>
<td>57</td>
<td>0</td>
<td>-80 80 61</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>-90</td>
<td>80</td>
<td>-66</td>
<td>63</td>
<td>0.302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>-150</td>
<td>130</td>
<td>-175</td>
<td>55</td>
<td>0.503</td>
<td>-150 150 60</td>
<td>0.351</td>
</tr>
<tr>
<td>A</td>
<td>-70</td>
<td>-50</td>
<td>-175</td>
<td>56</td>
<td>0.594</td>
<td>-70 -50 60</td>
<td>1.143</td>
</tr>
<tr>
<td>E</td>
<td>-150</td>
<td>160</td>
<td>63</td>
<td>59</td>
<td>0.750</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>-150</td>
<td>-60</td>
<td>-178</td>
<td>55</td>
<td>1.052</td>
<td>-160 -60 52</td>
<td>1.544</td>
</tr>
<tr>
<td>A</td>
<td>-80</td>
<td>-50</td>
<td>-67</td>
<td>62</td>
<td>1.118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>-130</td>
<td>150</td>
<td>-67</td>
<td>65</td>
<td>1.295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-150</td>
<td>40</td>
<td>55</td>
<td>53</td>
<td>2.305</td>
<td>-150 70 59</td>
<td>0.648</td>
</tr>
<tr>
<td>A*</td>
<td>60</td>
<td>70</td>
<td>-165</td>
<td>60</td>
<td>2.454</td>
<td>50 60 63</td>
<td>2.442</td>
</tr>
<tr>
<td>A*</td>
<td>60</td>
<td>60</td>
<td>-61</td>
<td>65</td>
<td>2.582</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>-60</td>
<td>150</td>
<td>73</td>
<td>67</td>
<td>2.604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F*</td>
<td>60</td>
<td>-180</td>
<td>-60a</td>
<td>60a</td>
<td>5.724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-70</td>
<td>-30</td>
<td>60a</td>
<td>60a</td>
<td>7.834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*</td>
<td>70</td>
<td>-80</td>
<td>-60a</td>
<td>60a</td>
<td>8.031</td>
<td>70 -70 60a</td>
<td>12.315</td>
</tr>
<tr>
<td>C*</td>
<td>70</td>
<td>-80</td>
<td>180a</td>
<td>60a</td>
<td>8.055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A*</td>
<td>40</td>
<td>50</td>
<td>60a</td>
<td>60a</td>
<td>12.423</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. All conformations listed except these five have been energy minimized with respect to side chain torsion angles, maintaining \( \phi \) and \( \psi \) at the values shown.
TABLE III.2 LOW ENERGY CONFORMATIONS DERIVED FROM ENERGY CONTOUR MAPS OF THE N-ACETYL-
N'-METHYLANIDE DERIVATIVES OF VALINE AND β-METHYLVAlINE

<table>
<thead>
<tr>
<th>Confl. Letter Code</th>
<th>Valine</th>
<th>B-Methylvaline</th>
<th>ΔE (kcal/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Torsion Angles</td>
<td></td>
<td>(E_o = -1.790)</td>
</tr>
<tr>
<td></td>
<td>φ</td>
<td>ψ</td>
<td>χ²</td>
</tr>
<tr>
<td>C</td>
<td>-90</td>
<td>100</td>
<td>178</td>
</tr>
<tr>
<td>E</td>
<td>-150</td>
<td>140</td>
<td>66</td>
</tr>
<tr>
<td>A</td>
<td>-80</td>
<td>-50</td>
<td>174</td>
</tr>
<tr>
<td>E</td>
<td>-140</td>
<td>160</td>
<td>-67</td>
</tr>
<tr>
<td>F</td>
<td>-60</td>
<td>140</td>
<td>73</td>
</tr>
<tr>
<td>D</td>
<td>-130</td>
<td>30</td>
<td>-67</td>
</tr>
<tr>
<td>A*</td>
<td>60</td>
<td>80</td>
<td>175</td>
</tr>
<tr>
<td>A</td>
<td>-60</td>
<td>-30</td>
<td>60^a</td>
</tr>
<tr>
<td>A</td>
<td>-70</td>
<td>-30</td>
<td>60^a</td>
</tr>
<tr>
<td>C*</td>
<td>70</td>
<td>-70</td>
<td>180^a</td>
</tr>
</tbody>
</table>

a. All conformations listed except these three have been energy minimized with respect to side chain torsion angles maintaining φ and ψ at the values shown.
Fig. III.1 Conformational energy contour maps of the N-acetyl-N'-methylamide derivatives of (a) alanine and (b) β-methylalanine. The contour lines are labelled with energies in kcal/mole above the minimum energy points of \((\phi, \psi) = (-80, 80)\) for alanine and \((\phi, \psi) = (-80, 90)\) for β-methylalanine. The map for β-methylalanine is constructed from the lowest of the energy values of its three rotamers at each \((\phi, \psi)\) point. Locations of the minima listed in Table III.1 are indicated by the filled circles. The regions defining the conformation letter code are marked on the alanine map.
Fig. III.1(a)
Fig. III.1(b)

ψ (degrees)

φ (degrees)

β-METHYL ALANINE
Fig.III.2  Conformational energy contour maps of the N-acetyl-N'-methylamide derivatives of (a) valine and (b) β-methylvaline. The contour lines are labelled with energies in kcal/mole above the minimum energy points of \((\phi,\psi) = (-90,100)\) for valine and \((\phi,\psi) = (-130,140)\) for β-methylvaline. The map for valine is constructed from the lowest of the energy values of its three rotamers at each \((\phi,\psi)\) point. Locations of the minima listed in Table III.2 are indicated by the filled circles.
Fig. III.2(a)
Fig. III.2 (b)
trans, gauche\(^+\) and gauche\(^-\) rotamers for these two residues are shown in Figures III.3, III.4 and III.5. The positions of the minimum energy conformations, and the values of the energy contours are marked in kcal/mole, with respect to a global minimum normalized to zero.

C. Rotamer Populations for the N-acetyl-N'-methylamides of \(\beta\)-methylalanine and Valine

The probability of the occurrence of each side chain rotamer summed over the entire conformational space available to the backbone for each of the two types of residues is shown in Table III.3, together with its statistical weight. Probabilities are expressed as percentages.

D. Conformational Probability Contour Diagrams

The conformational probability diagrams for the alanine, \(\beta\)-methylalanine, valine and \(\beta\)-methylvaline residues are shown in Figure III.6(a) and (b). Since there are no conformations for any of these residues with a probability of occurrence greater than 0.1\% for \(\phi > 0^\circ\), only the areas of the diagrams for \(\phi\) in the range \(-180^\circ\) to \(0^\circ\) are shown. The total probability (%) for each conformational region (using the designations of Zimmerman et al. (1977) is shown to the nearest whole number in Table III.4, for those regions with probabilities of at least one per cent.

The maxima on the probability diagrams for valine and \(\beta\)-methylvaline in Figure III.6(a) and (b) are compared, in Table III.5, with the energy values of the conformational minima at those points.
Conformational energy contour diagrams for the N-acetyl-N'-methy lamide derivatives of \( \beta \)-methylalanine and valine, in the trans rotameric state. Locations of minima are indicated by filled circles. The contour lines are labelled with energy in kcal/mole above the minimum energy points at \((\phi, \psi) = (-80, 90)\) for \( \beta \)-methylalanine and \((\phi, \psi) = (-90, 100)\) for valine.
Fig. III.3

\( \psi \) (degrees)

\( \phi \) (degrees)

\( \beta \)-METHYLALANINE
Fig. III.3 (cont'd.)
Fig.III.4 Conformational energy contour diagrams for the N-acetyl-N'-methylamide derivatives of β-methylalanine and valine in the gauche\(^+\) rotameric state. Locations of minima are indicated by filled circles. The contour lines are labelled with energy in kcal/mole above the minimum energy points at \((\phi,\psi) = (-150,160)\) for β-methylalanine and \((\phi,\psi) = (-150,140)\) for valine.
Fig. III.4

(β-METHYLALANINE)
Fig. III.4 (cont'd.)
Fig. III.5 Conformational energy contour diagrams for the N-acetyl-N'-methylamide derivatives of β-methylalanine and valine in the gauche-rotameric state. Locations of minima are indicated by filled circles. The contour lines are labelled with energy in kcal/mole above the minimum energy points at (ϕ,ψ) = (-90,80) for β-methylalanine and (ϕ,ψ) = (-140,160) for valine.
Fig. III.5
Fig.III.5 (cont'd)
### TABLE III.3

**SIDE CHAIN ROTAMER POPULATIONS FOR THE N-ACETYL-N'-METHYLAMIDE DERIVATIVES OF β-METHYLALANINE AND VALINE**

<table>
<thead>
<tr>
<th>Rotamer</th>
<th><strong>Statistical Weight</strong> ($w\times10^{-3}$)</th>
<th><strong>Probability</strong> (%$P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>β-Methylalanine</strong></td>
<td><strong>Valine</strong></td>
</tr>
<tr>
<td>t</td>
<td>2.102</td>
<td>0.496</td>
</tr>
<tr>
<td>$g^+$</td>
<td>0.179</td>
<td>0.049</td>
</tr>
<tr>
<td>$g^-$</td>
<td>1.123</td>
<td>0.015</td>
</tr>
</tbody>
</table>
### TABLE III.4

PROBABILITIES$^a$ FOR EACH ($\phi, \psi$) CONFORMATIONAL REGION FOR THE N-ACETYL-N'-METHYLAMIDE DERIVATIVES OF ALANINE, $\beta$-METHYLALANINE, VALINE, $\beta$-METHYLVALINE AND HYDROXYETHYLGLUTAMINE

<table>
<thead>
<tr>
<th>Conformation Letter Codes</th>
<th>$&gt; \phi &gt;$</th>
<th>$&gt; \psi &gt;$</th>
<th>Ala$^b$</th>
<th>$\beta$-Methylalanine</th>
<th>Val$^b$</th>
<th>$\beta$-Methylvaline</th>
<th>Hydroxyethylglutamine</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-40 -110</td>
<td>-10 -90</td>
<td>4(6)</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>-40 -110</td>
<td>130 50</td>
<td>37(30)</td>
<td>45</td>
<td>52</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>D</td>
<td>-110 -180</td>
<td>110 20</td>
<td>20(34)</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>E</td>
<td>-110 -180</td>
<td>180 110</td>
<td>24(28)</td>
<td>19</td>
<td>21</td>
<td>91</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>-40 -110</td>
<td>180 130</td>
<td>13(0)</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>-110 -180</td>
<td>-40 -90</td>
<td>1(1)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

---

*a. All values are given in % to the nearest whole figure, for those regions on the ($\phi, \psi$) surface that contain conformations which contribute to at least 1% of the total.

*b. The values in parenthesis are those calculated by Zimmerman et al. (1977a) using only the minima on the energy surface but taking into account the librational entropy, for the purposes of comparison his fractional values have been converted to %. No minima occurs in region F.*
TABLE III.5  PREFERRED BACKBONE CONFORMATIONS$^a$ OF THE VALINE AND $\beta$-METHYLALANINE RESIDUES:
STATISTICAL PROBABILITIES AND MINIMUM CONFORMATIONAL ENERGY VALUES

<table>
<thead>
<tr>
<th>Valine</th>
<th>$\phi$</th>
<th>$\psi$</th>
<th>$\Delta E$</th>
<th>$P(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-90</td>
<td>100</td>
<td>0</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>-150</td>
<td>140</td>
<td>0.54</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>-80</td>
<td>-50</td>
<td>0.72</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>-130</td>
<td>30</td>
<td>2.43</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>$\beta$-Methylalanine</td>
<td>$\phi$</td>
<td>$\psi$</td>
<td>$\Delta E$</td>
<td>$P(%)$</td>
</tr>
<tr>
<td>-80</td>
<td>90</td>
<td>0</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>-150</td>
<td>130</td>
<td>0.50</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>-70</td>
<td>-50</td>
<td>0.59</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>-150</td>
<td>-60</td>
<td>1.05</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

a. The probability maxima represent the most probable backbone conformations in the regions tested in Table III.4, taking into account all side chain conformations, whereas the energy minima are for specific $\chi$ values.
Conformational probability contour diagrams for the N-acetyl-N'-methyamide derivatives of (a) alanine and β-methylalanine, and (b) valine and β-methylvaline. The contour lines are labelled in percentages and the filled circles represent the backbone conformations with the maximum probability of occurrence for each residue (also see Table III.5).
Fig. III.6(a)

$\psi$ (degrees)

$\phi$ (degrees)

ALANINE

$\beta$-METHYL ALANINE
E. Conformational Energy Calculations for N-acetyl-N'-methyl amide of Hydroxyethylglutamine

The conformational space available to the backbone of the hydroxyethylglutamine residue is represented in Figure III.7 as the probability (%) to the nearest integer over all side chain conformations available to the molecule within the isomeric state model chosen for this residue. The \((\phi,\psi)\) map is divided into regions which are considered to contain similar backbone conformations and (as in Figure III.1(a)) labelled according to the conformational letter code devised by Zimmerman et al. (1977a). The most probable \((P = 22\%)\) backbone conformation occurs in region E, which contains extended conformations and the next most probable in region C, which contains the \(C_7\) \(_{eq}\) ring conformations. The probability of the backbone conformation of HEG falling into either of the regions A and A\(^*\) (which contain the right- and left-handed \(\alpha\)-helix) or the region G is less than 2%.

The conformation of lowest energy within each set of 2,916 side chain conformations, for each of the 31 calculated points on the \((\phi,\psi)\) plane is shown in Table III.6. Only four side chain conformations have a probability of occurrence of \(>1\%.\) These are listed in Table III.7. Gauche rotamers predominate for those side chain torsion angles which have three rotational isomeric states. The probability of occurrence of each state for each side chain torsion angle taken singly (i.e. independently of its neighbours) is shown in Table III.8.
Fig. III.7 Conformational probability diagram for the N-acetyl-N'-methylamide derivatives of hydroxyethylglutamine. The probability (%) over all side chain conformations available to the molecule are plotted as values to the nearest integer. The regions defined by the conformational letter code are marked.
<table>
<thead>
<tr>
<th>Dihedral Angles (Degrees)</th>
<th>$\phi$</th>
<th>$\psi$</th>
<th>$\chi^1$</th>
<th>$\chi^2$</th>
<th>$\chi^3$</th>
<th>$\chi^5$</th>
<th>$\chi^6$</th>
<th>$\chi^7$</th>
<th>$\Delta E^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-130</td>
<td>120</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>0.0d</td>
</tr>
<tr>
<td>-131.3</td>
<td>146.2</td>
<td>-67.3</td>
<td>-62.8</td>
<td>-31.9</td>
<td>100.5</td>
<td>-55.1</td>
<td>75.1</td>
<td>-3.02</td>
<td></td>
</tr>
<tr>
<td>-130</td>
<td>60</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>-115.4</td>
<td>90.6</td>
<td>-63.9</td>
<td>-62.0</td>
<td>-41.1</td>
<td>103.4</td>
<td>-60.2</td>
<td>81.6</td>
<td>-2.41</td>
<td></td>
</tr>
<tr>
<td>-110</td>
<td>150</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>-140</td>
<td>150</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>-60</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>-160</td>
<td>120</td>
<td>180</td>
<td>60</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td>-156.0</td>
<td>135.4</td>
<td>-169.0</td>
<td>56.0</td>
<td>-108.0</td>
<td>85.0</td>
<td>57.5</td>
<td>-84.5</td>
<td>-4.90e</td>
<td></td>
</tr>
<tr>
<td>-130</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>2.02</td>
</tr>
<tr>
<td>-140</td>
<td>90</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>2.16</td>
</tr>
<tr>
<td>-70</td>
<td>120</td>
<td>180</td>
<td>60</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>2.23</td>
<td></td>
</tr>
<tr>
<td>-74.8</td>
<td>124.8</td>
<td>-175.1</td>
<td>59.3</td>
<td>-103.2</td>
<td>79.1</td>
<td>55.9</td>
<td>-63.9</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>-130</td>
<td>180</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>2.26</td>
</tr>
<tr>
<td>-80</td>
<td>90</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>150</td>
<td>-60</td>
<td>60</td>
<td>2.45</td>
</tr>
<tr>
<td>-114.8</td>
<td>88.1</td>
<td>-67.0</td>
<td>-67.6</td>
<td>-72.2</td>
<td>158.9</td>
<td>-63.4</td>
<td>67.0</td>
<td>-3.20</td>
<td></td>
</tr>
<tr>
<td>-110</td>
<td>90</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>2.86</td>
</tr>
<tr>
<td>-80</td>
<td>150</td>
<td>-60</td>
<td>-60</td>
<td>-90</td>
<td>150</td>
<td>-60</td>
<td>60</td>
<td>2.87</td>
<td></td>
</tr>
<tr>
<td>-100</td>
<td>120</td>
<td>180</td>
<td>60</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td>-110</td>
<td>30</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td>$\phi$</td>
<td>$\psi$</td>
<td>$\chi^1$</td>
<td>$\chi^2$</td>
<td>$\chi^3$</td>
<td>$\chi^5$</td>
<td>$\chi^6$</td>
<td>$\chi^7$</td>
<td>$\Delta E$</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>-100</td>
<td>60</td>
<td>-60</td>
<td>180</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>3.26</td>
<td></td>
</tr>
<tr>
<td>-170</td>
<td>90</td>
<td>180</td>
<td>60</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>-170</td>
<td>60</td>
<td>180</td>
<td>-90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>3.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-140</td>
<td>30</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>-60</td>
<td>60</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td>-160</td>
<td>-60</td>
<td>180</td>
<td>180</td>
<td>90</td>
<td>90</td>
<td>-60</td>
<td>3.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-80</td>
<td>-30</td>
<td>-60</td>
<td>180</td>
<td>-90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td>-70</td>
<td>-60</td>
<td>180</td>
<td>180</td>
<td>90</td>
<td>-90</td>
<td>60</td>
<td>60</td>
<td>3.75</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>-60</td>
<td>180</td>
<td>90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>4.31</td>
<td></td>
</tr>
<tr>
<td>-100</td>
<td>-60</td>
<td>-60</td>
<td>180</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td>-160</td>
<td>180</td>
<td>60</td>
<td>180</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>4.40</td>
<td></td>
</tr>
<tr>
<td>-90</td>
<td>0</td>
<td>-60</td>
<td>180</td>
<td>-90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>5.30</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>30</td>
<td>-60</td>
<td>180</td>
<td>90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>5.57</td>
<td></td>
</tr>
<tr>
<td>-100</td>
<td>180</td>
<td>-60</td>
<td>-60</td>
<td>-30</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>5.68</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>90</td>
<td>-60</td>
<td>180</td>
<td>90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>5.71</td>
<td></td>
</tr>
<tr>
<td>-50</td>
<td>150</td>
<td>60</td>
<td>180</td>
<td>90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>6.32</td>
<td></td>
</tr>
<tr>
<td>-70</td>
<td>180</td>
<td>-60</td>
<td>180</td>
<td>-90</td>
<td>90</td>
<td>60</td>
<td>-60</td>
<td>6.86</td>
<td></td>
</tr>
<tr>
<td>-160</td>
<td>60</td>
<td>60</td>
<td>180</td>
<td>90</td>
<td>-90</td>
<td>-60</td>
<td>60</td>
<td>8.69</td>
<td></td>
</tr>
</tbody>
</table>
TABLE III.6 - FOOTNOTES

a. The underlined values refer to the conformations arrived at by energy minimizations involving all 8 torsion angles (see text).

b. \( \omega \) and \( \chi'' \) are set at 180° for all conformations.

c. Relative to \( E = -11.651 \) kcal/mole for conformation \( \phi = -130, \psi = 120 \).

d. A perspective drawing of this conformation is shown in Figure III.8(a).

e. A perspective drawing of this conformation is shown in Figure III.8(b).
### TABLE III.7

**PROBABILITY OF OCCURRENCE OF SIDE CHAIN CONFORMATIONS FOR THE HYDROXYETHYLGLUTAMINE RESIDUE**

<table>
<thead>
<tr>
<th>Dihedral Angles</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^1$ $x^2$ $x^3$ $x^4$ $x^5$ $x^6$ $x^7$</td>
<td></td>
</tr>
<tr>
<td>-60 -60 -30 180 90 -60 60</td>
<td>31.8</td>
</tr>
<tr>
<td>-60 -60 -30 180 90 60 -60</td>
<td>3.3</td>
</tr>
<tr>
<td>180 60 -90 180 90 60 -60</td>
<td>1.8</td>
</tr>
<tr>
<td>180 60 -90 180 90 180 180</td>
<td>1.2</td>
</tr>
<tr>
<td><strong>All others</strong></td>
<td>$\leq 1.0$</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Side-chain angle</th>
<th>Rotational isomeric state and probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^1$</td>
<td>$+60^\circ(2.5)$  $-60^\circ(59.7)$  $180^\circ(37.8)$</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>$+60^\circ(19.9)$ $-60^\circ(40.6)$  $180^\circ(39.5)$</td>
</tr>
<tr>
<td>$\chi^3$</td>
<td>$+30^\circ(8.2)$  $-30^\circ(43.5)$  $+90^\circ(15.3)$ $-90^\circ(33.0)$ $+150^\circ(0.0)$ $-150^\circ(0.0)$</td>
</tr>
<tr>
<td>$\chi^5$</td>
<td>$+30^\circ(1.3)$  $-30^\circ(1.1)$  $+90^\circ(59.3)$ $-90^\circ(19.1)$ $+150^\circ(10.0)$ $-150^\circ(9.2)$</td>
</tr>
<tr>
<td>$\chi^6$</td>
<td>$+60^\circ(27.5)$ $-60^\circ(55.0)$  $180^\circ(17.5)$</td>
</tr>
<tr>
<td>$\chi^7$</td>
<td>$+60^\circ(52.8)$ $-60^\circ(24.3)$  $180^\circ(22.9)$</td>
</tr>
</tbody>
</table>
The most probable backbone and side chain conformation is visualized in Figure III.8. This is also the conformation of lowest overall energy, before minimization, shown in Table III.6.

Since the purpose of including HEG in this study was to predict the characteristic ratio of copolymers containing HEG, a total description of the conformational space available to the hydroxyethylglutamine residue was required. Thus, no attempt was made to locate the global minimum energy conformation of the residue or all of the local minima. However, since the probabilities of occurrence of certain side chain and backbone conformations were very high, each backbone conformation was examined which had a probability of occurrence greater than 5% (i.e. five $(\phi, \psi)$ pairs in all), and those side chain conformations were selected which gave the lowest energy for the overall conformation. All dihedral angles (excepting $\omega$ and $\chi^4$ which were set at 180°) were also minimized. These minimized conformations are shown underlined in Table III.6 below the conformations that were the starting point for each minimization. In fact, the conformation of lowest energy of these five has neither the side chain nor the backbone conformations of highest probability and may not be the global minimum for this residue. This molecule is visualized in Figure III.8(b). Both the minimized and starting conformations were checked for the occurrence of hydrogen bonds using the criteria of Zimmerman et al. (1977a) for hydrogen bond formation. The HEG residue is theoretically capable of forming three types of hydrogen bond - (a) backbone-
Perspective drawings of two conformations of the N-acetyl-N'-methylamide derivatives of hydroxyethylglutamine (a) $\phi = -130^\circ$, $\psi = 120^\circ$ (b) $\phi = 156.0^\circ$, $\psi = 135.4^\circ$ (see Table III.2 for values of the other torsion angles). The dashed line represents a hydrogen bond.
Fig.III.8
backbone; (b) side chain-side chain, and (c) side chain-backbone. Each backbone conformation tested, except the one of highest energy (which did not show any hydrogen bond formation), has only one hydrogen bond, and this was in category (c). In each case the bond was between the side chain hydroxyl group as hydrogen donor and the backbone carbonyl group of the N-acetyl group as acceptor (see Figure III.8(a) and (b)). For the minimized conformations the H...O distance was 1.68 Å.

3. DISCUSSION

A. Conformational Energy Minima and the Total Conformational Space

The conformational energy minima and the (ϕ,ψ) maps for the glycine and alanine residues can be compared with the wealth of information obtained for these residues in previous studies using both empirical energy calculations and molecular orbital calculations. Surprisingly, the computed conformational space available to these residues has changed little since the early hard-sphere calculations (Ramachandran et al., 1963; Leach et al., 1966). The overall shapes of the maps have remained the same, the refinements in the calculations have generally led to an extension of the energetically allowed areas and the locations of the conformational energy minima within these areas have changed.

In most studies using empirical energy calculations as in this work, the global minimum for the glycine residue has been located in region C (Brant et al., 1967; Popov et al., 1968a, 1968b; Lewis et al., 1973a; Zimmerman et al., 1977a) an
exception being Ponnuwamy and Sasisekheran (1970) who computed the global minimum for unblocked glycine to be in the bridge region on the border of region A. Maigret et al. (1970a), using PCILO calculations also located the global glycine minima in region C.

Brant et al. (1967) found the global minimum for the alanine residue to be in region F (if coulombic interactions were ignored, it appeared in region A). Ponnuwamy and Sasisekheran (1970) located it in region B (close to region A), but other workers (Popov et al., 1968a, 1968b; Lewis et al., 1973a; Zimmerman et al., 1977a) located the lowest energy conformation for the alanine residue in region C as is the case in this study. The molecular orbital calculations (Maigret et al., 1970b) give the global minimum of alanine in the C* region with the next lowest energy conformation (differing by only 0.5 kcal/mole) in the C region. Refining these calculations to include minimization of $\phi$, $\psi$ and $\chi'$ (Maigret et al., 1971a), however, reversed the order of energetic preference of these two conformations.

Since the side chain of the $\beta$-methylalanine residue is larger than that of alanine its conformational space should lie within the low energy regions of alanine. The conformational energy contour diagrams for the individual rotamers of $\beta$-methylalanine each cover only a part of the conformational space available to alanine. Together, however, as represented by the composite map (Figure III.1) the conformational space available to $\beta$-methylalanine is only slightly more restricted than that for alanine. The minimum energy conformations for alanine do not have counterparts for all three rotamers in $\beta$-methylalanine. Only those minima
appearing in regions A, A* and E (see Table III.1) are also minima for each rotamer, but each backbone conformation which is a minimum for the alanyl residue has at least one side chain conformation available to it for β-methylalanine.

Other work has been published on β-methylalanine (Ponnuswamy and Sasisekheran, 1970) where the molecule was not blocked by end groups as in this study, but was considered in two ionic forms - as both a dipolar ion and a cation; the minimum energy conformation for each of the three side chain rotamers of each ionic form occurred in slightly different positions in region A, but close to region B. Miller and Goebel (1968) also produced (ϕ,ψ) maps of residues with CβH2 - CγH2 side chains, all three side chain rotamers had minima in region C. Neither of these studies are in agreement with the present results. The overall minimum for each rotamer does not fall in the same area of the (ϕ,ψ) map. Thus, for the trans and gauche- rotamers the minima fall into region C, the region that is most favoured for the alanine residue. For the gauche+ rotamer, the most preferred conformation is in region E. The discrepancies between the present findings and those in the two previous studies could well be due to the fact that in the former, the models used were unblocked and in the latter the "combined atom" approximation was used in which hydrogen atoms were included by using enlarged van der Waals carbon atomic radii.

It is obvious from the relative contributions of the three side chain populations for β-methylalanine and valine (see Table III.3) that the substitution of a second
methyl group on the β-carbon atom considerably restricts the permissible rotations about the Cα-Cβ bond. While the trans rotamer predominates for both types of residue, for valine it accounts for nearly 90% of the total side chain rotamer population, but only about 60% for the β-methylalanine residue in which there is a significant contribution (33%) from the gauche- rotamer. This finding is in conflict with the work of Popov et al. (1968a) on the valine residue, using potential energy functions, who found that the gauche- rotamer was the most favoured.

The addition of the second β-methyl group to β-methylalanine to give valine reduces the total conformational space available to the molecule (compare Figures III.1(b) and III.2(a)) and also reduces the total number of regional minima from 17 for β-methylalanine to 10 for valine. Only one region, A, that which contains the right-handed α-helix, contains a minimum for all three staggered rotamers, and the extended region E contains minima for the two gauche rotamers. The regions C, F, D and A* contain one minimum each and, unlike β-methylalanine, there are no minima in the regions F*, C* and G. The global minimum for both gauche rotamers occurs in region E and that of the trans rotamer occurs in region C. Miller and Goebel (1968) also studied the valine residue, but with the side chain only in the trans rotameric state. The global minima of this map was also in region C. However, the PCILO calculations on this molecule (Maigret et al., 1971b) suggest the C7 conformation is the most favoured for all orientations of the side chain and not just for the trans rotamer.
The effect on the conformational space available to the backbone when a third β-methyl group is added to valine, giving β-methylvaline, is profound. The energy contour diagram for this molecule (see Figure III.2(b)) shows that the only backbone conformations allowed for this molecule are, not surprisingly, those that are allowed for every one of the three valine rotamers. The global minimum for β-methylvaline (see Table III.2) occurs in the region E where the backbone is very extended and the only other minimum (in region A) has an energy 3 kcal/mole greater than this. Thus, although there are fewer backbone conformations available to valine than to β-methylalanine or alanine, the presence of a tert-butyl group on the Cα atom severely restricts the number of backbone conformations available to β-methylvaline so that it can adopt only a very extended conformation.

This property of the residue could make it a particularly useful tool in synthetic analogues of biologically active peptides, especially in combination with the α-amino iso-butyric acid residue which can only adopt the (ϕ,ψ) values close to those of the left- or right-handed α-helix (Burgess and Leach, 1973). Either of these two residues could be substituted for other hydrophobic residues such as valine, leucine and isoleucine, to manipulate the conformation of a molecule and restrict local (ϕ,ψ) values to extended or α-helical ones. The biological activity of such analogues could provide evidence about the conformation of the naturally-occurring molecule.
The potential energy functions applied in the study of Lewis et al. (1973a) on single residues are based on intermolecular interactions in X-ray crystal structures of model compounds (Momany et al., 1974; McGuire et al., 1972). These potential functions were refined from experimental data on intramolecular interactions (Momany et al., 1975) and it is this final set of parameters that are used in ECEPP, both in this study and in that by Zimmerman et al. (1977a). Thus, the conformational energy contour diagrams for the glycine and alanine residues reproduced here (Figures II.2 and III.1(a)) are identical to those of Zimmerman et al. and the low energy minima listed in Tables III.1 and III.2 for alanine and valine are as close as would be expected to the values given in the publication of Zimmerman et al. (1977a) considering that neither $\phi$ nor $\psi$ were minimized in the results presented here.

There are differences, however, in the low energy conformations of blocked amino acid residues calculated in this thesis, and those of Lewis et al. (1973a) and these differences are mainly attributable to two changes in the potential functions. The effect of "softening" the repulsive interactions between an amide nitrogen and an amide hydrogen in another peptide group results in the area around the 'bridge' region (region B) being energetically more allowed (3-4 kcal/mole) than previously (10 kcal/mole, Lewis et al., 1973a). The location of certain minima (e.g. $\phi = -160$, $\psi = -60$ for the alanine residue given in Table III.1) is also closer to the bridge region than in the study by Lewis.
et al (-160, -68) and the relative energies of these minima have decreased. This general effect of making the bridge region conformationally more favourable is in accord with the occurrences of amino acid residues with conformations in this region in X-ray crystal structures of protein (Ramachandran and Sasisekheran, 1968).

The second major difference between the work of Lewis et al. (1973a) and that of Zimmerman et al. (1977a) and the work presented in this thesis has been the introductions of large barriers to rotations about aliphatic C-C and C-N single bonds in the side chains of the residues; so that the values of $\chi^1$ for the alanine and valine energy minima in this study are closer to the values for strictly staggered conformations ($\pm 60^\circ, 180^\circ$) than in the study by Lewis et al. where some of the values adopted by $\chi^1$ deviated by about 20° from their presumed positions of minimum energy.

B. Probability of Occurrence of Backbone Conformations

From the conformational probability diagram of backbone conformations (Figure III.7) (for the hydroxyethyl-glutamine residue) the most likely backbone conformations are an extended one involving a distorted C$^5$ ring in region E or a distorted C$^7_{eq}$ ring in region C. Indeed, the most favoured ($\phi, \psi$) region calculated for all of the residues studied, except for $\beta$-methylvaline, is region C, which contains the above C$^7_{eq}$ ring with a hydrogen bond between the C = O and N-H groups adjacent to the central residues. This region accounts for 37 to 52% of the total conformational space of each type of residue (see Table III.4). The next
most favoured region is region E which contains extended backbone conformations. The distributions among the regions for the residues alanine, β-methylalanine, hydroxyethylglutamine and valine are somewhat similar but that for β-methylvaline is heavily biased towards the extended region of the map which accounts for 90% of the total conformational space. These results are consistent with empirical conformational energy calculations on the naturally occurring amino acids (Zimmerman et al., 1977a) whose lowest conformational energy minima tend to fall in regions C and E, and with experimental data that claim to have demonstrated the existence of the C\textsuperscript{5} and C\textsubscript{eq}\textsuperscript{7} structures for N-acetyl-N'-methylamide derivatives of amino acid residues.

Thus, the NMR and infrared spectroscopy on N-acetyl-N'-methylamides of several amino acids have been investigated by a number of workers (Avignon et al., 1969, 1972; Marraud et al., 1970; Cung et al., 1973; Bystrov et al., 1969) in solvents of low dielectric constant (e.g. CCl\textsubscript{4}); conditions which approximate most closely to the in vacuo status of the molecule assumed in these calculations. Their infrared spectra show, in addition to the main NH (unperturbed) stretching frequency at 3450-3460 cm\textsuperscript{-1} a number of peaks and shoulders at lower frequencies (between 3300 and 3450 cm\textsuperscript{-1}). Their assignments and interpretations differ somewhat from those of the earlier classical studies of Japanese workers (e.g. Mizushima et al., 1952; Tsuboi, 1959) but, nevertheless, claim to demonstrate the existence of C\textsuperscript{7} and C\textsuperscript{5} hydrogen-bonded conformations. Recent measurements by Maxfield et al.
(1978) on the basis of reassignments of NH stretching frequencies find some evidence for the $C^5$ ring, but none for a $C^7$ hydrogen bond. Their reassignments are consistent, however, with a population of conformers in which the $\beta$-carbon of the side chain is close enough to the NH group to perturb it. This applies to any conformation in which $\phi$ is about $-60^\circ$; region C of the alanine conformational energy map (see Figure III.1(a)) contains a variety of conformers which comply with this requirement. This region contains a broad shallow minimum (see conformational probability diagram Figure III.6(a)) in which those conformers that form a $C^7$ hydrogen bond are probably too small a fraction of the total statistical ensemble to provide an experimentally discernible NH band distinct from that of the conformers in region C as a whole. An infrared study of the N-acetyl-N'-methylamide of $\beta$-methylvaline in dilute nonpolar solvents could add weight to the assignments of Maxfield et al., since 91% of all backbone conformations lie in region E (Table III.4) and the direct steric effect of the additional methyl groups should render the $\beta$-carbon effect on the NH stretching frequency much less pronounced than in alanine.

Other recent experimental analyses of the conformational states of N-acetyl-N'-methylamides of amino acid residues with bulky, but polar, side chains show no preference for either the $C^7$ or $C^5$ conformation in the solid state (Koyama et al., 1977). Using Raman spectroscopy on crystals (Koyama et al., 1977) and X-ray diffraction analysis on crystals (Oyama et al., 1977) two stable conformers were
demonstrated for the seryl residue with backbone conformations in the B and A regions. Both conformers were stabilized by intramolecular side chain-backbone hydrogen bonds. None of the other residues in this study (tyrosine, histidine and proline) adopted $C^7$ or $C^5$ structures, but in all cases there was evidence that intermolecular hydrogen bonding in the solid state was stabilizing otherwise apparently energetically unfavourable structures.

The normalized statistical weights of the alanine residue can be calculated from the data of Zimmerman et al. (1977a) using only the minima on the energy surface, but taking into account the librational entropy at each minimum. These figures are given in parenthesis in Table III.4 for the regions in which each minimum falls. A study by Zimmerman et al. (1977b) on the statistical weights of dipeptides composed of glycine, alanine or proline suggested that the results obtained by a total search of the conformational space (the method used in this work) gives similar results to those derived by using only the local minima and the librational entropy. Although the figures for alanine are in general agreement, there is still some discrepancy. Using the approach of Zimmerman et al. (1977a) the most favoured regions are D>C>E, whereas our figures show C>E>D. Also, regions in which no minima occur, such as region F, are completely ignored when assessments are made solely on the basis of local minima and librational entropy. This region, by our calculations, makes a significant contribution (13%) to the total. Since the minimum energy conformations of single
amino acid residues are often used as starting points for predicting peptide conformations (e.g. Isogai et al., 1977) and \( \beta \) -bend probabilities for peptide sequences (Zimmerman and Scheraga, 1977) it is important to correctly assign the relative probabilities of the various conformational regions. The results presented here have suggested that (in contrast with the conclusions of Zimmerman et al. (1977b)) a total search of the conformational space is more accurate than using librational entropies in conjunction with local minima.

For the residues with side chain populations, namely valine and \( \beta \)-methylalalaine, conformational probability contour diagrams give a much clearer representation of the conformational space available to the residue than do the composite energy diagrams. The statistically most favoured backbone conformations are marked on these diagrams and these positions conform with the high contributions expected from the conformational energy minima. Nevertheless, the most probable backbone conformations do not rank in exactly the same order as the values of the conformational energy minima (see Table III.5). This is because each minimum is for discrete values of \( \phi, \psi \) and \( \chi^1 \), whereas the probabilities are calculated from the contributions of all three side chain rotamers. Hence, for \( \beta \)-methylalanine, although there is a value of \( \chi^1 \) for which the conformational energy at \((-150,130)\) is lower than that at \((-70,-50)\) the probability of the other two rotamers occurring at \((-150,130)\) is lower than at \((-70,-50)\). This is reflected in a higher overall probability for the latter conformation.
Probability values give a quick appraisal of the relative importance of favoured conformations. The values of the probabilities of the minimum energy conformations may seem low, the highest value for any one \((\phi,\psi)\) pair being only 4% of the total. These values, however, refer to the minima of regions which may contain many other conformations which are almost equally probable. If the total \((\phi,\psi)\) space is partitioned into regions containing similar conformations, by summing the contributions of each \((\phi,\psi)\) pair at each point in that region we obtain a clearer idea of the relative contributions of each type of conformation, rather than focusing attention on conformations narrowly defined by specific \((\phi,\psi)\) values. This has been done in Table III.4.

C. Probability of Occurrence of Side Chain Conformations

It is apparent from Tables III.4 and III.2 that backbone conformations exist for both valine and \(\beta\)-methylalanine for which all side chain rotamers are permitted. There have been many published observations on the side chain conformation of the valine residue in small molecules in which valine is the sole amino acid residue, as well as in other peptides and proteins; each of the side chain conformers of valine do occur to a significant degree both in the crystalline (Pullman and Pullman, 1974; Smith et al., 1975) and the solution state (Gibbons et al., 1972; Hansen et al., 1975). From our calculations on rotamer populations (see Table III.3) the trans rotamer should be strongly preferred for most regions of the \((\phi,\psi)\) map. Preferred populations of backbone \((\phi,\psi)\) values will require compatible side chain rotameric states.
Our calculations would indicate that for valine, extended backbone conformations would favour the gauche rotamers, whereas α-helical or \( C_{eq}^7 \) ring structures would demand the trans rotameric state for the side chain. On the other hand, where valine or β-methylalanine exist in a "structureless" situation in which near-neighbour interactions are reduced, e.g. in proteins or peptides in a randomising solvent, the trans rotamer should predominate. These generalisations should be of some value in current and future studies on side chain rotamer analysis with (for example) peptide antibiotics, such as gramicidin and valinomycin. [A great deal of effort is expended in seeking preferred rotameric states by measuring the temperature dependence of vicinal coupling constants \( J_{NH-C_H^a} \ J_{C_H^a-C_H^b} \) using 'H-NMR (Bovey, 1972; Steinfeld et al., 1972; Mauger, 1975; Dale and Jones, 1976).]

It should always be borne in mind that under experimental conditions the conformations adopted by the backbone will be influenced by factors that are not incorporated in these calculations, such as interactions with solvent and neighbouring amino acid residues.

Interactions with water have been incorporated into conformational energy calculations very recently. Hodes et al. (1978a) explored the effect of hydration on the conformational stability of the terminally blocked residues of the 20 naturally occurring amino acids using conformational energy calculations that incorporated a modified hydration shell model. Free energy hydration terms were introduced to account for "specific hydration" due to water-solute hydrogen
bonding and for "non-specific hydration" describing the interaction of the solute with the water molecules in the first-neighbour shell.

Using the conformational energy minima obtained in this study, these workers calculated bend probabilities of blocked dipeptides (Hodes et al., 1978b) and compared them with values obtained from the crystal structure of native proteins. Their results suggested that effects due to hydration may not be important for non polar amino acids, such as β-methylalanine and valine, since the inclusion of hydration in the conformational energy calculations improved the values of the computed bend probabilities for polar dipeptides, but yielded poorer values for non polar ones, whereas the omission of hydration leads to better results for non polar peptides.

Hodes et al., (1978b) also concluded that the close correlation between the minimum energy conformations of blocked dipeptides and those of their constituent amino acid residues supports the concept of the dominance of intrareside interactions in determining protein conformations.

All but four of the side chain conformations tested for the hydroxyethylglutamine residue had probabilities of occurrence of less than 1%, but one conformation was overwhelmingly preferred with a probability of occurrence of 32% (see Table III.6). The strong preference for this side chain conformation is reflected in the probability of occurrence of each rotational isomeric state for the side chain dihedral angles where, for the angles with torsional minima, the gauche- rotamers are most preferred for angles
\( \chi^1, \chi^2 \) and \( \chi^6 \), and the gauche\(^+\) rotamer is preferred for \( \chi^7 \), indicating that the side chain tends to bend back on itself towards the backbone. Computations of the interatomic distances between potential hydrogen bond donor and acceptor atoms for selected \((\phi,\psi)\) pairs with the most preferred side chain conformation indicated that a strong hydrogen bond between the terminal hydroxyl group side chain acting as proton donor to the carbonyl oxygen of the backbone is important in lowering the energy for these conformations. This forms an 11-membered ring. The high probability of this side chain conformation is almost certainly because it orientates the side chain in a position where this hydrogen bond can be formed for a large number of backbone conformations. Side chain-backbone hydrogen bonds have also been shown to play a prominent role in the minimum energy conformations of the N-acetyl-N'-methylamides of serine and threonine (Zimmerman et al., 1977a). In these two cases, the hydrogen bonding produces a 7-membered ring. In proteins such as haemoglobin larger hydrogen bonded rings are observed involving the hydroxy1 groups of serine and threonine interacting with main-chain carbonyl groups which are several residues removed (M.F. Perutz, 1969). The 7-membered ring is probably allowed for the hydroxyethylglutamine residue for certain backbone conformations. It does not appear, however, within side chain conformations which occur significantly frequently taking into account all \((\phi,\psi)\) backbone values considered. The reason for the observed preference for an 11-membered hydrogen bonded ring may be associated with additional favourable electrostatic and van der Waals energetic contributions.
Since solvent interactions are not incorporated in these calculations, it is important to consider whether conformations involving hydrogen bonds such as those described above, are likely to occur in the presence of polar solvents where the backbone peptide groups may be solvated.

In the study by Hodes et al. (1978a) inclusion of hydration terms tended to destabilize backbone-backbone hydrogen bonded conformations, but not the side chain-backbone hydrogen bonded conformations of serine, so that the four lowest energy conformations of the seryl residue had this type of hydrogen bond. In the study by the same workers (1978b) on the influence of hydration on the conformational stability and formation of bonds in blocked dipeptides, the three single residues occurring in the dipeptides examined that could form backbone-side chain hydrogen bonds, i.e. Asn, Asp and Ser, retained conformational energy minima that had this type of hydrogen bond, even on hydration. The results of Hodes et al. indicate that side chain-backbone hydrogen bonds involving OH...CO interaction is not precluded by other interactions of the molecule with water.

Energy calculations have been applied in computing the characteristic ratio of polymers in the unperturbed state at the $\Theta$-point, where interactions between neighbouring chain units are dominant over long range interactions between remote segments of the molecule and where polymer-solvent interactions are negligible (P.J. Flory, 1969). At the $\Theta$-point conformations with intraresidue hydrogen bonds will contribute significantly in the overall spectrum of
conformations available to the randomly coiled polymer. It may be that water is not the ideal solvent for deriving the unperturbed dimensions of polypeptides. However, the theoretically predicted value of the characteristic ratio of poly-L-hydroxyethylglutamine of 9.75 obtained in this study (and reported in Chapter IV, section 2.B) is in excellent agreement with the experimentally derived random coil dimensions (10 ± 1) of this polymer in water at 30°C (Mattice and Lo (1972)). The agreement between these values indicates that the inclusion of an intraresidue hydrogen bonding potential in the overall energy term for the estimation of chain dimensions in the unperturbed state is not inappropriate. The findings of Joubert et al. (1970), that the conformation of poly-L-hydroxyethylglutamine in water is that of a truly statistical random coil also suggests that in this solvent interactions of poly-L-hydroxyethylglutamine with water can be ignored.

4. CONCLUSIONS

The N-acetyl-N'-methylamides of a series of amino acids with increasingly bulky side chains, namely alanine, β-methylalanine, valine and β-methylvaline has been subjected to an exhaustive theoretical conformational analysis. The conformational energy values for 10° increments in \( \phi \) and \( \psi \) values over the entire range available to these angles have been calculated. For those of the series with rotameric states the side chain torsion angle \( \chi^1 \) value was allow to vary about the 'staggered' values of ±60° and ±180° until a minimum energy for that conformation was obtained. It was
found that the conformational space available to the 
$\beta$-methylalanine residue is only slightly less than that for alanine, but that a second C$^\gamma$ methyl group to give valine does, as is well known, restrict the conformational space of the residue. The addition of a third C$^\gamma$ methyl group to valine, to give $\beta$-methylvaline, dramatically confines the $(\phi,\psi)$ values available to the backbone so that only very extended conformations in the region (-130,140) are allowed. The conformational energy minima of the alanine residue have counterparts for at least one of the rotameric states of $\beta$-methylalanine - which has a total of 17 minimal energy conformations - but the alanine minimum in the position (-160,-60) does not have a counterpart for the valine residue.

The total conformational space of these molecules is visualized both by energy contour diagrams and conformational probability diagrams. The statistical probabilities of various side chain and backbone conformations are calculated. The region on the $(\psi,\psi)$ map that incorporates the C$^7_{eq}$ ring is the statistically most probable set of conformations for the alanine, $\beta$-methylalanine and valine residues (40, 45 and 52%, respectively), whereas the most probable conformational states of the $\beta$-methylvaline residue are extended conformations (91%). $\beta$-methylalanine and valine show favoured backbone conformations for each rotameric state. The C$^7_{eq}$ ring is most likely for the trans rotamer for both residues and for the gauche$^-$ rotamer of $\beta$-methylvaline while the gauche$^-$ rotamer of valine and the gauche$^+$ rotamers for both residues prefer extended conformations. Where $(\phi,\psi)$ values are randomly
assigned and the backbone of each residue can be considered to be structureless, the trans rotamer is strongly preferred with a probability of 62% for δ-methyalanine and 88% for valine.

The conformational space of the N-acetyl-N'-methylamide derivative of N⁵-(2-hydroxyethyl)-L-glutamine has been described in terms of the statistical percentage probabilities for 30° increments in φ and ψ within the areas of the (φ,ψ) map allowed for the alanine residue and averaged over all side chain conformations within the rotational isomeric state model chosen for this molecule. The most probable backbone conformation is an extended one and the most probable side chain conformation is one in which those gauche rotamers predominate, which facilitate the formation of a hydrogen bond between the terminal side chain hydroxyl hydrogen atom and the carbonyl oxygen of the N-acetyl group.
CHAPTER IV

THE AVERAGE RANDOM COIL DIMENSIONS

OF HOMO AND COPOLYPEPTIDES
1. **INTRODUCTION**

The identification of conformations of minimum energy of single amino acid residues is finding increasing use in studies on the tertiary structure of peptides (Burgess et al., 1973; Zimmerman et al., 1977b; Isogai et al., 1977). The validity of this type of calculation may be assessed by comparing the predicted structure with that derived experimentally, either by X-ray crystallographic (Zimmerman et al., 1977b) or solution studies (Jones et al., 1976; Roques et al., 1976). For polypeptides, interactions between residues which are nearest neighbours are dominant in the computations. This simplification may invalidate the comparisons so that experimental observations fail to agree, with the theoretically predicted structure of minimum energy. The theoretically predicted total conformational space available to a residue can also be tested for validity by comparing the characteristic ratio of a randomly coiled homopolypeptide (derived theoretically from the conformational energy surface of the monomer unit over the ($\phi,\psi$) plane) with that deduced experimentally from the hydrodynamic behaviour of the polymer. Excluding the case of amino acid residues containing a pyrrolidinone ring, there are few exceptions (Burgess et al., 1975; Hawkins and Holzer, 1972) to the close agreement between the theoretically predicted (Brant and Flory, 1965b; Miller and Goebel, 1968) and experimentally observed (Brant and Flory, 1965a; Fujita et al., 1966; Terbojevitch, 1967) unperturbed dimensions of polyamino acids. In such cases, intraresidue interactions alone are sufficient to account for the behaviour of randomly coiled unperturbed homopoly-
peptides. This is not the case with polymers containing the proline and hydroxyproline residues where some experimental measurements of the characteristic ratio (Mattice and Mandelkern, 1971a; Clark and Mattice, 1977) have been markedly lower than the ratios predicted theoretically for poly-L-proline using non-cooperative configurational statistics (Schimmel and Flory, 1967). However, these amino acids exhibit the phenomenon of cis-trans isomerism across the peptide bond and even a small number of cis-peptide units can markedly lower the chain dimensions (Schimmel and Flory, 1967). It is not permissible to ignore interactions between neighbouring residues in making theoretical estimates of residues when peptide groups are in the cis conformation. Where $\omega$, the dihedral angle about the peptide bond, is allowed to vary continuously, cooperative methods must be used to calculate the characteristic ratio (Tanaka and Scheraga, 1975d). Theoretical values obtained using such methods for poly-L-proline including a proportion of 5% cis residues in the chain (Tanaka and Scheraga, 1975d) agree closely with experimental estimates on this polymer (Mattice and Mandelkern, 1971a).

Most polyamino acids have a characteristic ratio similar to that of poly-L-alanine, namely 8 to 9 (Brant and Flory, 1965b; Miller and Goebel, 1968). Polyglycine, however, which has no side chains and in which there is, therefore, a much larger conformational space available to the backbone, has a calculated characteristic ratio of about 2. It is impossible to verify this theoretical prediction due to the
highly insoluble nature of polyglycine (Brant et al., 1967), but glycine residues have been incorporated, in varying proportions up to 46%, into random copolymers containing more soluble amino acid residues. The measured dimensions of these products show varying degrees of agreement with the theoretically predicted copolymer dimensions (Miller et al., 1967; Rao et al., 1975; Mattice and Mandelkern, 1971b).

The types of residues used so far have either been soluble only in strong acids (Rao et al., 1975) or have been electrolytes (Miller et al., 1967) or have contained a pyrollidine ring (Mattice and Mandelkern, 1971b). Hydroxyalkylglutamines have been used extensively in copolymers to 'solubilise' amino acid residues in order to measure their helix-coil stability constants in water (von Dreele et al., 1971; Dygert et al., 1976). The advantages of using polymers containing this type of residue are that they are (a) likely to be water soluble and thus provide a better model for the behaviour of denatured proteins in aqueous solutions than would copolymers in non-aqueous solvents; (b) they are non-ionic and charge effects need not, therefore, be considered, and (c) in the case of hydroxyethylglutamine, its polymer shows no tendency to form helical structures in water (Lotan et al., 1966; Joubert et al., 1970), and its conformation in this solvent is considered to be typical for a truly statistical random coil (Joubert et al., 1970).

The present study set out to compare the effect of side chain length on the experimentally determined and theoretically predicted characteristic ratio of polypeptides in the unperturbed state. Copolymers were chosen containing
hydroxethylglutamine, together with residues carrying increasingly long side chains as the major component. Three sequential copolymers were selected, with the repeating sequence \([\text{hydroxyethylglutamine-}X-X]\), where \(X = \text{glycine, alanine or } \beta\text{-methylalanine}\). The use of sequential rather than random copolymers eliminates the possibility of incorporating blocks of one type of residue. This may occur in random polymerisation leading to problems with solubility, localised formation of secondary structures and difficulties in the theoretical assessment of unperturbed dimensions of such conformationally inhomogeneous polymers. This chapter describes the theoretical prediction of the unperturbed dimensions of these copolymers.

In order to determine the theoretical characteristic ratio of copolypeptides, the weighted average transformation matrix for each type of monomer must be determined (P.J. Fory, 1969) using the total conformational energy space available to the backbone. Since hydroxyethylglutamine has nearly 3,000 side chain conformations for each backbone conformation some approximations must be made in the description of the conformational space used in the weighting of the unperturbed dimensions of this molecule in order to reduce the requirements for computer time. Approximations used by other workers have not always been well verified, e.g. Miller and Goebel (1968) neglected side chain atoms remote from the backbone and incorporated them in their calculations as an enlarged \(C^\beta\) atom. The choice of grid for the (\(\phi,\psi\)) conformational space has previously seemed arbitrary, thus some workers have chosen a \(10^\circ\) grid for the (\(\phi,\psi\)) conformational space (Burgess
et al., 1975) some 20° (Brant et al., 1967) and others 30° (Brant and Flory, 1965b). To establish when such approximations were justified the methods of energetic statistical weighting and the effects of refinements in the description of the conformational space on the computed chain dimensions of polypeptides was explored and the results used to select a suitable method for estimating the unperturbed dimensions of poly-L-hydroxyethylglutamine and thus its copolymers. A series of amino acids with short side chains and a maximum of three side chain conformations for each point on the (ϕ,ψ) map was chosen so that an exhaustive description of their conformational space could be made using a fairly modest amount of computer time. The results of this comparative study are also given in this chapter.

2. RESULTS

A. Unperturbed Chain Dimensions of Polypeptides with Increasingly Branched Side Chains

The characteristic ratio obtained in this study for a freely-rotating peptide chain with no energetic weighting is 1.89. The characteristic ratios, C_0, for the homopolymers composed of glycine, alanine, β-methylalanine and valine are listed in Table IV.1. The characteristic ratios of β-methylvaline for various values of n are given in Table IV.2. In both tables values derived from minimized and non-minimized energy surfaces at 10° and 30° intervals are shown. For those residues with non-symmetric side chains either the type of weighting used where all three rotamers are being considered or the individual rotameric state of the side chain, is indicated under the subheading "side chain".
<table>
<thead>
<tr>
<th>Residue</th>
<th>Side-chain</th>
<th>With Energy Minimization</th>
<th>Without Energy Minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10° Grid</td>
<td>30° Grid</td>
</tr>
<tr>
<td>Glycine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alanine</td>
<td></td>
<td>8.06</td>
<td>8.44</td>
</tr>
<tr>
<td>3-Methylalanine</td>
<td>gauche+</td>
<td>31.29</td>
<td>34.64</td>
</tr>
<tr>
<td></td>
<td>gauche-</td>
<td>8.31</td>
<td>8.58</td>
</tr>
<tr>
<td></td>
<td>trans</td>
<td>7.15</td>
<td>6.63</td>
</tr>
<tr>
<td></td>
<td>composite</td>
<td>6.80</td>
<td>6.68</td>
</tr>
<tr>
<td></td>
<td>statistical</td>
<td>7.02</td>
<td>6.82</td>
</tr>
<tr>
<td>Valine</td>
<td>gauche+</td>
<td>90.78</td>
<td>313.52</td>
</tr>
<tr>
<td></td>
<td>gauche-</td>
<td>27.24</td>
<td>14.79</td>
</tr>
<tr>
<td></td>
<td>trans</td>
<td>10.51</td>
<td>11.61</td>
</tr>
<tr>
<td></td>
<td>composite</td>
<td>9.27</td>
<td>9.65</td>
</tr>
<tr>
<td></td>
<td>statistical</td>
<td>9.21</td>
<td>9.58</td>
</tr>
<tr>
<td>Number of amino acid residues in the chain</td>
<td>With Energy Minimisation</td>
<td>Without Energy Minimisation</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10° Grid</td>
<td>30° Grid</td>
<td>10° Grid</td>
</tr>
<tr>
<td>100</td>
<td>52.41</td>
<td>36.01</td>
<td>72.98</td>
</tr>
<tr>
<td>500</td>
<td>94.82</td>
<td>49.04</td>
<td>249.70</td>
</tr>
<tr>
<td>1000</td>
<td>102.22</td>
<td>50.76</td>
<td>344.99</td>
</tr>
<tr>
<td>1500</td>
<td>104.69</td>
<td>51.33</td>
<td>388.57</td>
</tr>
<tr>
<td>∞</td>
<td>109.63</td>
<td>52.48</td>
<td>483.83</td>
</tr>
</tbody>
</table>
The statistically weighted average matrices $\langle T \rangle$ used to calculate the characteristic ratios shown in Tables IV.1 and IV.2 are given in Tables IV.3 to IV.6.

Table IV.3 - $\langle T \rangle$ elements for polyamino acids with different minimized energetic weightings and a $10^\circ$ integration interval.

Table IV.4 - $\langle T \rangle$ elements for polyamino acids with different minimized energetic weightings and a $30^\circ$ integration interval.

Table IV.5 - $\langle T \rangle$ elements for polyamino acids with different non-minimized energetic weighting and a $10^\circ$ integration interval.

Table IV.6 - $\langle T \rangle$ elements for polyamino acids with different non-minimized energetic weightings and a $30^\circ$ integration interval.

The rate of convergence of the characteristic ratio with chain length $n_p$ is shown in Figure IV.1 for all of the homopolypeptides, except poly-L-$\beta$-methylvaline which is shown in Figure IV.2. These characteristic ratios were calculated with the most refined techniques, that is, weighting with a $10^\circ$ grid using minimised energies and the statistical procedure for those amino acid residues with three side chain rotameric states.
<table>
<thead>
<tr>
<th>ELEMENTS FOR POLYAMINOACIDS WITH DIFFERENT MINIMIZED ENERGETIC WEIGHTINGS AND A 10° INTEGRATION INTERVAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Alanine</td>
</tr>
<tr>
<td>$\beta$-methylalanine: SIDE CHAIN</td>
</tr>
<tr>
<td>gauche$^+$</td>
</tr>
<tr>
<td>gauche$^-$</td>
</tr>
<tr>
<td>trans</td>
</tr>
<tr>
<td>composite</td>
</tr>
<tr>
<td>statistical</td>
</tr>
<tr>
<td>Valine:</td>
</tr>
<tr>
<td>SIDE CHAIN</td>
</tr>
<tr>
<td>gauche$^+$</td>
</tr>
<tr>
<td>gauche$^-$</td>
</tr>
<tr>
<td>trans</td>
</tr>
<tr>
<td>composite</td>
</tr>
<tr>
<td>statistical</td>
</tr>
<tr>
<td>$\beta$-methylvaline</td>
</tr>
</tbody>
</table>
TABLE IV.4  $<T>$ ELEMENTS FOR POLYAMINOACIDS WITH DIFFERENT MINIMIZED ENERGETIC
WEIGHTINGS AND A 30° INTEGRATION INTERVAL

<table>
<thead>
<tr>
<th></th>
<th>$a_{11}$</th>
<th>$a_{12}$</th>
<th>$a_{13}$</th>
<th>$a_{21}$</th>
<th>$a_{22}$</th>
<th>$a_{23}$</th>
<th>$a_{31}$</th>
<th>$a_{32}$</th>
<th>$a_{33}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alanine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.01</td>
<td>0.75</td>
<td>-0.12</td>
<td>-0.67</td>
<td>-0.07</td>
<td>0.66</td>
<td>-0.13</td>
<td>-0.27</td>
</tr>
<tr>
<td><strong>β-methylalanine:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche$^+$</td>
<td>0.69</td>
<td>0.38</td>
<td>0.55</td>
<td>0.37</td>
<td>-0.82</td>
<td>0.04</td>
<td>0.50</td>
<td>0.16</td>
<td>-0.70</td>
</tr>
<tr>
<td>gauche$^-$</td>
<td>0.30</td>
<td>-0.24</td>
<td>0.84</td>
<td>-0.16</td>
<td>-0.59</td>
<td>-0.29</td>
<td>0.63</td>
<td>-0.06</td>
<td>-0.21</td>
</tr>
<tr>
<td>trans</td>
<td>0.34</td>
<td>-0.06</td>
<td>0.76</td>
<td>-0.17</td>
<td>-0.57</td>
<td>-0.15</td>
<td>0.59</td>
<td>-0.12</td>
<td>-0.21</td>
</tr>
<tr>
<td>composite</td>
<td>0.37</td>
<td>-0.05</td>
<td>0.76</td>
<td>-0.12</td>
<td>-0.58</td>
<td>-0.16</td>
<td>0.59</td>
<td>-0.08</td>
<td>-0.24</td>
</tr>
<tr>
<td>statistical</td>
<td>0.35</td>
<td>-0.09</td>
<td>0.78</td>
<td>-0.13</td>
<td>-0.59</td>
<td>-0.18</td>
<td>0.60</td>
<td>-0.08</td>
<td>-0.24</td>
</tr>
<tr>
<td><strong>Valine:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche$^+$</td>
<td>0.72</td>
<td>0.39</td>
<td>0.57</td>
<td>0.40</td>
<td>-0.89</td>
<td>0.11</td>
<td>0.56</td>
<td>0.15</td>
<td>-0.80</td>
</tr>
<tr>
<td>gauche$^-$</td>
<td>0.55</td>
<td>0.07</td>
<td>0.78</td>
<td>0.24</td>
<td>-0.69</td>
<td>-0.27</td>
<td>0.54</td>
<td>0.33</td>
<td>-0.37</td>
</tr>
<tr>
<td>trans</td>
<td>0.31</td>
<td>-0.24</td>
<td>0.86</td>
<td>-0.17</td>
<td>-0.68</td>
<td>-0.27</td>
<td>0.70</td>
<td>-0.07</td>
<td>-0.26</td>
</tr>
<tr>
<td>composite</td>
<td>0.35</td>
<td>-0.18</td>
<td>0.83</td>
<td>-0.12</td>
<td>-0.70</td>
<td>-0.24</td>
<td>0.68</td>
<td>-0.06</td>
<td>-0.30</td>
</tr>
<tr>
<td>statistical</td>
<td>0.35</td>
<td>-0.17</td>
<td>0.83</td>
<td>-0.11</td>
<td>-0.70</td>
<td>-0.24</td>
<td>0.68</td>
<td>-0.05</td>
<td>-0.31</td>
</tr>
<tr>
<td><strong>β-methylvaline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.04</td>
<td>0.78</td>
<td>0.38</td>
<td>-0.81</td>
<td>-0.28</td>
<td>0.63</td>
<td>0.46</td>
<td>-0.52</td>
</tr>
<tr>
<td></td>
<td>$a_{11}$</td>
<td>$a_{12}$</td>
<td>$a_{13}$</td>
<td>$a_{21}$</td>
<td>$a_{22}$</td>
<td>$a_{23}$</td>
<td>$a_{31}$</td>
<td>$a_{32}$</td>
<td>$a_{33}$</td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Glycine</td>
<td>0.37</td>
<td>0.07</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.20</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.38</td>
<td>-0.03</td>
<td>0.76</td>
<td>-0.13</td>
<td>-0.66</td>
<td>-0.09</td>
<td>0.67</td>
<td>-0.13</td>
<td>-0.27</td>
</tr>
<tr>
<td>β-methylalanine:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche$^+$</td>
<td>0.73</td>
<td>0.44</td>
<td>0.49</td>
<td>0.42</td>
<td>-0.84</td>
<td>0.05</td>
<td>0.44</td>
<td>0.16</td>
<td>-0.80</td>
</tr>
<tr>
<td>gauche$^-$</td>
<td>0.30</td>
<td>-0.24</td>
<td>0.84</td>
<td>-0.17</td>
<td>-0.63</td>
<td>-0.28</td>
<td>0.66</td>
<td>-0.07</td>
<td>-0.23</td>
</tr>
<tr>
<td>trans</td>
<td>0.35</td>
<td>-0.12</td>
<td>0.77</td>
<td>-0.10</td>
<td>-0.53</td>
<td>-0.19</td>
<td>0.56</td>
<td>-0.06</td>
<td>-0.23</td>
</tr>
<tr>
<td>composite</td>
<td>0.37</td>
<td>-0.09</td>
<td>0.76</td>
<td>-0.08</td>
<td>-0.57</td>
<td>-0.18</td>
<td>0.57</td>
<td>-0.05</td>
<td>-0.28</td>
</tr>
<tr>
<td>statistical</td>
<td>0.36</td>
<td>-0.12</td>
<td>0.77</td>
<td>-0.00</td>
<td>-0.59</td>
<td>-0.20</td>
<td>0.59</td>
<td>-0.05</td>
<td>-0.28</td>
</tr>
<tr>
<td>Valine:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche$^+$</td>
<td>0.69</td>
<td>0.41</td>
<td>0.58</td>
<td>0.32</td>
<td>-0.89</td>
<td>0.26</td>
<td>0.64</td>
<td>0.00</td>
<td>-0.75</td>
</tr>
<tr>
<td>gauche$^-$</td>
<td>0.66</td>
<td>0.23</td>
<td>0.69</td>
<td>0.39</td>
<td>-0.78</td>
<td>-0.20</td>
<td>0.50</td>
<td>0.39</td>
<td>-0.61</td>
</tr>
<tr>
<td>trans</td>
<td>0.33</td>
<td>-0.21</td>
<td>0.86</td>
<td>-0.13</td>
<td>-0.72</td>
<td>-0.26</td>
<td>0.72</td>
<td>-0.04</td>
<td>-0.28</td>
</tr>
<tr>
<td>composite</td>
<td>0.36</td>
<td>-0.17</td>
<td>0.84</td>
<td>-0.10</td>
<td>-0.73</td>
<td>-0.23</td>
<td>0.71</td>
<td>-0.04</td>
<td>-0.31</td>
</tr>
<tr>
<td>statistical</td>
<td>0.36</td>
<td>-0.17</td>
<td>0.84</td>
<td>-0.09</td>
<td>-0.73</td>
<td>-0.22</td>
<td>0.71</td>
<td>-0.03</td>
<td>-0.31</td>
</tr>
<tr>
<td>β-methylvaline</td>
<td>0.63</td>
<td>0.19</td>
<td>0.75</td>
<td>0.30</td>
<td>-0.95</td>
<td>-0.01</td>
<td>0.71</td>
<td>0.23</td>
<td>-0.65</td>
</tr>
</tbody>
</table>
### TABLE IV.6 <T> ELEMENTS FOR POLYAMINOACIDS WITH DIFFERENT NON-MINIMIZED ENERGETIC WEIGHTINGS AND A 30° INTEGRATION INTERVAL

<table>
<thead>
<tr>
<th></th>
<th>a₁₁</th>
<th>a₁₂</th>
<th>a₁₃</th>
<th>a₂₁</th>
<th>a₂₂</th>
<th>a₂₃</th>
<th>a₃₁</th>
<th>a₃₂</th>
<th>a₃₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>0.37</td>
<td>0.11</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.20</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.39</td>
<td>0.01</td>
<td>0.76</td>
<td>-0.13</td>
<td>-0.67</td>
<td>-0.06</td>
<td>0.67</td>
<td>-0.13</td>
<td>-0.27</td>
</tr>
<tr>
<td>β-methylalanine:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche⁺</td>
<td>0.72</td>
<td>0.42</td>
<td>0.54</td>
<td>0.40</td>
<td>-0.86</td>
<td>0.07</td>
<td>0.49</td>
<td>0.17</td>
<td>-0.79</td>
</tr>
<tr>
<td>gauche⁻</td>
<td>0.30</td>
<td>-0.22</td>
<td>0.85</td>
<td>-0.18</td>
<td>-0.64</td>
<td>-0.28</td>
<td>0.67</td>
<td>-0.08</td>
<td>-0.24</td>
</tr>
<tr>
<td>trans</td>
<td>0.35</td>
<td>-0.05</td>
<td>0.76</td>
<td>-0.14</td>
<td>-0.51</td>
<td>-0.17</td>
<td>0.52</td>
<td>-0.08</td>
<td>-0.19</td>
</tr>
<tr>
<td>composite</td>
<td>0.39</td>
<td>-0.02</td>
<td>0.74</td>
<td>-0.08</td>
<td>-0.55</td>
<td>-0.16</td>
<td>0.53</td>
<td>-0.04</td>
<td>-0.25</td>
</tr>
<tr>
<td>statistical</td>
<td>0.37</td>
<td>-0.07</td>
<td>0.77</td>
<td>-0.11</td>
<td>-0.58</td>
<td>-0.19</td>
<td>0.57</td>
<td>-0.06</td>
<td>-0.25</td>
</tr>
<tr>
<td>Valine:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDE CHAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gauche⁺</td>
<td>0.72</td>
<td>0.41</td>
<td>0.56</td>
<td>0.40</td>
<td>-0.90</td>
<td>0.13</td>
<td>0.56</td>
<td>0.13</td>
<td>-0.80</td>
</tr>
<tr>
<td>gauche⁻</td>
<td>0.59</td>
<td>0.10</td>
<td>0.77</td>
<td>0.32</td>
<td>-0.75</td>
<td>-0.27</td>
<td>0.56</td>
<td>0.40</td>
<td>-0.47</td>
</tr>
<tr>
<td>trans</td>
<td>0.31</td>
<td>-0.23</td>
<td>0.87</td>
<td>-0.18</td>
<td>-0.72</td>
<td>-0.27</td>
<td>0.73</td>
<td>-0.09</td>
<td>-0.27</td>
</tr>
<tr>
<td>composite</td>
<td>0.32</td>
<td>-0.21</td>
<td>0.86</td>
<td>-0.16</td>
<td>-0.73</td>
<td>-0.26</td>
<td>0.72</td>
<td>-0.08</td>
<td>-0.29</td>
</tr>
<tr>
<td>statistical</td>
<td>0.32</td>
<td>-0.21</td>
<td>0.86</td>
<td>-0.16</td>
<td>-0.72</td>
<td>-0.26</td>
<td>0.72</td>
<td>-0.07</td>
<td>-0.29</td>
</tr>
<tr>
<td>β-methylvaline</td>
<td>0.62</td>
<td>0.06</td>
<td>0.78</td>
<td>0.39</td>
<td>-0.87</td>
<td>-0.25</td>
<td>0.66</td>
<td>0.46</td>
<td>-0.56</td>
</tr>
</tbody>
</table>
Fig. IV.1 Characteristic ratios as functions of \( n_p \) for homopolypeptides. The limiting values of \( C_\infty \) as \( n_p \to \infty \) are marked.
Fig. IV.2 Characteristic ratio as a function of $n_p$ for poly-L-β-methylvaline compared to poly-L-alanine. The limiting values of $C_\infty$ as $n_p \to \infty$ are marked.
B. **Unperturbed Chain Dimensions of Poly-L-hydroxyethylglutamine and its Copolymers**

$\langle T \rangle$ matrices for glycine, alanine and $\beta$-methylalanine which were judged to be the most accurate were selected from Tables IV.3 and IV.5 to be used to evaluate the characteristic ratios of the copolypeptides. These matrices were produced by weighting with minimized energy maps at $10^\circ (\phi, \psi)$ intervals. For the residues with side chains, i.e. alanine and $\beta$-methylalanine, these maps contained minimized energy values and for $\beta$-methylalanine a statistical method of energetic weighting was used to incorporate contributions from all three side chain rotameric states. The elements of these matrices together with those for the average transformation matrix for poly-L-hydroxyethylglutamine are shown in Table IV.7.

The characteristic ratio of the randomly coiled homopolypeptide, poly-L-hydroxyethylglutamine, calculated from the $\langle T \rangle$ shown in Table IV.7 is 9.75. The characteristic ratios for the three copolypeptides $(\text{HEG-Gly-Gly})_n$, $(\text{HEG-Ala-Ala})_n$ and $(\text{HEG-}\beta\text{-methylAla-}\beta\text{-methylAla})_n$ calculated for increasing values of $n_p$ (where $n_p = 3n$) are shown in Figure IV.3. The limiting values $C_\infty$ as $n_p \rightarrow \infty$ are marked on the Figure and are 2.27, 8.56 and 7.73, respectively.

3. **DISCUSSION**

A. **The Effect on Calculated Characteristic Ratios of Changes in the Energy Parameters and Covalent Geometry**

The effect of using more 'refined' energy parameters and the most recent covalent geometry on the unperturbed dimensions of polypeptide chains is best
Fig. IV.3 Characteristic ratios as functions of $n_p$ for copolymers containing the hydroxyethylglutamine (HEG) residue. The limiting values of $C_\infty$ as $n_p \to \infty$ are marked.
<table>
<thead>
<tr>
<th>Residue</th>
<th>$a_{11}$</th>
<th>$a_{12}$</th>
<th>$a_{13}$</th>
<th>$a_{21}$</th>
<th>$a_{22}$</th>
<th>$a_{23}$</th>
<th>$a_{31}$</th>
<th>$a_{32}$</th>
<th>$a_{33}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEG</td>
<td>0.393</td>
<td>-0.037</td>
<td>0.770</td>
<td>-0.091</td>
<td>-0.699</td>
<td>-0.096</td>
<td>0.698</td>
<td>-0.086</td>
<td>-0.299</td>
</tr>
<tr>
<td>GLY</td>
<td>0.368</td>
<td>0.074</td>
<td>0.000</td>
<td>-0.166</td>
<td>-0.451</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.202</td>
</tr>
<tr>
<td>ALA</td>
<td>0.379</td>
<td>-0.023</td>
<td>0.752</td>
<td>-0.126</td>
<td>-0.652</td>
<td>-0.089</td>
<td>0.661</td>
<td>-0.124</td>
<td>-0.271</td>
</tr>
<tr>
<td>β-METHYL ALANINE</td>
<td>0.341</td>
<td>-0.144</td>
<td>0.782</td>
<td>-0.123</td>
<td>-0.591</td>
<td>-0.202</td>
<td>0.614</td>
<td>-0.077</td>
<td>-0.257</td>
</tr>
</tbody>
</table>
demonstrated using glycine where side chain considerations do not apply. The values for \( C_m \) obtained in this study (2.12 and 2.13) are slightly lower than those obtained by earlier workers (2.16 (Brant et al., 1967) and 2.17 (Tanaka and Nakajima, 1971a)). Since this decrease is also reflected in the characteristic ratio of the freely rotating chain (we obtained a value of 1.89 compared with the earlier value of 1.93 (P.J. Flory, 1969) the new geometry, rather than the refined energy parameters, has the greater effect on the flexibility of the polymer chain.

B. The Choice of Integration Interval in Characteristic Ratio Calculations

The earliest calculations on homopolypeptides by P.J. Flory's group used discrete values \( \phi \) and \( \psi \) at 30° intervals to evaluate the characteristic ratios of the polymers, since a larger increment of 60° yielded comparable results (Brant and Flory, 1965b). Since then most workers have chosen this interval in their calculations and where smaller increments of say 20° (Brant et al., 1967) or 10° (Miller and Goebel, 1968; Burgess et al., 1975) have been used, no justification for this refinement has been given. Our calculations show that for the majority of amino acid residues with characteristic ratios of 8 ± 2, the 30° and 10° grids of the \((\phi,\psi)\) surface give results that fall within 10% of each other. As might be expected, the exceptions to this result appear to be for residues with very restricted \((\phi,\psi)\) surfaces where the regions that most heavily weight the characteristic ratio are in one small area of the map.
Thus, the $C_\infty$ of the gauche$^+$ rotamer for valine is greatly reduced by the choice of a $10^\circ$ over a $30^\circ$ grid, whereas both the gauche$^-$ rotamer for this residue (see Table IV.1) and the $\beta$-methylvaline residue (see Table IV.2) show very significant increases in their $C_\infty$ values when a $10^\circ$ grid is used. It seems that in the majority of cases, the use of a $30^\circ$ interval is adequate, except for residues with characteristic ratios that depart from the "alanine-like" value of 8-9.

C. The Effect of Incorporating Energy Minimization in the Energetic Statistical Weighting

Energy minimization of the side chain conformations of the homopolymers influences the mean square end-to-end distance in a similar way to the choice of integration interval. Thus, for those residues with "alanine-like" polymer chain dimensions, the extra side chain freedom affects the $C_\infty$ values by less than $10\%$, but where extended polymer chain dimensions are involved, e.g. for $\beta$-methylvaline and for the gauche rotamers of valine, the effect is to dramatically decrease the characteristic ratio. Other factors involved in the calculations can compound the error arising from the use of non-minimized energy values. For example, the $C_\infty$ for the gauche$^+$ rotamer of $\beta$-methylvaline weighted with different choices of integration interval show only a $10\%$ difference when minimized energy surfaces are used, whereas the difference is nearly $30\%$ where minimized energy values are not used.
The effect of using minimized energy values on all residues except β-methylalanine is to decrease the dimensions to some extent. For β-methylalanine, however, there is a slight increase in the C\textsubscript{0} of the trans rotamer; since this rotamer is strongly preferred (see Table III.3) there is a net increase in the chain dimensions of the molecule weighted for all three rotamers.

D. Treatment of Side Chains and Comparison with Experimental Results

Experimental measurements on homopolyamino acids in various solvents have shown that polymers composed of different residues have similar chain dimensions (Terbojevitch \textit{et al.}, 1967; Brant and Flory, 1965a; Fujita \textit{et al.}, 1966), which, within the greatest quoted (Fujita \textit{et al.}, 1966) experimental error of 16\%, are remarkably alike to the theoretically-predicted (Brant and Flory, 1965b) dimensions for alanine (see Table IV.8). Thus, it was postulated (Brant and Flory, 1965b) that substituents along the side chain beyond the C\textsubscript{β} atom exert little influence on the conformational space available to the backbone and that bond rotations beyond the C\textsuperscript{α}-C\textsubscript{β} bond need not be taken into account in these calculations.

A later study (Miller and Goebel, 1968) compared a 'structureless' side chain approach, where the side chain atoms were represented by enlarged C\textsubscript{β} atoms, with an approach where the C\textsuperscript{γ} atoms, but not their adjoined atoms, were included in the calculations. This study demonstrated a significant reduction (nearly 20\%) in the unperturbed
**TABLE IV.8 EXPERIMENTALLY DERIVED CHARACTERISTIC RATIOS OF SOME POLYPEPTIDES PUBLISHED BY OTHER WORKERS**

<table>
<thead>
<tr>
<th>Polymer</th>
<th>Characteristic Ratio</th>
<th>Worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Lys)_n</td>
<td>8.6 ± 0.9</td>
<td>Brant &amp; Flory (1965a)</td>
</tr>
<tr>
<td>(Glu)_n</td>
<td>8.8 ± 0.9</td>
<td></td>
</tr>
<tr>
<td>(γ-Benzyl-Glu)_n</td>
<td>8.8 ± 0.9</td>
<td>Fujita et al. (1966)</td>
</tr>
<tr>
<td>(β-Benzyl-Asp)_n</td>
<td>9.6 ± 1.0</td>
<td></td>
</tr>
<tr>
<td>(γ-Benzyl-Glu)_n</td>
<td>7.5 ± 1.2</td>
<td></td>
</tr>
<tr>
<td>(γ-Ethyl-Glu)_n</td>
<td>8.2 ± 0.3</td>
<td>Terbojevitch et al. (1967)</td>
</tr>
<tr>
<td>(HEG)_n</td>
<td>10.1 ± 1.0</td>
<td>Mattice &amp; Lo (1972)</td>
</tr>
<tr>
<td>(Glu&lt;sub&gt;0.35&lt;/sub&gt;Lys&lt;sub&gt;0.35&lt;/sub&gt;Gly&lt;sub&gt;0.27&lt;/sub&gt;)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>8.1</td>
<td>Rao et al. (1975)</td>
</tr>
<tr>
<td>(Glu&lt;sub&gt;0.33&lt;/sub&gt;Lys&lt;sub&gt;0.21&lt;/sub&gt;Gly&lt;sub&gt;0.46&lt;/sub&gt;)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>(Glu&lt;sub&gt;0.87&lt;/sub&gt;Gly&lt;sub&gt;0.13&lt;/sub&gt;)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>4.0</td>
<td>Miller et al. (1967)</td>
</tr>
<tr>
<td>(Glu&lt;sub&gt;0.75&lt;/sub&gt;Gly&lt;sub&gt;0.25&lt;/sub&gt;)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>4.1</td>
<td></td>
</tr>
</tbody>
</table>

N.B. Values of n are often not quoted in the above references.
dimensions of chain molecules with branched side chains when the $\gamma$ and its hydrogen atoms were explicitly taken into account. When weighting with the energy maps of the rotamers of residues with unbranched side chains, however, these workers used a composite rather than a statistical method in which they neglected the fact that there are $(\phi,\psi)$ values for which more than one side chain rotamer is equally probable and for residues with branched side chains $\chi^1$ was fixed in only one rotameric state, i.e. 180°. Nor did they allow flexibility about the 'staggered' $\chi^1$ values by allowing these values to vary so as to minimize the conformational energy. Thus, even using a semi-structured representation, the characteristic ratios of residues with branched side chains are significantly different (27%) from those of alanine (Miller and Goebel, 1968).

Our results show that a composite weighting with only one energy value for each $(\phi,\psi)$ pair is not as inaccurate as one might expect (compare the $C_\infty$ obtained with composite versus statistical weightings for $\beta$-methylalanine and valine in Table VI.1). The results obtained using either method are very close and fall well within the errors involved in estimating the unperturbed dimensions experimentally. In our calculations, all of the atoms in the residue were explicitly defined, and the effect of this, plus the use of the most recent (Momany et al., 1975) covalent peptide geometry and energy parameters is to increase slightly the characteristic ratios of alanine and valine in this study (8.3 and 11.2 respectively) compared to those of Miller and Goebel (8.0 and 10.7 respectively). However, by allowing the side chains more freedom by including minimization in our calculations, the
chain dimensions for both β-methylalanine (7.0) and valine (9.2) fall within 14% of those for alanine, in accord with Brant and Flory's hypothesis (1965b). The fact that they are not identical with that of alanine may imply that a three state approximation does not adequately describe the rotational degrees of freedom of the side chain bonds.

A characteristic ratio for poly-L-hydroxyethylglutamine of 9.8 is also consistent with the hypothesis of Brant and Flory that all amino acid residues with side chains longer than alanine have similar dimensions. This value is somewhat higher than the best estimates for alanine, β-methylalanine and valine reported in this chapter, but due to the large number of side chain conformations of this residue less rigorous techniques for the energetic weighting and less side chain freedom was allowed for this molecule. Using a 30° grid and no minimization for polyalanine (the conditions used for poly-L-hydroxyethylglutamine) a characteristic value of 8.8 was obtained which is about 10% higher than the most accurate estimate for this polymer of 8.1 (see Table IV.1). Compared to this value (8.8) the computed chain dimensions of poly-L-hydroxyethylglutamine are well within the reported accuracy for experimental estimates of the characteristic ratio for various homopolypeptides (see Table IV.8) and is in excellent agreement with an experimentally derived characteristic ratio of 10 ± 1 reported by Mattice and Lo (1972) for poly-L-hydroxyethylglutamine. This result was produced for the polymer in water at 30°. Under these conditions, circular dichroism showed
that the polymer was randomly coiled and in the unperturbed state to which the theoretical predictions apply. Thus, it can be said that the hydrodynamic behaviour of the homopolypeptides considered in this study is dominated by intraresidue interactions and that the behaviour of the polymer backbone is not greatly influenced by side chain atoms remote from it. The only exception to this generalisation is poly-β-methylvaline which, of the residues with side chains that were chosen for this study is the only one with an atypical characteristic ratio. The presence of a tert-butyl group on the Cβ atom has a profound steric effect on the conformation of the backbone. Only very extended conformations are available to this molecule on the (φ,ψ) energy surface (see Fig.III.2(b)) and this dramatically increases the chain dimensions.

It is possible that poly-β-methylvaline is unable to adopt a random coil [i.e. inhabit a wide spectrum of configurations] in any solvent. The only other homopolypeptide that has been reported to have such extended computed chain dimensions is that of poly-L-proline (Schimmel and Flory, 1967; Tanaka and Scheraga, 1975d) and this is considered to be a consequence of the confinement of all residues to one conformational energy well. Short sequences of trans poly-L-proline have computed chain dimensions that approximate to those expected for a rigid poly-L-proline II (Cowan and McGavin, 1955) helix. Thus \(<r^2>^\frac{1}{2}\) calculated for a sequence of 16 residues is 98% of its value for the rigid helix; for 128 residues it is 80% of the length of the rigid helix.
(Schimmel and Flory, 1967). Dilute solution viscosity of trans poly-L-proline II for short chains (50 residues, Steinberg et al., 1960) indicate that its length approximates to that of the Cowan-McGavin helix, whereas random coil behaviour is indicated for long chain polymers. The configurational character of trans poly-L-proline may well be better represented (P.J. Flory, 1969) by the Porod-Krolky model (a 'worm-like' chain which incorporates the concept of continuous curvature of the chain skeleton, the direction of curvature at any point of the trajectory being random; it has been used successfully to represent 'stiff' chains in synthetic polymers). The unperturbed dimensions of poly-L-proline derived experimentally and theoretically have not always been in agreement. This may not be due entirely to the pre-disposition for cis-trans isomerism about the peptide bond in this polymer (see the brief review in the Introduction to this chapter) but could be attributed to the difficulties described above in the theoretical treatment of rigid chains. Since β-methylvaline would not be expected to undergo cis-trans isomerism of the type seen in proline peptides, a polymer composed of this residue would provide a simpler model to compare the hydrodynamic behaviour with theoretical treatments of a randomly coiled rigid chain, than does poly-L-proline.

Such a study on poly-β-methylvaline would also be expected to be a more severe test of the various approximations inherent in these calculations than would a flexible poly-peptide. It was demonstrated that slight differences in the conformational energy maps used to weight the transformation matrices can make differences in the characteristic ratio
which are much more profound the more restricted the map. For example, the effects of grid size and energy minimisation on characteristic ratios are much greater for poly-β-methylvaline (see Table IV.2) than for polyalanine (Table IV.1).

In addition, in generating $\langle r^2 \rangle_0$, the statistical probability of runs of identical ($\phi, \psi$) values (and therefore runs of helix) occurring for any polypeptide is greatly increased if the overall set of ($\phi, \psi$) pairs to choose from is very small. This effect would result in extended chain dimensions as $n \to \infty$ for most polypeptides, but would be less pronounced for poly-β-methylvaline because the limited conformational space available incorporates regions of opposite chirality (compare the region of low energy in Fig.III.2(b) with pitch and residue per turn plots of Ramakrishnan (1964)). It should be possible to test these speculations by deriving experimentally the dimensions for increasingly longer chains of this polymer based on a random coil model and comparing these values with the theoretical results obtained in this study.

E. The Unperturbed Dimensions of Copolypeptides Containing Hydroxyethylglutamine

The chain dimensions of the copolymers containing hydroxyethylglutamine are only slightly higher than the dimensions of the major components. The characteristic ratios of polyglycine, polyalanine and poly-β-methylalanine are 2.12, 8.06 and 7.02, respectively; those of the sequential copolypeptides containing two of each of these residues to one of the hydroxyethylglutamine are 2.27, 8.56 and 7.73.
Although the β-methylalanine residue has a bulkier side chain and its conformational energy space is thus slightly more restricted than that of alanine (see Fig.III.1(a) and (b)), poly-β-methylalanine has a slightly lower characteristic ratio than polyalanine which is reflected in the chain dimensions of the copolymers containing these residues.

However, chain dimensions of polyamino acids are influenced not so much by the total space available to the residue, as by the energetic distributions of conformations within that space. The results reported in this chapter have shown that differences in the methods and parameters used to define the conformational energy space and the methods of weighting the chain dimensions using such data can give markedly different values for the characteristic ratio. The slightly lower dimensions of the β-methylalanine-containing polymers, therefore, may be the result of the approximations involved in using a three-state model for side chain conformations of this molecule. However, the computed chain dimensions of the homopolymer of β-methylalanine are as close to those of polyalanine as the experimental limit in experimentally estimating chain dimensions. The published results for theoretical estimates of the characteristic ratios of sequential copolypeptides have so far been concerned with the effect of N-methyl amino acids (Tanaka and Nakajima, 1971b) or proline (Schimmel and Flory, 1968; Mattice and Mandelkern, 1971b) on chain dimensions. Our findings are consistent with those of Miller et al. (1967) who showed that a small percentage of randomly distributed glycine has a large
effect on the chain dimensions of randomly coiled polyalanine. A random copolymer containing 66% glycine would be expected from their results to have a characteristic ratio of 2.4.

There are few experimental estimates of the unperturbed dimensions of copolypeptides that can be confidently compared with the theoretical results presented in this chapter. The estimates that are available are from studies exploring the effect of the incorporation of glycine residues into copolypeptides and refer to random rather than sequential copolymers (these are shown in Table IV.8). The results of Miller et al. (1967) are most consistent with our results on poly(HEG-Gly-Gly) and these experimental results are in good agreement with their own theoretical predictions quoted above. The work of Rao et al. (1975) is not in accord either with the work of Miller et al. or with ours. Their higher (than expected) values for the chain dimensions of glycine-containing copolymers may be due to 'blocks' of amino acid residues being incorporated during polymerisation so that their copolymers may not be entirely random; a series of glutamic acid residues could take up the α-helical conformation which would increase the observed chain dimensions.

Other published experimental work is on copolypeptides containing proline and hydroxyproline, e.g. the work by Mattice and Mandelkern (1971b) on the hydrodynamic behaviour of sequential copolypeptides containing either proline or hydroxyproline as one of the monomer units and this is in good agreement with the theoretical estimates on similar copolymers made by the Flory group (Schimmel and Flory, 1967, 1968; Brant et al., 1967).
Lapanje and Tanford (1967) measured the viscosity and osmotic pressure of protein chains under reducing (0.1 M β-mercaptoethanol) and randomising (6 M guanidine hydrochloride) conditions and from these hydrodynamic parameters estimated a C.R. of $4.2 \pm 0.7$ for denatured proteins. The proteins that they used contained 2-10% glycine and 3-6% proline. Miller et al. (1967) showed a theoretical reduction of the characteristic ratio to 6.6 when glycine is present in a polypeptide at the level of 10%. The incorporation of proline into polypeptides also depresses chain dimensions (Schimmel and Flory, 1968) and the presence of 6% proline residues in the chain reduces the chain dimensions by 3 to 20% depending on which configuration is adopted by the proline residue about the peptide bond. However, even if all the proline residues were in the unfavoured cis configuration, the characteristic ratio would still not be as low as that measured by Lapanje and Tanford. A more realistic model of a denatured protein chain than that of the earlier work of the Flory group simulating actual amino acid sequences with 9 types of side chain was treated theoretically by Miller and Goebel (1968) and their results are in reasonable agreement with the experimental results of Lapanje and Tanford, although the latter are still somewhat lower by 15%.

4. CONCLUSIONS

In order to establish more rigorously the effect of side chain length and branching on the theoretically derived unperturbed dimensions of homopolypeptides, the most recently revised covalent geometry has been used to
calculate the characteristic ratios of polymers composed solely of glycine, alanine, β-methylalanine, valine or β-methylvaline. All atoms in the side chains were included in the calculations and different methods of weighting for multiple side chain conformations and of incorporating side chain flexibility were compared. The effects of weighting with the conformational space of the backbone at 30° and 10° intervals of φ and ψ were compared. The characteristic ratios of alanine, β-methylalanine and valine using the most rigorous of the methods explored were found to be very similar (8.06, 7.02 and 9.21). That of poly-L-hydroxyethylglutamine using a 30° grid with no minimization was 9.75. The side chain of β-methylvaline imposed profound steric restrictions on the backbone which resulted in a slowly converging characteristic ratio (94.8 at a chain length of 500 residues).

The characteristic ratios of the copolypeptides (HEG-X-X)ₙ, where X = glycine, alanine or β-methylalanine have been derived and are, respectively, 2.27, 8.56 and 7.73.

With the exception of β-methylvaline these results are consistent with P.J. Flory's hypothesis that the side chain atoms further removed from the backbone than the Cβ atom need not be considered when calculating the unperturbed dimension of most homopolypeptides. Comparison with the experimental results by other workers shows that intraresidue interactions alone are sufficient to compute the chain dimensions of polypeptides and to account for their hydrodynamic behaviour.
CHAPTER V

STUDIES ON PROCEDURES FOR CONVERTING
POLYGLUTAMIC ACID TO POLYHYDROXYETHYLGLUTAMINE
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDC</td>
<td>1-benzyl,3-dimethylaminopropylcarbodiimide</td>
</tr>
<tr>
<td>Bz</td>
<td>Benzyl</td>
</tr>
<tr>
<td>DCC</td>
<td>Dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DCU</td>
<td>Dicyclohexylurea</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulphoxide</td>
</tr>
<tr>
<td>EDMAP</td>
<td>N-ethyl,N'-(3-dimethylaminopropyl)carbodiimide</td>
</tr>
<tr>
<td>HEG</td>
<td>N(^5)-(2 hydroxyethyl)-L-glutamine</td>
</tr>
<tr>
<td>HOSu</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>M.W.</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>(\bar{M}_w)</td>
<td>Average molecular weight</td>
</tr>
<tr>
<td>Mn</td>
<td>Number-average molecular weight</td>
</tr>
<tr>
<td>(\bar{M}_v)</td>
<td>Viscosity-average molecular weight</td>
</tr>
<tr>
<td>(\bar{M}_w)</td>
<td>Weight-average molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>ORD</td>
<td>Optical rotatory dispersion</td>
</tr>
<tr>
<td>OSu</td>
<td>N-hydroxysuccinimide ester</td>
</tr>
<tr>
<td>(O^t)Bu</td>
<td>tert-butyl ester</td>
</tr>
<tr>
<td>PAP</td>
<td>Pyroglutanyl aminopeptidase</td>
</tr>
<tr>
<td>Pr</td>
<td>Propionyl</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin-layer chromatography</td>
</tr>
<tr>
<td>Z</td>
<td>Benzylxoycarbonyl</td>
</tr>
</tbody>
</table>
INTRODUCTION

Since polyhydroxyalkylglutamines were first prepared by the aminolysis of poly(γ-benzylglutamate) by Lotan et al. (1965) (see Rn.1), this method has been used exclusively by workers who have followed their procedure either exactly (Adler et al., 1968; Joubert et al., 1970) or with slight modifications (Mattice and Lo, 1972; and a group of workers at Cornell University: Von Dreele et al., 1971; Platzer et al., 1972; Alter et al., 1973; Maxfield et al., 1975).

This method of synthesis, however, involves heating the parent polymer at 60° for up to 9 days with high concentrations of the very basic amino alcohol (10 to 20 ml/g of polymer). These conditions lead to two serious disadvantages which are particularly important when the resultant polymer is to be the subject of experimental measurements on polymer dimensions.

It was recognised by Lotan et al. that the conversion from benzyl to hydroxyalkyl side chains was accompanied by a decrease in the degree of polymerisation of the chain. This decrease was also observed by the Cornell group who used these water soluble polyamino acids as "host" polymers for the determination of the helix-coil stability constants for the 20 naturally occurring "guest" amino acids in water. For alanine-containing co-polymers, for example, (Platzer et al., 1972) reported chain fission of up to 60% occurred; this effect was ascribed to a transaminolysis reaction in which the aminopropanol reacts with the peptide backbone as well as the side chain ester (see Rn.2). For tyrosine containing co-polymers, (Scheule et al., 1976), the extent
Aminolysis of poly(γ-benzylglutamate) by ethanolanine to yield polyhydroxyethylglutamine:

\[
\begin{align*}
\text{OBz} & \quad + \quad n\text{H}_2\text{N(\(\text{CH}_2\))}_2\text{OH} \\
\text{CO} & \quad \rightarrow \\
(\text{CH}_2)_2 & \quad \rightarrow \\
\text{NH} & \quad \rightarrow \\
-\text{NH} & \quad \rightarrow \\
-\text{CH} & \quad \rightarrow \\
\text{CO} & \quad \rightarrow \\
\text{H}_2\text{N(\(\text{CH}_2\))}_2\text{OH} & \quad + \quad n\text{BzOH}
\end{align*}
\]

Side reaction by which chain fission can occur:

\[
\begin{align*}
-\left[\text{R} - \text{NH} - \text{CH} - \text{CO} - \text{NH} - \text{CH} - \text{CO}\right]_n & \quad \rightarrow \\
+ \quad \text{H}_2\text{N(\(\text{CH}_2\))}_2\text{OH} & \quad \text{NH} - \text{CH} - \text{CO} - \text{NH} - (\text{CH}_2)_2\text{OH}_x \\
& \quad + \quad \text{NH}_2 - \text{CH} - \text{CO}\right]_{n-x}
\end{align*}
\]

of chain fission reported was such that the degrees of polymerisation in the polymer products were from 9 to 35% of the parent polymers. Lowering the temperature to 40°C until 90% of the aminolysis reaction was complete, and then increasing it to 60° improved the recovery slightly, increasing the degree of polymerisation from 9-15% to 17-40% for two polymers of similar composition and degree of polymerisation.

Although no detailed molecular weight distribution data have been reported in synthetic studies such as those reported above on hydroxyalkylglutamine, it seems reasonable to suppose that random chain fission will produce increased polydispersity in polymer samples.
The other problem associated with the aminolysis procedure arises from the exposure of the polymer to highly racemising conditions. The Cornell group routinely detected up to 2 mol% of D-amino acid residues in their polyhydroxyalkyl-glutamine co-polymers and in the case of poly(hydroxybutyl-glutamine co-L-aspartic acid) as many as 15% of the aspartic acid residues were converted to the D stereo isomer (Kobayishi et al., 1977).

The unperturbed dimensions of a chain molecule may be estimated experimentally in good solvents by a series of relationships involving the intrinsic viscosity, molecular weight and second virial coefficient of the polymer (Brant and Flory, 1965a). These relationships are valid only for polymers with a sufficiently high degree of polymerization ($n_p > 100$) and which are monodisperse, i.e. $\bar{M}_n = \bar{M}_v = \bar{M}_w$ (Flory, 1969). It has also been predicted (Miller et al., 1967) that the incorporation of D-amino acid residues into 'alanine-like' poly-L-peptides markedly lower the chain dimensions; even 2% of D-residues in the chain were predicted to reduce the chain dimensions by about 15%. Thus, for the purposes of the present study, a method of synthesizing polyhydroxyalkyl-glutamines that includes the hazards of chain fission with an increase in polydispersity and of racemisation cannot be considered suitable for preparing polymers if the objectives are investigations into their configurational statistics.

There are many methods now available for coupling carboxyl and amino groups to form amide bonds and they are used extensively in peptide synthesis. These methods have been designed to avoid the problems of racemisation. The two
procedures most commonly used are direct coupling using a
carbodiimide and preliminary carboxyl group activation by
formation of an active ester intermediate (see e.g. Bodanszky
et al. (1976) for a review of these methods). One of the most
successful active ester procedures is that of Anderson et al.
(1964) using esters of N-hydroxysuccinimide; of the various
active esters available for peptide synthesis these are reported
to be the least susceptible to racemisation (Anderson et al.,
1965) and, being soluble in a wide variety of solvents, can be
coupled under many different conditions. When applied to the
synthesis of polyhydroxyethylglutamine from poly-L-glutamic
acid, the first step in the reaction (Rn.3) would be the
conversion of the side chain carboxyl groups to hydroxy-
succinimide esters, a reaction catalysed by dicyclohexyl-
carbodiimide. This could be followed by the subsequent
aminolysis of the highly reactive N-hydroxysuccinimide ester
to give polyhydroxyethylglutamine (Rn.4).

\[
\begin{align*}
\text{COOH} & \quad \text{(CH}_2\text{)}_2 \\
\text{\text{-NH-CH-CO-}} & \quad \text{n}
\end{align*}
\]

\[
+ \quad \text{nHON} \quad \text{CO-CH}_2 \\
\text{CO-CH}_2 & \quad \text{DCC}\rightarrow \\
\text{\text{-NH-CH-CO-}} & \quad \text{n}
\]

\[
\begin{align*}
\text{COO-N} & \quad \text{(CO}_2\text{)}_2 \\
\text{CO-CH}_2 & \quad \text{nH}_2\text{O}
\end{align*}
\]

\[
\begin{align*}
\text{nH}_2\text{N-(CH}_2\text{)}_2\text{-OH} & \\
\text{CO-NH-(CH}_2\text{)}_2\text{-OH} & \quad \text{(CH}_2\text{)}_2 \\
\text{\text{-NH-CH-CO-}} & \quad \text{n}
\end{align*}
\]

\[
+ \quad \text{nHOSu}
\]
Since OSu esters are highly reactive and susceptible to hydrolysis by water, the reaction (Rn.3) must be performed under non-aqueous condition. Poly-L-glutamic acid is freely soluble in water at neutral pH but has only limited solubility in other solvents, so an alternative method was investigated to circumvent the necessity for non-aqueous solvents.

A direct coupling of the side chain carboxyl group with ethanolamine should be possible using the carbodiimide coupling procedure, but the most commonly used carbodiimides are very insoluble in water (Bodanszky et al., 1976). Water soluble carbodiimides are available, in particular N-ethyl-N'- (γ-dimethylamino propyl) carbodiimide (Sheehan et al., 1956), but these have found only limited use in peptide synthesis, e.g. in the linking of acidic peptides by their carboxyl ends to insoluble resins (Previero et al., 1973).

Water soluble carbodiimides of this type have been used, however, to modify the carboxyl side chains of peptides and proteins (Riehm and Scheraga, 1966; Hoare and Koshland, 1966, 1967) and would thus appear to offer good possibilities for coupling the carboxyl side chain of polyglutamic acid to amino alcohols.

The reaction sequence is initiated (Rn.5) by condensation of the carboxyl with the carbodiimide to give an O-acylurea (H.G. Khorana, 1953). The activated carboxyl group of this adduct can then react by one of three routes as shown in Rns.(6), (7) or (8).
An attack by the nucleophile ethanolamine will yield hydroxyethylglutamine plus the urea derived from the carbodiimide Rn.(6). Alternatively, the O-acylurea can rearrange to an N-acylurea via an intramolecular acyl transfer or water can act as the nucleophile in a reaction analogous to Rn.(6), in which case the carboxyl will be regenerated with the conversion of 1 molecule of carbodiimide to its corresponding urea. Kinetic studies on model carbodiimide-carboxyl-nucleophile systems have shown that the alternative reactions (Rns. 7 and 8) can be made comparatively insignificant if the concentrations of amino nucleophile and carbodiimide are high (Hoare and Koshland, 1967), thus driving the coupling to completion (Rn.6).
The work presented in this chapter explores the use of the N-hydroxysuccinimide active ester intermediate and also the direct coupling procedure using N'-(γ-dimethylaminopropyl) carbodiimide in water for the modification of poly-L-glutamic acid side chains. To facilitate the identification of products by NMR, the N-hydroxysuccinimide ester of propionic acid was prepared as a model for the ester of the side chain of poly-L-glutamic acid.

3. METHODS

A. Materials

All solvents were of analytical grade and used as purchased (from BDH Chemicals Ltd., Poole, England), except:

DMF - This was gassed with N₂ overnight to remove volatile amines, stored under N₂ and then filtered through an alumina column immediately before use. It was checked before use that it was free of dimethylamine using the method of Stewart and Young (1969).

1,4-Dioxane - This was freshly filtered through an alumina column.

Diethylether - This was redistilled then stored over sodium.

Pyridine - This was shaken with Å₄ molecular sieves (purchased from BDH Chemicals Ltd., Poole, England) for at least 24 hr, filtered through a glass sinter and stored over Å₄ sieves.

Water - Glass distilled water was used throughout this work.

Deuterated solvents for NMR - DMSO-d₆:100 atom% D

D₂O:99.8 atom% D

Both were obtained from Aldrich Chemical Company Inc., Milwaukee, Wisconsin, U.S.A.
Poly-L-glutamic acid was purchased from the Protein Research Foundation, Japan, as the sodium salt. No information was available from the supplier on its degree of polymerisation. Its molecular weight by gel filtration and by end-group titration was found to be approximately 25,000 as received. The sodium salt was desalted before use in the preparation of poly-L-HEG by dissolving in water, titrating to a pH of 2.0 with 1 M HCl, and then dialysing for two days against 4 x 4 l of distilled water, or until chloride tests with AgNO₃ were negative. The product was then freeze dried. The yield from this procedure was 80%. Due to a loss of low molecular weight material during dialysis, the molecular weight of the free poly-L-glutamic acid after desalting was 27,000 by end-group titration.

EDMAP, NHS and dithiothreitol were purchased from the Sigma Chemical Company, St. Louis, Missouri, U.S.A. DCC was purchased from the Protein Research Foundation, Japan. o-Tolidine was purchased from Ajax Chemicals Ltd., Sydney-Melbourne.

Bovine β-lactoglobulin was purchased from Koch-Light Laboratories Ltd., Buckinghamshire, England.

L-Pyroglutamyl peptide hydrolase (EC 3.4.11.8) calf liver was purchased from Boehringer Mannheim, The Boehringer Corp. (London) Ltd., London, England.

Sephadex gels were purchased from Pharmacia, Uppsala, Sweden.

The origins of some other materials are referenced in the text in the procedure in which they are used.
The following materials were generously donated:
PGA peptide (Pyroglu.Ala.Tyr.Ile) by Dr. G. Tregear, Howard Florey Institute of Experimental Physiology and Medical Research, University of Melbourne, Parkville, Victoria.
\( \alpha S_2 \) and \( \alpha S_1 \) by Mr. R. Beeby, C.S.I.R.O., Division of Dairy Research, Highton, Victoria.
Insulin A chain (prepared by the method of Fittkau (1963) by Dr. B. Davidson, Russell Grimwade School of Biochemistry, University of Melbourne, Parkville, Victoria.

B. Syntheses

(i) Thin-Layer Chromatography

Thin-layer chromatography was used to ascertain the purity of various derivatives. Three solvent systems were routinely used: "ME" = methanol:ethylacetate, 1:2; "CAM" - chloroform:glacial acetic acid:methanol, 90:2:8, and "BAWP" - n-butanol:glacial acetic acid:water:pyridine, 15:3:12:10, with silica gel aluminium backed plates (Merck F-254) of 0.2 mm thickness. Components were detected on the plates by UV quenching of fluorescence (after exposure at 254 nm), ninhydrin reagent and chlorine/o-tolidine-potassium iodide reagent.

(ii) Preparation of the N-Hydroxysuccinimide Ester of Propionic Acid

\[
\begin{align*}
\text{CH}_2\text{CH}_2\text{COON} & \quad \text{C} \quad \text{CH}_2 \\
\text{C} & \quad \text{CH}_2
\end{align*}
\]
0.37 g (5 mmoles) of propionic acid and 0.863 g (7.5 mmoles) of \( \text{N-hydroxysuccinimide} \) were dissolved in 10 ml of dioxane:ethylacetate, 1:1, and chilled in ice. 1.55 g (7.5 mmoles) of DCU in 5 ml of dioxane:ethylacetate were added. The reaction was followed by TLC in CAM and ME for 18 h. The insoluble DCU that formed was then removed by filtration. The reaction mixture was concentrated by rotary evaporation, taken up in ethylacetate and washed with saturated sodium bicarbonate four times, to remove unreacted propionic acid. It was then dried with anhydrous sodium sulphate and recrystallised overnight with petroleum ether:diethyl ether, 9:1. This procedure yielded 0.446 g of crystals (52% yield) which gave one spot on TLC in CAM (\( R_f \) 0.57) and ME (\( R_f \) 0.59). The melting point, taken with a Büchi melting point apparatus was 40-42°C (uncorrected). The 100 mHz \(^1H\)-NMR spectrum of this compound in DMSO-\( d_6 \) was recorded with a Varian HA100 NMR spectrometer in a continuous wave mode with a sweep width of 1000 cps at room temperature and was consistent with a pure product of the required structure (see Fig.V.1). Although microanalysis figures were not in exact agreement (calculated for \( \text{C}_7\text{H}_9\text{O}_4\text{N} \): C, 49.12; H, 5.26; N, 8.19. Found: C, 51.09; H, 5.87; N, 8.50) the compound was judged pure enough to serve as a model compound for proton assignments in the NMR spectra of the side chain glutamyl OSu ester.
FIG. V.1 NMR SPECTRUM OF THE N-HYDROXYSUCCINIMIDE ESTER OF N PROPYONIC ACID IN DMSO; (...) INTEGRATED SPECTRUM
(iii) Preparation of the N-Hydroxysuccinimide Ester of Poly-L-Glutamic Acid

The method used was based on that of Anderson et al. (1964). Desalted poly-L-glutamic acid (0.14 moles) was reacted for 16 hr with quantities of HOSu and DCC (0.28 moles each) solvent and temperature were varied (see Table V.2). Any insoluble precipitate formed was identified on the basis of melting point. DCU has a melting point of 228°C. Precipitates with melting points below this was indicative of insolubilised polymer and the latter was extracted from any DCU by washing exhaustively with hot ethanol and filtering on a glass fibre disc. The product was precipitated from the reaction mixture with ether, redissolved in solvent (see Table V.2) and reprecipitated with ether.

The extent to which the polymer had reacted was assessed by estimating the OSu ester bound to the polymer as described in reaction (4.A(i)b).

(iv) Preparation of Poly-Hydroxyethylglutamine using a Water Soluble Carbodiimide

The method used was a modification of that of Carroway and Koshland (1972) for the estimation of carboxyl groups in proteins. A typical procedure was as follows:

Desalted poly-L-glutamic acid (1 mmole) was dissolved in ethanolamine HCl (pH 8.0, 1.0 M, 10 ml), EDMAP was added

1. When referring to molar concentrations of polypeptides in this chapter, the molarity quoted is with respect to the residue molecular weight which, of course, differs for each of the homopolypeptides, poly-L-HEG and poly-L-glutamic acid. For native proteins, a mean residue molecular weight of 120 is used.
(1.5 mmole) and the pH adjusted to 5.0 with 1 M HCl. The reaction was allowed to proceed with stirring at room temperature and the pH was kept in the range 4.5 to 5.0 by titration with ethanolamine (1.0 M). (As the reaction proceeded the pH fell and would drop by approximately 3 pH units during the incubation. This is presumably due to HCl being released as the ethanolamine is coupled to the polymer.) After two hours, a further 1.5 mmoles of EDMAP was added, the pH adjusted to 5.0 and the reaction mixture was left stirring for up to a further 18 hr. Since the extent of conversion of the side chain groups was not appreciably increased by the second addition of EDMAP this was omitted in later preparations because of the problem of removing the urea of this compound from the preparation except by exhaustive dialysis, which resulted in loss of polymer.

At the end of the reaction period the polymer was purified in three ways:

(a) Dialysis: Dialysis tubing (Union Carbide) was either soaked for 1 hr in distilled water or brought to the boil in distilled water and then transferred to fresh water, cooled and soaked for 1 hr. The retention properties of the soaked tubing were checked. It retained 80% of the poly-L-glutamic acid Na salt (M.W. 25,000) but only 26% of lysozyme (M.W. 14,000). Dialysis was against 4 l of distilled water at 4°C, the water was changed twice daily until the dialysate was free of ninhydrin positive material (about 6 changes of water were necessary for this).
(b) Ultrafiltration: The polymer was filtered onto Amicon Diaflo ultrafilters in an Amicon Diaflo apparatus, under nitrogen at a pressure of 80 psi at 4°C. Three different sizes of ultrafilters were used: UM2 which retains molecular weights >1,000, UM10 which retains materials with a molecular weight >10,000 and UM20E which holds back polymers in the 10,000-20,000 range. The preparation was washed on the membrane with about 1 l of distilled water and then freeze dried.

(c) Gel Filtration: This was generally used in conjunction with the Diaflo purification. The freeze dried preparation from that step was taken up in about 5 ml of water and applied to a Sephadex G25 (fine) column (3.5 x 100 cm). The material was eluted with water at a rate of 28 ml/hr. The column was continually monitored at 205 and 280 nm with an LKB UVCORD III model 2089 flow monitor. Neither the contaminating EDMAP-urea nor the polymer absorbed at 280 nm, but the 210 absorbance peak of the EDMAP-urea was broader than that of the polymer and when present in large amounts it was detectable even at 280 nm. This was sufficient to roughly estimate the position of the late eluting EDMAP-urea, but since the molecular weight range of the poly-HEG was so polydisperse, 3 ml fractions were collected and absorbances measured manually at 230 nm to be certain that the pooled polymer reactions were quite free of contaminating EDMAP. The fractions containing the 230 absorbing material were discarded and the remainder were freeze dried.
Estimation of the Yield of Poly-Hydroxyethylglutamine

At each stage of the purification procedure the \(^1H\)-NMR spectrum of the product at 60 mHz was obtained in \(D_2O\) at 308° K on a Perkin-Elmer R12 NMR spectrometer operated in continuous wave mode with a sweep width of 10 ppm. Both the conversion of the poly-L-glutamic acid side chains and the degree of contamination by EDMAP-urea were assessed quantitatively from the NMR spectra of the final products. Instrument integration of the spectrum was done six times and the results averaged; this should give an accuracy of about ±5% (Kasler, 1973). A typical spectrum (of preparation E, Table V.3) is shown in Figure V.2. The assignment of the side chain methylenes of poly-HEG were made according to the published results of Joubert et al. (1970) on poly-HEG and the protons are numbered using his system. Those of the EDMAP molecule and its urea are substantiated in Section 4.C (iii) of this chapter.

The percentage conversion of side chains was calculated by taking the ratio of the areas under the resonances at 3.15-4.00 ppm of the \(CH_2(6 + 3)\) to those at 1.70-2.75 of the \(CH_2(1 + 2)\) groups of the original polyglutamic acid. The areas under these two peaks should be equal when the reaction is 100% complete and all side chains are converted to the hydroxyethyl-glutamine.

The degree of contamination by EDMAP-urea was assessed using the resonance at 2.9-3.05 arising from the dimethyl groups of that compound. The area under this peak is the product of six protons and must be scaled down by 2/3 before taking the
FIG.V.2 NMR SPECTRUM OF POLYHYDROXYETHYLGLUTAMINE PREPARATION E (TABLE V.3) IN D₂O

(HEG)ₙ \( \text{EDMAP-urea} \)

H-(NH-CH-CO)ₙ-OH

\( \alpha \)

| \( \text{CH}_2 \)₁ |
| \( \text{CH}_2 \)₂ |
| CO |
| NH |
| \( \text{CH}_2 \)₃ |
| \( \text{CH}_2 \)₆ |
| OH |

\( 1 \ 2 \ 3 \ 4 \ 5 \ \text{CH}_3 \)₆

\( \text{H} \ \text{H} \ \text{H} \)
ratio of it to the area under the CH$_2$(1 + 2) peak. The mole percentage contamination, thus derived, was converted to %w/w on the basis of the molecular weight of the EDMAP-urea (210) and of the hydroxyethylglutamyl residues (172) and was corrected for the percentage of side chains converted.

The overall yield of the product was calculated on the basis of dry weight compared to expected yield and was corrected for the assessed contamination by EDMAP-urea.

(v) Preparation of the Urea Derivative of N-ethyl,N'-
(3-dimethylaminopropyl) Carbodiimide

The urea derivative of EDMAP (1 mmole) was readily made by incubating with hydrochloric acid (1.5 mmole) for 12 hr at room temperature. Unreacted hydrochloric acid was removed by drying the product in a vacuum dessicator for several days. The resulting compound was analysed in a V-G Micromass double-focussing mass spectrometer and the molecular weight was found to be 173.1525 ± 0.0003 which is in excellent agreement with the calculated molecular weight of C$_8$H$_{19}$N$_3$O, namely 173.15270 Daltons.

(vi) Cleavage of Polypeptides Containing Glutamic Acid with a Water Soluble Carbodiimide

In most experiments using poly-L-glutamic acid this was previously fractionated to produce samples with a narrower spectrum of molecular weights than that supplied by the manufacturer. This was done on a Sephadex G75 standard grade column (3.5 x 100 cm) and eluted at a rate of 25 ml/hr with 0.1 M KCl to reduce trailing due to charge effects and monitored at 205 nm with an LKB UVICORD III flow monitor.
The conditions used for the cleavage of polypeptides are similar to those used in the previous section (3.B(iv)) for the preparation of poly-L-hydroxyethylglutamine. For poly-L-glutamic acid, a 3 molar excess of EDMAP was added to the polymer dissolved in water at a concentration of 25 μmoles/ml, the pH was adjusted to 6.0 with 0.1 M HCl and the pH monitored between 5 and 6 for the course of the incubation (20 hr).

For the proteins, insulin A-chain and β-lactoglobulin, which contain only a few residues of glutamic acid, smaller amounts of EDMAP (i.e. equimolar quantities) were used, but the other conditions were essentially the same.

The products were purified in a variety of ways, depending on the purpose of the experiment. The methods included washing on a Diaflo UM2 membrane and gel filtration on Sephadex G25 or G75 preparative columns as described previously.

C. Quantitative Estimations

(i) Estimation of Molecular Weights of Polymers

Two methods of estimating the molecular weights of poly-L-glutamic acid and polyHEG preparations were used depending on the degree of accuracy required.

(a) Ninhydrin Determination of Amino Terminal Groups

The procedure followed was a modification of that of Moore (1968). Samples containing up to 0.2 μmoles of amino groups were dissolved in 1 ml of n-propanol:water (1:1) and heated on a boiling water bath in stoppered glass tubes for 10 min to remove dissolved air. After cooling, 1 ml of
ninhydrin reagent (DMSO: 4 M lithium acetate buffer, pH 5.2 3:1 containing 2% ninhydrin and 0.06% hydridantin) was added and the stoppered tubes heated in the boiling water bath for 20 min, cooled and 8 ml of n-proponal:water (1:1) added.

Absorbances were measured at 570 nm using a spectrophotometer. Standard amounts of free glutamic acid were reacted at the same time to provide calibration curves for colour yields and, hence, estimates of the free amino groups in the polymer sample. It was found that even after intensive purification, traces of EDMAP or its urea were present as contaminants in polymer samples. Fortunately, this compound was found not to react with ninhydrin under the analytical conditions used, even when present at the maximum concentrations known (from NMR estimates) to be present in polymer samples (5 mg/ml).

(b) Gel Filtration

Molecular weights of polypeptides in the range 3,000 to 60,000 Daltons can be estimated by gel filtration on a Sephadex G75 column (Andrews, 1964). A Sephadex G75 superfine column (1.5 x 145 cm) equilibrated with Tris/HCl buffer (pH 7.5, 0.05 M Tris, 0.1 M KCl) was calibrated by observing the elution volumes of proteins (5.10 g in 2 ml of eluting buffer) of known molecular weight (see Table V.9) at a flow rate of 12 ml/hr. The column was monitored at 205 and 280 nm.

\[ \log_{10} \text{M.W.} = 5.78 - 0.0116 \text{Ve} \]
## TABLE V.1  ELUTION VOLUMES OF CRISTALLINE PROTEINS OF KNOWN MOLECULAR WEIGHT ON A SEPHADEX G75 SUPERFINE COLUMN

<table>
<thead>
<tr>
<th>Protein</th>
<th>M.W.</th>
<th>Log$_{10}$ M.W.</th>
<th>Ve(mls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucagon ( Sigma)</td>
<td>3,500</td>
<td>3.54</td>
<td>193</td>
</tr>
<tr>
<td>Insulin monomer (B.D.H., Bovine)</td>
<td>5,733</td>
<td>3.76</td>
<td>175</td>
</tr>
<tr>
<td>Insulin dimer*</td>
<td>11,466</td>
<td>4.06</td>
<td>145</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sigma, Beef Heart Type V)</td>
<td>12,400</td>
<td>4.09</td>
<td>141</td>
</tr>
<tr>
<td>Myoglobin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sigma, Whale Type II)</td>
<td>17,800</td>
<td>4.25</td>
<td>131</td>
</tr>
<tr>
<td>β-Lactoglobulin A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Koch-Light, Bovine)</td>
<td>36,500</td>
<td>4.56</td>
<td>106</td>
</tr>
<tr>
<td>Serum Albumin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sigma, Bovine Type F)</td>
<td>67,000</td>
<td>4.83</td>
<td>86</td>
</tr>
</tbody>
</table>

* Inferred from its elution volume.

By carefully standardising the buffers and flow rate of the column and sample application, a reproducibility in elution volume of ±1 ml could be obtained. Thus, molecular weights could be estimated to the nearest 1000 Daltons in the middle of the range of the column.
(ii) **Estimation of Polypeptide Concentrations**

It was not always possible to calculate the concentration of a polymer sample on the basis of dry weight because of contamination by EDMAP-urea or, more seriously, by KCl present after purification by column chromatography. Attempts to estimate the concentration by polarimetry and by U.V. absorption between 200 and 210 nm were not successful due to interference by contaminating compounds.

Peptide bond concentration can be determined by changes in the U.V. spectrum when polypeptides undergo the Biuret reaction (Ellman, 1962). Differences between the absorbance of alkaline polyglutamic acid solutions in the presence and absence of copper were found here to be greatest at 273 nm. Neither EDMAP or KCl at the maximum degree of contamination gave any colour reaction. The following procedure was thus routinely adopted:

Alkali or alkaline copper sulphate (2 M NaOH, 3 ml with or without 0.045% CuSO₄·5H₂O) were added to duplicate samples of polypeptide (0-1 mg in 0.3 ml water) and to water blanks (0.3 ml). The contents of the tubes were mixed thoroughly, those containing polypeptide were read at 273 nm against the appropriate water blanks. The difference between the absorbance for each sample in the presence and absence of copper were computed (ΔA₂₇₃) and the polymer content estimated from samples containing a known amount of pure poly-L-glutamic acid or poly-L-HEG treated at the same time. Under these conditions the ΔA₂₇₃ was linearly proportional to the amount of polymer present up to 0.4 mg.
The colour yields for poly-L-HEG was found to be higher than that of poly-L-glutamic acid, presumably due to the side chain amide bond also undergoing a Biuret reaction (see Fig.V.3).

**FIG.V.3** COLOUR YIELDS FOR THE BIURET REACTION WITH VARYING AMOUNTS OF POLYGLUTAMIC ACID AND POLYHYDROXYETHYLGLUTAMINE
Estimation of Pyroglutamyl N-Terminal Groups

The possible occurrence of pyroglutamyl residues at the N-termini of polypeptide chains was investigated by determining the number of free amino terminal groups before and after treating the polymer with L-pyroglutamyl peptide hydrolase using the method of Podell and Abraham (1978).

Materials:

Deblocking buffer (DBB) = 0.1 M phosphate buffer, pH 8.0, containing 5 mM dithiothreitol, 10 mM EDTA and 5% glycine.

Enzyme = L-pyroglutamyl peptide hydrolase.

The crude lyophilisate was dissolved in DBB at a concentration of 40 mg/ml just before use.

PGA Peptide = Pyroglu.Ala.Tyr.Ile was used at a concentration of 5 mg (0.5 μmoles)/ml in DBB.

Procedure:

Polymer (5 mg) was weighed out in duplicate in glass stoppered tubes and dissolved in DBB (100 μl). One set of duplicates were stored at 4 °C; to the other set enzyme solutions (50 μl) were added. The tubes were gassed with N₂ and left to shake gently at 4°C for 9 hr, a further 50 μl of enzyme solution was added, the tubes regassed and the reaction mixtures left to incubate with gentle shaking. After 16 hr absolute ethanol (500 μl) was added to the incubation mixtures. The other set of duplicates were used as "zero time" blanks; enzyme solution (100 μl) was added to each, followed immediately
by ethanol (500 μl). To remove the enzyme protein, all of the tubes were left for 30 min at 4°C, dried on a rotary evaporator, resuspended in water (500 μl) and boiled for 30 min. The extracts were filtered through a glass sinter and diluted to 1 ml with water. 500 μl of these solutions were used for the ninhydrin estimation of free amino groups by the method described previously.

Using this procedure with PGA peptide, 0.5 μmoles routinely provided 0.3-0.35 μmoles of amino terminal groups, indicating an efficiency of 60-70% for removing pyroglutamic acid end-groups from peptides with this preparation of enzyme.

4. RESULTS AND DISCUSSION

A. Conversion of Poly-L-Glutamic Acid Side Chains to their N-Hydroxysuccinimide Esters

(i) Estimation of the Unreacted Side Chain Carboxylic Groups of Poly-L-Glutamic Acid

(a) Stability of N-Hydroxysuccinimide Esters to Hydrolysis

It should be feasible to estimate the number of unreacted side chain carboxylic acid groups of poly-L-glutamic acid by pH titration (Maxfield et al. (1975), for example, estimated the composition of random copolymers of hydroxybutyl-glutamine and glutamic acid in this way). However, OSu esters are unstable in aqueous solutions and readily hydrolyse to the free acid and N-hydroxysuccinimide (Löw and Kisfaludy, 1965). To determine whether this factor would appreciably interfere with a titrametric estimation of unesterified carboxylic acid groups, the rate of hydrolysis of the N-hydroxysuccinimide ester of propionic acid (PrOSu) was measured at various pH's.
The hydrolysis of the N-hydroxysuccinimide ester to give free N-hydroxysuccinimide was followed by NMR by following the shift in the resonance of the former at 2.79 ppm to the latter of 2.55 ppm (see Fig.V.1) and thus calculating the ratio of covalently bound to free N-hydroxysuccinimide. This was done at four pD values, 1.5, 4.5, 7.3 and 10.0, over a period of 6 hr. It was found at low pD values (1.5 and 4.5) no hydrolysis took place over this period. At pH 10.0, however, all of the ester had hydrolysed within 30 min. The ratio of bound to free N-hydroxysuccinimide is plotted against time at pD 7.3 in Figure V.4. At this pD, 30% of the PrOSu ester had hydrolysed within 1 hr.

**FIG.V.4** RATE OF HYDROLYSIS OF PROPIONYL OSu ESTER ESTIMATED USING $^1$H-NMR
The rate of hydrolysis of PrOSu in the pH range 4.5 to 7.5 was measured using a Radiometer pH-stat, using known quantities of the ester (20-30 mg) and maintaining a constant pH with 0.1 M NaOH at 20°C. The percentages of ester hydrolysed in the first hour (calculated from mmoles of NaOH consumed) are plotted in Figure V.5 against the pH of the hydrolysis. (Initial rates (at t = 0) followed the same pH dependence.)
Since the rate of hydrolysis rises rapidly above pH 4.0 and the $pK_o$ of the glutamic acid side chain carboxylic acid is 4.3-4.6, depending on ionic strength (Olander and Holzer, 1968) a titrametric method of estimating unreacted side chain carboxyl groups is not ideal and a method based on the quantitative estimation of characteristic resonances in the $^1$H-NMR spectrum of PrOSu was developed.

(b) Quantitative Assay of N-Hydroxysuccinimide Esters by $^1$H-NMR

Since the cyclic methylene groups of N-hydroxysuccinimide have a single characteristic magnetic resonance at 2.55 ppm which shifts to ~2.8 ppm when the hydroxyl group is esterified (see Fig.V.1, the $^1$H-NMR spectrum of PrOSu ester on which the position of the methylene resonances are marked) it is possible to distinguish N-hydroxysuccinimide which is bound to poly-L-glutamic acid and that present as free N-hydroxysuccinimide. The model compound, propionoxysuccinimide, was used to quantitate this resonance.

The calibration curves were constructed using ester solutions containing measured weights rather than concentrations of ester as NMR measures the quantity (i.e. number) of protons rather than their concentration (Kasler, 1973). Four samples of PrOSu ester were weighed in tared NMR tubes which were then filled to a height of approximately 3 cm with CDCl$_3$. The spectra were integrated at a sensitivity of 1, the peak areas of the N-hydroxysuccinimide ester resonances at 2.79 ppm measured for each sample and plotted against quantity in mmoles (see Fig.V.6).
(ii) Effect of Reaction Conditions on Yields of N-Hydroxysuccinimide Esters of Poly-L-Glutamic Acid

The range of conditions and the degree of success in esterifying the polymer using dicyclohexylcarbodiimide as discussed in 3.B(iii) are shown in Table V.2.

The solvents for the reaction were those in which poly-L-glutamic acid showed the best solubility (in the order pyridine > DMF > acetonitrile). In acetonitrile the polymer was only sparingly soluble and frequently precipitated during the reaction period. Since esterification did not take place on the polymer in the solid-phase, other solvents in which poly-L-glutamic acid was less soluble were not tried (DMSO, methanol, dioxane, chloroform).
TABLE V.2  CONDITIONS AND YIELDS FOR POLY-L-GLUTAMYLM OSu SYNTHESIS

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Temperature</th>
<th>% Esterification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetonitrile</td>
<td>4°C</td>
<td>0</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>R.T.</td>
<td>0</td>
</tr>
<tr>
<td>DMF</td>
<td>R.T.</td>
<td>0</td>
</tr>
<tr>
<td>DMF</td>
<td>60°C</td>
<td>0</td>
</tr>
<tr>
<td>Pyridine*</td>
<td>R.T.</td>
<td>50</td>
</tr>
<tr>
<td>Pyridine</td>
<td>60°C</td>
<td>20</td>
</tr>
</tbody>
</table>

* This preparation was retreated under the same conditions with no increase in the number of side chains converted.

The only solvent in which any reaction took place was pyridine and even here all attempts to increase the yield of esterified side chains above 50% were fruitless. The reason for this could be three-fold. It is known from classical solution peptide synthesis and the fragment condensation method (Protein Synthesis Group, 1975; Meienhofer, 1973) that reaction with α-carboxyl groups decreases in rate as the peptide chain gets longer and the yields of peptides of greater than 20 residues by solution methods are very low. Thus, the γ-carboxylic acid groups of a long chain polymer may also be markedly less reactive than those in small peptides.

There is also evidence that since the γ-glutamyl carboxylic acid is a weaker acid than the α-carboxylic acid it is less reactive than the latter. Thus, in the solid-phase synthesis of poly-γ-L-glutamates of folic acid (Krumdieck and
Baugh, 1968) it was found that the resin-ester linkage through the γ-carboxyl group was much weaker than the usual carboxyl-resin ester bond, and that DCC γ peptide bond couplings, under conditions that would have been more than adequate for the formation of an α peptide bond, resulted in incomplete couplings. Similar results have been obtained in attempted solid-phase synthesis of a number of oligo-γ-L-glutamyl peptides (Nitecki and Goodman, 1970).

Meienhoffer et al. (1970) attempted to synthesise hepta-γ-L-glutamic acid by several routes. Using the standard solid-phase procedure of Marshall and Merrifield (1965) they also found incomplete couplings and turned to classical solution methods. The standard DCC coupling procedure was abandoned because the N-benzyl rearrangement (Rn.10) of the activated side chain carboxyl group which, with α-carboxylic acids is a suppressible side reaction, appeared to be the major reaction route.

\[
\begin{align*}
Z-NH-CH-Co-O^tBu \quad & \quad Z-NH-CH-CC-O^tBu \quad O-acyl \text{ urea} \\
(\text{CH}_2)_2 \quad & \quad (\text{CH}_2)_2 \\
COOH \quad & \quad (\text{active}) \quad (\text{inactive}) \\
\end{align*}
\]
Of the active ester methods of coupling Meienhofer et al. found that the γ-pentachlorophenyl esters were not very reactive and the OSu esters were unstable. Since γ-glutamyl peptide bonds are more labile to various hydrolytic agents than α-peptide bonds (Nitecki and Goodman, 1970) ester linkages may also be more unstable than the corresponding α-linkage. They finally succeeded in synthesizing the γ-L-glutamyl peptides using the mixed carbonic anhydride method which is noted for its speed of reaction and ability to produce products with a high yield and purity (Anderson et al., 1967).

Thus, the lack of success in completely converting all of the poly-L-glutamic acid side chains to the active N-hydroxysuccinimide ester could well be due to the comparatively low reactivity of the γ-carboxyl side chain, exacerbated by its incorporation into a long chain polymer, the instability of the OSu esters formed and the undesirable side reaction to give the N-acyl urea. These problems are not encountered in the synthesis of other esters of poly-L-glutamic acid (e.g. poly-γ-benzyl glutamate) since the most commonly used method is to polymerise the highly reactive N-carboxyanhydrides of the previously synthesized γ-glutamyl ester monomer (see, for example, Nagayama and Wada, 1973; Bradbury et al., 1973). The mixed carbonic anhydride method, which was successful in the studies by Meienhofer et al. (1970), should be explored in future studies in producing poly-γ-L-glutamyl derivatives of the type discussed in these pages.
B. Preparation of Poly-Hydroxyethylglutamine using a Water Soluble Carbodiimide

Several batches of poly-L-HEG were prepared with variations in some of the conditions in the general method described in 3.B(iv). Table V.3 summarizes the degree of success of these preparations and the conditions of preparation and purification of the polymer.

Table V.3 shows a wide variation in yield of recovered polymer, depending upon the exclusion limits of the membranes used for removal of low M.W. reagents. Thus, a 65% yield was obtained with a UM2 Diaflo membrane that is claimed to retain molecules >1000 M.W., a 35% yield with a UM10 membrane that retains molecular weights >10,000 and a 12-16% yield with either a UM20 membrane or with dialysis tubing, both of which have exclusion limits of about 20,000 M.W. Since the M.W. of the starting material (dialysed, desalted poly-L-glutamic acid) is about 27,000, one would expect the resulting poly-L-HEG to have a molecular weight of 36,000 - which, like the starting polymer, should easily be retained by dialysis tubing.

Initially, it was considered possible that the highly reactive excess EDMAP could be coupling the polymer covalently to the cellophane dialysis tubing which is known to contain free -SH groups. Boiling the tubing to remove sulphhydryl compounds and adding acetic acid at the end of the coupling to convert the EDMAP to its inert urea, however, made little improvement to the yield (c.f. preparation C and D, Table V.3). The molecular weights of these preparations were in the range expected from the molecular weight of the starting polymer.
<table>
<thead>
<tr>
<th>Time of Coupling (h)</th>
<th>Yield %</th>
<th>Side chains % Converted</th>
<th>Mol. Wt. *Ninhydrin + Gel Filtration</th>
<th>Methods of Purification</th>
<th>Contamination by EDMAP-urea %</th>
<th>Modifications to the General Method of 3.B.(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.5</td>
<td>32</td>
<td>65</td>
<td>Diaflo on a UM10 membrane</td>
<td>7</td>
<td>1.5 moles EDMAP per mole COOH groups added at T₀</td>
</tr>
<tr>
<td>B</td>
<td>4.5</td>
<td>62</td>
<td>80</td>
<td>Diaflo on a UM2 membrane followed by gel filtration (Sephadex G25)</td>
<td>11</td>
<td>1 mole of glacial acetic acid per mole of EDMAP present added at T∞</td>
</tr>
<tr>
<td>C</td>
<td>6.0</td>
<td>16</td>
<td>100</td>
<td>Dialysis through boiled cellophane tubing</td>
<td>&lt;2</td>
<td>As in preparation B</td>
</tr>
<tr>
<td>D</td>
<td>20</td>
<td>12</td>
<td>100</td>
<td>Dialysis through soaked cellophane tubing</td>
<td>&lt;2</td>
<td>No modifications</td>
</tr>
<tr>
<td>E</td>
<td>20</td>
<td>55</td>
<td>100</td>
<td>As in preparation B</td>
<td>15</td>
<td>As in preparation A</td>
</tr>
<tr>
<td>F</td>
<td>12</td>
<td>+21,000</td>
<td></td>
<td>Diaflo on a UM20 membrane followed by gel filtration (Sephadex G25)</td>
<td>7</td>
<td>1.5 moles of EDMAP and 20 moles of ethanolamine per mole COOH added at T₀ The preparation was separated into two fractions by retention on two Diaflo membranes of widely different size and investigated independently</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>35</td>
<td>100</td>
<td></td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>
However, it seemed irrefutable that some 85% of the material lost during dialysis must have a molecular weight far lower than that of the starting polymer. Preparation E purified by retention on a UM2 membrane confirmed this conclusion. The yield was 65%, but the \(\bar{M}_W\) determined by two methods was 12-13,000 - about 1/3 of that expected. Preparation F, retained on a UM20 membrane also had a decreased \(\bar{M}_W\) (about 21,000 by gel filtration).

The elution profiles of these two preparations on gel filtration chromatography using Sephadex G75 Superfine indicated that the degradation of the polymer chains may not be completely random. The starting material shows a broad symmetrical band indicating a very polydisperse but symmetrical distribution of chain lengths in this polymer (see Fig.V.7).

**FIG.V.7** ELUTION PROFILES ON SEPHADEX G75 SUPERFINE OF POLY-L-HEG PREPARATIONS DESCRIBED IN TABLE V.3 \([-\cdots F, -E]\) AND OF STARTING MATERIAL POLY-L-GLUTAMIC ACID \([\cdots]\)
[The poly-L-glutamic acid shown in Figure V.7 was the undialysed, undesalted sodium salt with a $\text{MW}$ of 21,000.]

Both poly-L-HEG preparations show a bimodal $\text{MW}$ distribution. This bimodality manifested itself only as a slight shoulder on the profile of preparation F; because this low molecular weight material was probably not retained by the UM20 membrane used to collect this preparation. Thus, the value for $\text{MW}$ reported in Table V.3 is the median value of the main peak of preparation F. In the case of preparation E, the value 13,000 refers to the median of the whole elution profile, the mode of the first peak occurring at a $\text{MW}$ of 21,000 (as for prep. F) and of the second peak at about 3,000 (this is just beyond the resolution of G75).

It is common practice in peptide synthesis for couplings to proceed for 12 hr or more. For the preparation of this compound it was thought advisable to reduce the reaction time to the minimum required to complete the conversion of the side chain, in the hope that the chain cleavage reaction was much slower than the coupling reaction and could, thereby, be minimised. Preparations A, B and C show that 6 hr incubation time is necessary to completely modify the side chains, but in that time extensive chain fission had occurred. Preparations C and D were incubated for 6 hr and 20 hr, respectively, with very similar low yields of polymer with $\text{MW} > 25,000$. Even after 1.5 hr, when 65% of the side chains had been modified, only 34% of the polymer was retained by a Diaflo UM10 membrane - suggesting that 65% had been cleaved to chains of $<10,000$ in molecular weight.
The first step in the coupling pathway is the formation of the O-acyl urea of the EDMAP with the side chain carboxyl group of the poly-L-glutamic acid (Khorana, 1953) [see Eq. (5) in Section 2]. Whatever the nature of the subsequent reaction leading to peptide fission, it seemed likely that it could be minimised by ensuring preferential nucleophilic attack by the ethanolamine. Thus, in an endeavour to repress the chain cleavage a two-fold excess of ethanolamine was used in preparation F. This appears to have been ineffective, the $\bar{M}_W$ of even the highest molecular weight fraction being low enough to indicate substantial peptide bond fission.

In summary, the method for preparing poly-L-HEG using the water soluble carbodiimide coupling procedure may be satisfactory if high molecular weight preparations are not required. It has the advantage of not only using far less severe conditions, but of being much quicker than the conventional method of synthesis (Scheule et al., 1976) - only 3 days for the complete preparation, rather than 8 to 12 days. However, the problems of freeing the polymer of the side product EDMAP-urea and still obtaining reasonable yields is serious, as well as the losses on removal of the low M.W. fraction produced by peptide bond fission during the coupling.

The ureas of diimides are acknowledged to be difficult to remove from synthesized peptides (Bodanszky et al., 1976). There is the added problem with EDMAP-urea that it is soluble in water, the only effective solvent for the polymeric reactant and product, and that it contains an amino nitrogen which is readily protonated and is thus capable of binding to the polymer electrostatically (the characterisation of EDMAP and its urea derivative will be presented in section C (iii) of this chapter).
Exhaustive dialysis completely eliminated the urea from preparations C and D but provided yields of only 10-20% of polymer. With preparation A the amount of EDMAP-urea was decreased to less than 10% by washing on a UM10 membrane with a polymer yield of 32%. The use of the highly retentive UM2 ultrafilter in the purification of the preparations B and E gave higher yields of polymer (55-65%) but contamination by EDMAP-urea was more extensive (10-15%).

These and other data on the purification procedures for preparation F (see Table V.3) might indicate that the EDMAP is covalently bound to the polymer, but the stoichiometry is not exact and investigations into the conversion of EDMAP to its urea, reported in section C (iii) of this chapter showed that the urea is certainly free.

The next section will describe some of the features of the reaction responsible for peptide-bond cleavage of poly-L-glutamic acid by EDMAP and discuss possible mechanisms that could be responsible for it.

C. Investigations into the Mode of Peptide Chain Fission by a Water Soluble Carbodiimide

(i) Changes in the Molecular Weight Distribution of the Poly-L-Glutamic Acid on Treatment with N-ethyl,N'-(3-dimethylaminopropyl) Carbodiimide

To establish whether the reduction in the degree of polymerisation of poly-L-glutamic acid during its conversion to polyhydroxyethylglutamine was due entirely to the action of EDMAP, the reaction was carried out in the absence of ethanolamine. The product was purified as in the poly-L-HEG
preparations. The molecular weight of this preparation by ninhydrin end-group estimation was 3,500, but estimates of the molecular weight by gel filtration showed a bimodal distribution of molecular weights with some material even higher in molecular weight than the starting material (see Fig.V.8). The presence of this high molecular weight material suggests that in the absence of a nucleophile to react with the activated side chain some fragment condensation of the polymer chains occurs - probably between amino termini and side chain carboxylic acid groups.

![Diagram of elution profiles](image)

**FIG.V.8** ELUTION PROFILES OF POLY-L-GLUTAMIC ACID BEFORE (---) AND AFTER (——) TREATMENT WITH N-ETHYL, N'-(3-DIMETHYLAMINOPROPYL) CARBODIIMIDE
Most of the material, however, had a Ve that corresponded to a molecular weight of about 4,000 in agreement with the ninhydrin end-group estimations. The degree of chain fission in the absence of ethanolamine was greater than in its presence; in the latter case the average chain length was halved. This indicates that in the absence of a nucleophile activated side chains react more extensively with peptide bonds to cause their fission. One possible mechanism (Rn.11 and 12, below) would produce new pyroglutamyl N-terminal groups.

\[ \text{R}_1 \text{NH-C-NHR}_2 + \left( \text{CH}_2 \right)_2 \text{COOH} \leftrightarrow \text{CH-COOH} + \text{HN-CH-CO-} \]  
\[ \text{R}_1 \text{NH-C-NHR}_2 + \left( \text{CH}_2 \right)_2 \text{COOH} \leftrightarrow \text{CH-COOH} + \text{HN-CH-CO-} \]

However, treating both this preparation of poly-L-glutamic acid and preparation E (Table V.3) of poly-L-hydroxyethyl glutamine with the enzyme pyroglutamyl peptide hydrolase did not result in any increase in free amino terminal groups (the activity of the enzyme and the sensitivity of the method were checked as described in section 3.C (iii)). The occurrence of fragment condensation of the poly-L-glutamic acid was established more convincingly by fractionating the starting material and using a sample with a much narrower spectrum of molecular weights, and a higher degree of polymerisation. The elution profile of
this material on the analytical Sephadex G75 column before and after treatment with EDMAP is shown in Figure V.9. The molecular weight of the starting material was beyond the resolution of the column, being >60,000. Chain fission resulted in a highly polydisperse fraction of product with an average molecular weight of about 20,000.

FIG. V.9  ELUTION PROFILE OF FRACTIONATED POLY-L-GLUTAMIC ACID BEFORE (---) AND AFTER (——) TREATMENT WITH EDMAP

The purpose of ensuing experiments was to investigate the specificity of the carbodiimide for peptide bonds adjacent to glutamic acid residues; in particular, to see whether the fission was specific for bonds adjacent to single, double or perhaps multiple glutamic acid residues.
(ii) The Effect of N-Ethyl,N'-(3-dimethylaminopropyl) Carbodiimide on Polypeptides and Proteins Containing Single (Glu) and Double (Glu-Glu) Sequences

Insulin A chain has the sequence:

\[
\begin{align*}
1 & \quad \text{A}_4 \quad \text{B} \\
\text{Gly-Ile-Val}^{\text{A}_4} \text{Glu}^{\text{B}} \text{Cys-Cys-Thr-Ser-Ile-Cys-Ser-Leu-Tyr-Glu-} \\
\text{A}_{17} & \quad \text{B} \\
\text{Leu}^{\text{A}} \text{Glu}^{\text{B}} \text{Asn-Tyr-Lys-Asn}
\end{align*}
\]

If EDMAP cleaves peptide chains at points adjacent to single glutamic acid residues then its effect on insulin A chain would be to produce three peptide fragments; a 13 residue fragment and two smaller peptides (3 and 5 residues if fission is specific for only the NH side of the Glu residue, but two tetrapeptides if the CO side is split). In either case, the short peptide fragments should be easily separated from the longer fragment and from the unsplit insulin A chain (molecular weight 2,400) by gel filtration on Sephadex G25. However, on treating insulin A chain with the carbodiimide no fragments were detectable by thin-layer chromatography either over a reaction period of 4 hr or on separating the reaction mixture on Sephadex G25 after incubating overnight. This is not surprising since, as was discussed in the Introduction to this chapter, water soluble carbodiimides have been used extensively to modify the carboxylic side chains of proteins at isolated glutamyl residues (Carroway and Koshland, 1972) and chain cleavage during this process has never been reported.

There are many examples of 'runs' of glutamic acid residues occurring in proteins, notably the bovine milk proteins. One of these, αS₂ casein, has two sets of Glu-Glu-Glu
sequences (Brignan et al., 1977). Another, αS₁ casein, has two
sets of Glu-Glu sequences (Mercier et al., 1972). Preliminary
experiments with these two proteins were unsuccessful due to
solubility problems under the conditions used in the cleavage
procedure. Solubility problems did not arise with bovine
β-lactoglobulin. This protein has 162 residues with two Glu-Glu
sequences at positions (44,45) and (157,158) in the chain
(Braunitzer et al., 1972) and S-S bridges between residues
(106-119) and (66-160). Thus, if two adjacent glutamic acid
residues were sufficient for the cleavage reaction, treatment
of β-lactoglobulin with EDMAP would produce a mixture of peptide
fragments as listed in Table V.4. The products of chain fission
would be two fragments of molecular weight 5,000 and 13,500 and
the unsplit material of molecular weight 18,400. These fractions
should elute distinctly at the appropriate volumes for their
molecular weights when applied to the calibrated analytical
Sephadex G75 column.

After treating β-lactoglobulin with EDMAP under
conditions which would result in extensive chain cleavage for
poly-L-glutamic acid there was no evidence of any fragmentation
from the elution profile of the treated protein on Sephadex G75.
In fact, all of the material eluted with the void volume of the
column indicating an increase in effective molecular weight due
either to non-covalent aggregation or intermolecular covalent
bonding. A probable cause for the latter is that covalent cross-
linking has occurred between carboxyl and amino acid side chains
and end-groups, coupled by the carbodiimide. β-Lactoglobulin
contains 15 lysine residues as well as 11 aspartic acid and
16 glutamic acid residues.
TABLE V.4 PEPTIDE FRAGMENTS OF β-LACTOGLOBULIN CLEAVED AT GLU-GLU SEQUENCES

<table>
<thead>
<tr>
<th>Number of Residues</th>
<th>Approximate M.W. a</th>
<th>Degree of Cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td>18,400</td>
<td>Unsplit</td>
</tr>
<tr>
<td>157b</td>
<td>17,900</td>
<td>Split only at Glu-Glu</td>
</tr>
<tr>
<td>118</td>
<td>13,500</td>
<td>Split only at Glu-Glu</td>
</tr>
<tr>
<td>113b</td>
<td>12,900</td>
<td>Split at both Glu-Glu sequences</td>
</tr>
<tr>
<td>44</td>
<td>5,000</td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>570</td>
<td></td>
</tr>
</tbody>
</table>

a. Based on a mean residue weight of 114.

b. Due to an S-S bridge between residues 66 and 160 these fragments will not be released.

Protein amino groups are reported to be unreactive with the carbodiimide under the acid pH conditions used in the present experiments (Carraway and Koshland, 1972) and although water soluble carbodiimides have been used to cross-link peptides (Goodfriend et al., 1964) and to couple proteins to cell surfaces (Johnson et al., 1966; Linscott et al., 1969) or insoluble supports (Wiliky and Weetall, 1965; Previero et al., 1973) these reactions are usually performed near pH 8.0. Nevertheless, the $pK_a$ values and nucleophilicity of lysine amino groups in proteins are known to vary and it is more than likely that some lysine and $\alpha$ amino groups with sufficiently low $pK_a$ values react with carboxyl groups in the same or other peptide chains (even at pH 5 to 6) after the latter have been activated by EDNAP. To establish or eliminate this possibility
one could reversibly block all free amino groups, for example, by trifluoroacetylation, before treating the protein with EDMA. This has not been done in the present study.

(iii) Changes in the $^1$H-NMR Spectrum of Poly-L-Glutamic Acid on Reaction with N-Ethyl,N'-(3-dimethylaminopropyl) Carbodiimide

Before following the reaction of poly-L-glutamic acid with EDMA using $^1$H-NMR spectroscopy it was necessary to assign the proton resonances in the spectra of EDMA and its urea derivative - a likely product of the reaction. $^1$H-NMR spectra of these compounds in D$_2$O were recorded at 301° K on a Bruker WH90 NMR spectrometer operated at 90 MHz in the pulse-FT mode with sweep width of 1200 Hz and 4 K real data points (i.e. a digital resolution of 0.3 Hz) and assignments were checked by spin decoupling. It was observed (Table V.5) that the resonances of protons in the alkyl substituents of the tertiary nitrogen atom (-C$_5$H$_2$,C$_6$H$_2$,C$_7$H$_2$) are shifted considerably upfield when EDMA is converted to its urea derivative; these shifts are diminished for protons attached to carbons more remote to the nitrogen atom (C$_5$H$_2$ > C$_4$H$_2$ > C$_3$H$_2$). This might suggest that the changes in the $^1$H-NMR shifts on converting EDMA to its urea are due to a change in the state of ionization of the trialkyl substituted nitrogen atom. In an effort to simulate this shift the $^1$H-NMR spectrum of EDMA was measured at higher pD values (>11) at which deprotonation of the EDMA might occur, resulting in a comparable upfield shift analogous to that observed on urea formation. Experimentally, such shifts could not be observed without simultaneous conversion to the
<table>
<thead>
<tr>
<th>Functional Group(^b)</th>
<th>EDMAP(^c)</th>
<th>EDMAP UREA(^c)</th>
<th>Change(^d) (\Delta\delta) (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mathrm{ClH}_3)(^-)</td>
<td>1.138 (t)</td>
<td>1.066 (t)</td>
<td>0.072</td>
</tr>
<tr>
<td>(-\mathrm{C2H}_2)(^-)</td>
<td>3.141 (q)</td>
<td>3.104 (q)</td>
<td>0.037</td>
</tr>
<tr>
<td>(-\mathrm{C3H}_2)(^-)</td>
<td>3.452 (t)</td>
<td>3.164 (t)</td>
<td>0.289</td>
</tr>
<tr>
<td>(-\mathrm{C4H}_2)(^-)</td>
<td>2.17 (m)</td>
<td>1.790 (m)</td>
<td>0.380</td>
</tr>
<tr>
<td>(-\mathrm{C5H}_2)(^-)</td>
<td>3.834 (t)</td>
<td>2.863 (t)</td>
<td>0.971</td>
</tr>
<tr>
<td>(\mathrm{C6H}_3)(^&gt;)</td>
<td>3.394 (s)</td>
<td>2.618 (s)</td>
<td>0.776</td>
</tr>
<tr>
<td>(\mathrm{C7H}_3)(^\uparrow)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. In ppm from an internal standard of Tiers Salt.

b. Numbered thus \(\mathrm{ClH}_3\)-C\(2\mathrm{H}_2\)-N = C = \(\mathrm{N-C3H}_2\)-C\(4\mathrm{H}_2\)-C\(5\mathrm{H}_2\)\(\mathrm{N}\)\(\mathrm{C6H}_3\)\(\mathrm{C7H}_3\)

and \(\mathrm{ClH}_3\)-C\(2\mathrm{H}_2\)-N-C-N-C\(3\mathrm{H}_2\)-C\(4\mathrm{H}_2\)-C\(5\mathrm{H}_2\)\(\mathrm{N}\)\(\mathrm{C6H}_3\)\(\mathrm{C7H}_2\)

c. Abbreviations: \(t\) = triplet, \(q\) = quartet, \(m\) = multiplet, \(s\) = singlet

d. Change in shifts between EDMAP and its urea for each functional group.

urea derivative. However, proton shifts in the expected direction were observed when the urea was deprotonated with 10% NaOD. Thus, the \(\mathrm{C6H}_2\)\(\mathrm{C7H}_2\) chemical shift moved upfield from 2.618 to 2.280 ppm. This suggests that N-deprotonation is the major cause of the differences \(\Delta\delta\) between the chemical shifts of the functional groups of EDMAP and its urea.
Having assigned the proton resonances due to ELMAP and its urea, poly-L-glutamic acid was mixed with an equimolar amount of ELMAP in D₂O at a concentration of 10 mM in an NMR tube. Spectra were recorded at 308° K on a Perkin-Elmer R12 spectrometer in continuous wave mode with a sweep width of 10 ppm at zero time, and then after 2 hr, 4 hr and 20 hr. The only changes observable were the disappearance of the ELMAP and the concomitant appearance of its urea. The resonances of these compounds were sharp and at the positions characteristic for the free compounds, suggesting that they were in no way bound to the polymer. The portion of the spectrum arising from the poly-L-glutamic acid showed no changes. The conversion of the ELMAP to its urea was monitored by integrating the areas under the single resonances due to the \( \text{C}_6\text{H}_3 \) and \( \text{C}_7\text{H}_3 \) groups at 3.39 ppm for the ELMAP and at 2.62 for its urea. The ratios of these two peaks during the course of the reaction are shown in Table V.6. The reaction is seen to start immediately an addition of the poly-L-glutamic acid and more than half the ELMAP present has reacted to give the urea derivative after 4 hr. The spectrum of this reacting system after 2 hr is shown in Figure V.10.

(iv) Possible Reaction Mechanisms for the Fission of Poly-L-Glutamic Acid by N-Ethyl,N'-(3-dimethylaminopropyl) Carbodiimide

The body of experimental evidence accumulated from the preparation of polyhydroxyethylglutamine (section 4.B) and the experiments reported in this section of this chapter (4.C) can be summarised as follows:
TABLE V.6  CONVERSION OF EDMAP TO ITS UREA AS MEASURED BY $^1$H-NMR

<table>
<thead>
<tr>
<th>Time of Reaction (hr)</th>
<th>Ratio of Areas of Resonances</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (before addition of poly-L-glutamic)</td>
<td>0.62:3.39</td>
</tr>
<tr>
<td>0 (before addition of poly-L-glutamic)</td>
<td>1:4</td>
</tr>
<tr>
<td>2</td>
<td>1:1.5</td>
</tr>
<tr>
<td>4</td>
<td>1:1</td>
</tr>
<tr>
<td>20</td>
<td>5:1</td>
</tr>
</tbody>
</table>

1. The molecular weight of poly-L-glutamic acid is decreased by reaction with N-ethyl,N'-(3-dimethylaminopropyl) carbodiimide. The molecular weight distribution shows a spread at both ends of the distribution and the average molecular weight decreases from about 25,000 for the untreated polymer to 4,000 after reaction with EDMAP (see section 4.C(i)).

2. In the presence of ethanolamine the average molecular weight of the treated polymer decreases to 12,500, indicating that the presence of this nucleophile reduces the degree of fission (see section 4.B).

3. Degradation of the poly-L-glutamic acid is not accompanied by the appearance of pyroglutamic acid N-terminal groups, whether ethanolamine is present or absent (see section 4.C(i)).

4. Experiments with insulin A chain and $\beta$-lactoglobulin indicate that more than one consecutive glutamic acid residue is necessary for cleavage of peptide bonds by EDMAP and it is possible that more than two consecutive residues are required (see section 4.C(ii)).
FIG. V.10 $^1$H-NMR SPECTRUM OF A REACTION MIXTURE CONTAINING POLY-L-GLUTAMIC ACID AND EDMAP AFTER 2 hr [E REFERS TO RESONANCES DUE TO EDMAP, U TO THOSE DUE TO ITS UREA AND PG TO THOSE ASSIGNED TO POLY-L-GLUTAMIC ACID; (... ) INTEGRATED SPECTRUM]
FIG. V.10

C5E  C2,3,E&U  C5U  C6,7 E  C β,γ PG plus C4 E&U  C1E  C1U  TMS external standard

p.p.m. from TMS

0 1 2 3 4 5
5. Both the coupling of the glutamic side chain to ethanolamine and the cleavage of the peptide backbone are extensive after 1½ hr and complete after 6 hr (see Table V.3).

6. The urea of EDMAP is formed during the fission of poly-L-glutamic acid by EDMAP. Any adduct of this urea with poly-L-glutamic does not have a sufficiently long lifetime to be detectable in the NMR time scale. Similarly, the $^1$H-NMR spectra show no evidence of an O-acyl-poly-L-glutamic acid urea (or of any other intermediate) and thus, if present, their concentrations must be low at any one time (see section 4.C(iii)).

Any proposed reaction scheme for the cleavage mechanism must be consistent with these experimental observations.

Aspartic and glutamic acid have long been recognised as potential sites for the acid catalysed chemical cleavage of peptide bonds in proteins (Leach, 1955); the peptide bonds adjacent to these residues being particularly susceptible to hydrolysis. Most of the mechanisms involved, however, require far more rigorous conditions of pH and temperature and far more reactive reagents than were used in these experiments (selective cleavage and modification of peptides and proteins has been reviewed by Spande et al. (1970)). The mechanisms proposed for cleavage of peptide bonds at glutamic and aspartic acid residues are:

a. The reduction of esters (formed by methanol and hydrochloric acid) by lithium tetraborohydride in tetrahydrofuran with subsequent cyclization under acid conditions (Burstein et al., 1967).
b. The formation of glutarimides with thionyl chloride and their subsequent alkaline hydrolysis to give a γ-glutamyl peptide bond (Battersby and Reynolds, 1961). This can then be cleaved by hydantoin (Stark, 1968).

Since these two schemes, (a) and (b), require very stringent conditions, they are unlikely mechanisms for the cleavage process with EDMAP.

c. l-Acylpyrollidin-2-ones

This mechanism was discussed previous in section 4.C(i) (see Rns. (11) and (12)), but was rejected on the grounds that the N-terminal residue of the peptide chain produced at the point of cleavage is not a pyroglutamic acid residue.

d. Lössen Rearrangement

This series of reactions proceeds under mild conditions similar to those under which the cleavage of peptides by EDMAP is observed. Rns. (17) and (18) are known to be mediated by a water soluble carbodiimide under the mild conditions that pertain to this work (Hoare et al., 1968).

However, the glutamic acid must form a hydroxamate (Rn.16) before the carbodiimide can effect the Lössen rearrangement and this does not seem possible in our system. The net result of cleavage (under alkaline conditions - whereas our experiments are conducted under acid conditions) is a peptide chain blocked by an N-terminal imidazolidone. The enzyme pyroglutamyl peptide hydrolase would not be expected to release free amino end-groups from N-terminal imidazolidones. In any case, the close agreement between the molecular weights of cleaved chains derived from gel filtration and amino end group estimations by ninhydrin suggests that the cleaved chains have free NH₂-groups.
The reactions which the active O-acyl urea of glutamic acid and EMDAP would be expected to undergo spontaneously are given in the Introduction to this chapter (Rns. (7) and (8)). It has been suggested (Previero et al., 1973) that the O-acyl urea of a terminal carboxyl group may form an active oxazolinane (Rn.(21)) before reacting with a nucleophile (Rn.(22)), thus:

[Rns.(16), Seifter et al., 1960; Rns. (17) and (18), Hoare et al., 1968; Rns. (19) and (20), Gross and Morrell, 1966]

[N.B. A dotted line across a covalent bond denotes a point of cleavage throughout this discussion.]
If such a reaction occurred between the adduct of the side chain of glutamic acid with EDMAP, and the adjacent peptide backbone, it could render the peptide bond involved susceptible to mild hydrolysis (Rns. (23) and (24)). This may seem unlikely, but it should not be overlooked that poly-L-glutamic acid is a polyfunctional carboxylic acid and that carboxylic acids are well known as catalysts in a variety of aminolysis, cyclisation and rearrangement reactions. Thus, the cyclisation of glutamic acid esters to pyroglutamic acid is catalysed by carboxylic acids (Hubert et al., 1963). The spontaneous decomposition of D-Val-Pro-Sar to yield D-Val-Pro diketopiperazine and sarcosine (Meienhofer, 1970) and the cleavage of D-Val-Pro from a solid-phase resin with the formation of its diketopiperazine (Gisin and Merrifield, 1972) were thought to be catalysed by the carboxylic acid group of the C-terminal sarcosine in the former case and of Boc-D-Pro-OH in the latter case (which was present to couple to the resin bound dipeptide). The early release of aspartic acid from peptides and proteins during hydrolysis of proteins with
weakly acidic solutions is thought to be mediated through an internal mechanism involving the nearby unionised carboxyl group of the aspartic acid side chain (Leach and Lindley, 1953a,b). The catalysed aminolysis of active esters by carboxylic acids has also been widely reported (Handford et al., 1965; Nakamizo, 1969). In all these cases, where carboxylic acids have acted as catalysts the proposed mechanisms of action involved a concerted push-pull reaction in which the unionised carboxyl group of the catalyst acted through a hydrogen bonded cyclic intermediate.

Since, in the experiments reported here, peptide bonds adjacent to single glutamic acid residues were not cleaved, it could be that spatially neighbouring side chain carboxylic acid groups (but not necessarily those immediately adjacent in sequence) must be present as catalysts for the reaction described by Rns. (23) to occur.

The fact that the presence of ethanolamine reduces the degree of polypeptide chain fission is probably due to competition by the amine with the peptide carbonyl groups for reaction with the O-acyl urea. Rapid reaction of activated carboxyl groups with ethanolamine reduces the number of free protonated carboxyl groups available for catalysing peptide chain fission which would further diminish the degree of cleavage.

Further studies would be required to establish whether Rns. (23) and (24) are the correct mechanism for the cleavage reaction. These could include:
(i) Measuring the rate of chain fission as a function of pH to verify the requirement for $H^+$ and for the catalytic carboxylic acid group to be in the protonated form.

(ii) The addition of other carboxylic acids in cleavage studies on simple peptides containing single glutamic acid residues to determine if these would act as catalysts.

(iii) Cleavage studies on peptides containing sequences of glutamic acid (e.g. gastrin which contains a Glu-Glu-Glu-Glu sequence) and on sequential copolypeptides containing a fairly high proportion of glutamic acid residues (e.g. polymers of the form \((-\text{Glu-X-X-})_n\)) to determine whether proximity or concentration of carboxylic acid groups are important in the reaction.


Avignon, M. and Huong, P.V. (1970), Biopolymers 9, 427.

Bovey, F.A., Ryan, J.J. and Hood, F.P. (1968), Macromolecules 1, 305.
Bradbury, E.M., Crane-Robinson, C. and Hartman, P.G. (1973), Polymer 14, 543.
Burgess, A.W. and Leach, S.J. (1973), Biopolymers 12, 2599.
Burgess, A.W., Paterson, Y. and Leach, S.J. (1975),
J. Polymer Sci., Symp. No. 49, 75.
Israel Jnl. Chem. 12, 239.
Sci., U.S. 72, 1221.
Burstein, Y., Fridkin, M. and Patchornik, A. (1967), Israel
Jnl. Chem. 5, 64P.
Bystrov, V.F., Portnova, S.L., Isetlin, V.I., Ivanov, V.T.
Carroway, K.L. and Koshland, D.E. (1972), in Methods in
Enzymology VXXV, Ed. S.P. Colowick and N.O. Kaplan,
123, 129.

1976, 92.


Dorman, D.R., Torchia, D.A. and Bovey, F.A. (1973), 
Macromolecules 6, 80.

Dygert, M.K., Taylor, G.T., Cardinaux, F. and Scheraga, H.A. 
(1976), Macromolecules 9, 794.


62, 613.


Fitzwater, S., Hodes, Z.I. and Scheraga, H.A. (1978), 
Macromolecules, in press.

Flory, P.J. (1953), Principles of Polymer Chemistry, 
Chapter XIV, Cornell University Press, Ithaca, N.Y.

Flory, P.J. (1969), 'Statistical Mechanics of Chain Molecules', 
New York: Interscience.

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

89, 6807.


Goodman, M., Listowsky, I., Masuda, Y. and Boardman, F. (1963), Biopolymers 1, 33.


Goodman, M., Ueyama, N. and Naider, F. (1975a), Biopolymers 14, 901.


Hodes, Z.I., Némethy, G. and Scheraga, H.A. (1978a), Biopolymers, in press.

Hodes, Z.I., Némethy, G. and Scheraga, H.A. (1978b), Biopolymers, in press, accompanying previous paper.


IUPAC-IUB Commission on Biochemical Nomenclature (1970), Biochemistry 9, 3471.
Krumdieck, C.L. and Baugh, C.M. (1968), Biochemistry 8, 1568.
Markley, J.L. (1975a), Biochemistry 14, 3546.
Markley, J.L. (1975b), Biochemistry 14, 3554.
Mattice, W.L. (1975), Macromolecules 8, 644.
Mattice, W.L. (1977a), Macromolecules 10, 511.
Mercier, J.C., Grosclaude, F. and Ribadeau-Dumas, B. (1972), Milchwissenschaft 27, 402.
Miller, W.G. and Goebel, C.V. (1968), Biochemistry 7, 3925.
Nagayama, K. and Wada, A. (1973), Biopolymers 12, 2443.
Nall, B.T. and Baldwin, R.L. (1977), Biochemistry 16, 3572.
Oyama, S., Yoshioka, H., Koyama, Y. and Nakatsu, K. (1978),
to be published, quoted in Koyama et al. (1977).
Palumbo, M., Darin, S., Bonora, G.M. and Toniolo, C. (1976),
Makromol. Chem. 177, 1477.
Perutz, M.F. (1969), The Harvey Lectures (1967-68), Series 63,
Platzer, K.E.B., Ananthanarayanan, V.S., Andreatta, R.H. and
Scheraga, H.A. (1972), Macromolecules 5, 177.
Commun. 81, 176.
Transitions in Biopolymers', Chapters 4, 8 and 9,
60, 481.
Ponnuswamy, P.K. and Sasisekheran, V. (1970), Int. J. Protein
Res. 2, 37.
Popov, E.M., Dashevskii, V.G., Lipkind, G.M. and Arkhipova, S.F.
(1968a), Mol. Biol. (USSR) 2, 491.
Popov, E.M., Lipkind, G.M., Arkhipova, S.F. and Dashevskii, V.G.
(1968b), Mol. Biol. (USSR) 2, 498.
Commun. 42, 1069.

Proteins Synthesis Group (1975), Scientia Sinica XVIII, 745.


213.


Schimmel, P.R. and Flory, P.J. (1968), J. Mol. Biol. 34, 105.


Stark, G.R. (1968), Biochemistry 7, 1796.


Tanaka, S. and Nakajima, A. (1971a), Polymer J. (Japan) 2, 717.
Tanaka, S. and Nakajima, A. (1971b), Polymer J. (Japan) 2, 725.


Von Dreele, P.H., Rae, I.D. and Scheraga, H.A. (1978), Biochemistry 17, 956.
Wu, C.C., Komoraski, R.A. and Mandelkern, L. (1975), Macromolecules 8, 635.


APPENDICES
## APPENDIX I

### STANDARD RESIDUE DATA USED IN CHAPTER II

### SECTION 2.C

### VALINE

<table>
<thead>
<tr>
<th>16</th>
<th>3-0.4437972</th>
<th>0.3305252-0.0993245</th>
<th>0.9950551</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.35</td>
<td>3 1 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.35</td>
<td>3 1 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.35</td>
<td>3 1 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### ALANINE

<table>
<thead>
<tr>
<th>10</th>
<th>1-0.9437972</th>
<th>0.3315252-0.0993245</th>
<th>0.9950551</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.35</td>
<td>3 1 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### GLYCINE

<table>
<thead>
<tr>
<th>7</th>
<th>0-0.9335424</th>
<th>0.3583627-0.097349</th>
<th>0.997556</th>
</tr>
</thead>
</table>

### BETA-METHYL-ALA

<table>
<thead>
<tr>
<th>13</th>
<th>2-0.9437972</th>
<th>0.3305252-0.0993245</th>
<th>0.9950551</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.35</td>
<td>3 1 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.35</td>
<td>3 1 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table Details

- The table includes detailed residue data for Valine, Alanine, Glycine, and Beta-Methyl-Ala.
- Each residue is represented with specific values for different conditions.
- The data is structured in a tabular format with columns for residue type, values, and conditions.
- Each residue type has multiple entries indicating different conditions or values.
- The table is organized to show a clear comparison of residue data across different conditions.
### H-E-GLUTATHIONE

<table>
<thead>
<tr>
<th></th>
<th>24</th>
<th>7-0.9378933</th>
<th>0.3469239-0.081984</th>
<th>0.996419</th>
</tr>
</thead>
<tbody>
<tr>
<td>-14</td>
<td>5</td>
<td>1.35</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.358356-0.518654-0.778409</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000000-0.000000</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.358354-0.518676-0.774339</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.396465-0.396531-0.00151</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.258444-0.532065-0.805950</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.808039-0.324771-0.41569</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2443649-0.534667-0.92193</td>
<td>1</td>
<td>21</td>
</tr>
</tbody>
</table>

### BETA-METHYL-VAL

<table>
<thead>
<tr>
<th></th>
<th>19</th>
<th>4-0.9437972</th>
<th>0.3305252-0.093245</th>
<th>0.9950551</th>
</tr>
</thead>
<tbody>
<tr>
<td>-18</td>
<td>4</td>
<td>1.35</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.55</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.342018-0.533775-0.773375</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9999354-0.0005248-0.007595</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.329374-0.402477-0.954275</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.329374-0.941490-0.973314</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

N 13 | -4.59 | 11 | 7 | 10 |
<table>
<thead>
<tr>
<th></th>
<th>1.9763</th>
<th>-0.8167</th>
<th>-1.1833</th>
<th>CB 9</th>
<th>0.10</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.9587</td>
<td>1.4440</td>
<td>0.00</td>
<td>C 7</td>
<td>5.80</td>
<td>11</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2.1848</td>
<td>2.3835</td>
<td>0.37</td>
<td>O 17</td>
<td>-4.95</td>
<td>2</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5.0802</td>
<td>0.8087</td>
<td>-1.1717</td>
<td>CG 6</td>
<td>-0.86</td>
<td>0</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1.4730</td>
<td>0.2009</td>
<td>-2.4903</td>
<td>CG 2</td>
<td>-0.86</td>
<td>0</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>1.4730</td>
<td>2.2572</td>
<td>-1.0711</td>
<td>CG 3</td>
<td>-0.86</td>
<td>0</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3.8790</td>
<td>1.3905</td>
<td>-2.0147</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3.8667</td>
<td>-1.2474</td>
<td>-0.2495</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3.8667</td>
<td>0.2176</td>
<td>-1.2516</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.8458</td>
<td>-0.7427</td>
<td>-3.3333</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.8316</td>
<td>0.4253</td>
<td>-2.5793</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.3380</td>
<td>0.2966</td>
<td>-2.4986</td>
<td>H2G1</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.8458</td>
<td>-2.3340</td>
<td>-1.9141</td>
<td>H2G3</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.3830</td>
<td>-2.2629</td>
<td>-1.0794</td>
<td>H2G3</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.8316</td>
<td>-2.6959</td>
<td>-0.1400</td>
<td>H2G3</td>
<td>0.32</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3.2771</td>
<td>1.5756</td>
<td>9.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**AMINO - H2**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**AMINO - COCH3**

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>6</th>
<th>7</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**GARBOXYL - COOH**

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**GARBOXYL - NHCH3**

<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>2</th>
<th>0</th>
<th>2</th>
<th>2</th>
<th>4</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**COCH3**

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
APPENDIX I(b)  INPUT SPECIFICATIONS FOR THE STANDARD RESIDUE DATA

1. Atomic Numbering System

Atomic coordinates are given in the standard residue data set for each amino acid residue, N-terminal end group and C-terminal end group, in the local coordinate system. Within each residue or end group data set, the atoms are arranged in the order required for generation.

For amino acid residues, number 1-7 are assigned to the backbone atoms and those directly attached to the backbone, as shown in the illustration for β-methylalanine (p.223) (for proline-type residues, no.2 is assigned to Cδ). Numbers 8 and higher are used for the remaining side chain atoms, beginning with the H atoms attached to Cβ, and moving out the side chain. When the side chain is generated, the dihedral angles are considered in order (χ1, χ2, etc.), except that in a case of unequal branch length the shorter branch is generated first (e.g. isoleucine, where the order is χ1, χ2, χ2', χ3'). All atoms whose positions are determined by a particular dihedral angle must be numbered consecutively, although the order within that group is arbitrary. For example, in β-methylalanine χ1 fixes the two Hβ atoms and Cγ (nos.8, 9, 10), χ2 fixes the three HY atoms (nos.11, 12, 13).

The numbering of atoms in the end groups follows the "left-to-right" order in which their dihedral angles are considered during generation. The scheme also reflects the fact that when a given dihedral angle is adjusted, the coordinates of all dependent atoms are adjusted, i.e. all...
atoms "before" the corresponding bond in N-terminal end groups, and "after" it in C-terminal end groups. This requires that in an N-terminal end group the first (i.e. leftmost) atom of each rotatable bond must be numbered higher than all preceding atoms in the chain. In a C-terminal end group the second (i.e. rightmost) atom of each rotatable bond must be numbered lower than all following atoms in the chain. For example, in the N-acetyl end group, the three methyl protons depend on the dihedral angle for rotation about the C-C bond, and the protons are assigned nos. 1-3 and the methyl C is no. 4. Further, the coordinates of the whole methyl group plus the carbonyl oxygen depend on the carbonyl C-N dihedral angle, so the carbonyl oxygen is no. 5 and the carbonyl C is no. 6.

The atomic numbering scheme described above is also the basis for the indexing of the atoms in the energy computation, i.e. the "limiting numbers" used to select 1-4 and 1-5 interactions. The numbering scheme is generally adequate for this purpose for the amino acid residues and end groups in the standard data sets supplied. However, special coding is required in certain cases, e.g. β-methyl-valine. Because the 1-4 non-bonded interactions are calculated for all atoms between NSN14 and NFN14 (see Card Type 6, Section (2)) specific modifications were required for ECEPP to cope with a tert-butyl group. The numbering system for the tert-butyl group is as follows:
Thus, atom no. 9 has 1-4 interactions with atoms 11, 12, 13 and with atoms 17, 18, 19, but not those between 13 and 17. Thus, in the data set for \( \beta \)-methylvaline 1-4 interactions for CG2 are given for atoms 11-13 and those for atoms 17-19 are explicitly calculated in the modified form of ECEPP in sub-routine Int.1-4.

Atomic Numbering System for Amino Acid Residues

(example: \( \beta \)-methylalanine)
Atomic Numbering System for End Groups

Amino end group (Amino-COCH₃)

\[
\begin{align*}
\text{H}^1 & \quad \text{O}^5 & \quad \text{H}^8 \\
\text{H}^2 & \quad \text{C}^4 & \quad \text{C}^6 & \quad \text{N}^7 & \quad \text{C}^9 \\
\text{H}^3 & \quad & \quad & \quad & \\
\end{align*}
\]

Carboxyl end group (Carboxyl-NHCH₃)

\[
\begin{align*}
\text{O}^1 & \quad \text{H}^3 & \quad \text{H}^5 \\
\quad & \quad \text{C}^0 & \quad \text{N}^2 & \quad \text{C}^4 & \quad \text{H}^6 \\
\quad & \quad & \quad & \quad \text{H}^7 & \\
\end{align*}
\]

(data set numbering system)

\[
\begin{align*}
\text{H}^2 & \quad \text{H}^4 \\
\quad & \quad \text{N}^1 & \quad \text{C}^3 & \quad \text{H}^5 \\
\quad & \quad & \quad & \quad \text{H}^6 & \\
\end{align*}
\]

(residue numbering system)

2. Amino Acid Residue Example: β-methylalanine

Card Type 1 (one card of this type):

Variables: (TITLE(L, I), L=1,4)

Format: (4A4)

TITLE: Name of the residue.

Card Type 2 (one card):

Variables: NATOMS(I), NCHI(I), SNT2H(I), CSTH2(I), SDEL(I), CDEL(I)

Format: (2I5, 4F10.7)
NATOMS  Number of atoms in the residue (13 in this example). Maximum allowed = 26.
NCHI  Number of variable side chain dihedral angles (rotatable bonds) (3 in the example). Maximum allowed = 7.
SNTH2  Sine of the bond angle N\alpha C' minus \pi.
CSTH2  Cosine of the preceding quantity.
SDEL  Sine of the negative of the angle between the C'-N bond vector and the x-axis of the local coordinate system.
CDEL  Cosine of the preceding quantity.

Card Type 3 (one card):

Variables:  KNDRES(I), NTOR(I), NGEOM(I)

Format:  (3I4)

KNDRES  The residue type number, which if negative indicates that there is at least one side chain intrinsic torsional term (-2 in the example).
NTOR  The number of side chain intrinsic torsional terms (2 in the example). Maximum allowed = 8.
NGEOM  Supplied for proline-type residues only; specifies the pyrrolidine ring geometry as follows: 1 (Planar), 2 (Up) or 3 (Down).

Card Type 4 (one card for each side chain intrinsic torsional term):

Variables:  AR(J,I), NBB(J,I), NSS(J,I), NANG(J,I)

Format:  (F10.4, 3I5)

Each side chain torsional term will ultimately be calculated from the first three quantities as follows:

UTOR = (AR) [1 + (NSS)ccs(NBB)\chi]
AR \( U_o/2 \), where \( U_o \) is the torsional barrier height in kcal/mole.

NBB Number defining torsional symmetry; has values 2 or 3.

NSS Defines the position of the torsional minimum; has values +1 or -1.

NANG Number of the angle to which this term applies, in the numbering system of the ANGLES array.

In the example, card no.4 indicates that a torsional term is to be computed for \( \chi^1 \) (angle no.4), as follows:

\[ \text{UTOR} = (1.35)(1 + \cos 3\chi^1) \]

**Card Type 5** (one card for each variable side chain dihedral angle \( \chi \)):

Variables: (CHIANG(L,J,I), L=1,3), NDPT1(J,I), NDPT2(J,I), NUM(J,I), LRT1(J,I)

Format: (5x, 3F9.6, 5x, 3I3, 110)

CHIANG Direction cosines of the bond to which the dihedral angle applies, in the local coordinate system, with respect to the x-, y- and z- axes.

NDPT1 Number of the first atom forming the bond.

NDPT2 Number of the second atom forming the bond.

NUM Number of atoms whose coordinates depend only on the value of that dihedral angle.

LRT1 The lowest-numbered atom following the bond in the chain (not used in the present program).

In the example, card no.6 gives the three direction cosines of the first rotatable bond (angle \( \chi^1 \)), and indicates that it connects atom no.3 (\( c^\alpha \)) to atom no.5 (\( c^\beta \)), that the
coordinates of 3 other atoms are defined by the value of $\chi^1$
(they are atoms no. 8, 9 and 10, but this is not stated
explicitly), and that the lowest-numbered atom following
the C$^\alpha$-C$^\beta$ bond is no. 8.

Card Type 6 (one card):

Variables:  ALPHA(l,I), LTYPE(l,I), CHG(l,I), NSN15(l,I),
            NSN14(l,I), NFN14(l,I)

Format:  (37X, A3, I2, F10.6, I5, 2X, 4I3)

This card contains data for the amino nitrogen atom.

No coordinates are given, since these will be assigned during
the generation of the preceding residue.

ALPHA  Characters giving the chemical type of the atom.

LTYPE  Atom type for nonbonded interactions.

CHG  Partial charge, as $q(332/2)^{1/2}$, where $q$ is in
electronic charge units.

The following three integer variables give the limiting
numbers for the intraresidues 1-4 and 1-5 interactions.

NSN15  The number of the first atom which interacts 1-5.

NSN14  The number of the first atom which interacts 1-4.

NPN14  The number of the last atom which interacts 1-4.

In the present example, the limiting numbers are 11, 7 and
10, respectively, meaning that the amino N has 1-5
interactions with atoms no. 11 and higher (12 and 13), and
1-4 interactions with atoms no. 7, 8, 9 and 10.

Card Type 7 (one card for each atom, except amino N):

Variables:  (XCOORD(L,J-1,I),L=1,3), ALPHA(J,I), LTYPE(J,I),
            CHG(J,I), NSN15(J,I), NSN14(J,I), NFN14(J,I)

Format:  (3F10.3, 7X, A3, I2, F10.6, I5, 2X, 4I3)
With the exception of \textit{XCOORD}, these variables have been described in connection with card type 6. \textbf{XCOORD} \hspace{1cm} Coordinates \((x, y \ and \ z)\) of the atom in the local coordinate system.

\underline{Card Type 8} (one card):

\textbf{Variables:} \((\text{XCOORD}(L,NATOM,I),L=1,3)\)

\textbf{Format:} \((3F10.3)\)

This card contains only the coordinates of the amino \(N\) of the following residue, in the local coordinate system of the present residue. Its purpose is to provide a starting point for the generation of the following residue.

3. \textbf{Amino End Group Example: Amino-COCH}_3

\underline{Card Type 1} (one card):

\textbf{Variables:} \((\text{TITLE}(L,I+NREG),L=1,4)\)

\textbf{Format:} \((4A4)\)

\textbf{TITLE: Name of the end group.}

\underline{Card Type 2} (one card):

\textbf{Variables:} \(\text{NDATOM}(I), \text{NBOND}(I), (\text{NNDPT1}(L,I),\text{NNDPT2}(L,I),L=1,6)\)

\textbf{Format:} \((14I5)\)

\textbf{NDATOM} \hspace{1cm} The total number (9 in the example) of atoms for which data will be given, including the atoms in the residue itself (6 in the example) and the \(N, H\) and \(Ca\) of the following full residue. \((\text{NDATOM} \hspace{0.5cm} \text{is limited by array dimensions to a maximum of 22.})\)

\textbf{NBOND} \hspace{1cm} The number (2 in the example) of rotatable bonds \((i.e. \text{the number of variable dihedral angles})\) including the bond preceding the amino \(N\), but not including the \(N-Ca\) bond.
NNDPT1 For each rotatable bond, the number of the first atom forming the bond.

NNDPT2 For each rotatable bond, the number of the second atom forming the bond.

In the present example, the sequence 4,6,6,7,0,0, etc., indicates that the rotatable bonds connect atoms no.4 and 6, (C-C' and atoms no.6 and 7 (C'-N).

Card Type 3 (one card):

Variables:  KNDEND(I), AAN(I), NBBN(I), NSSN(I), NANGN(I)

Format:  (I4, F10.4, 3I5)

KNDEND The group number (-4), which if negative indicates that there is a torsional term.

The torsional term will ultimately be calculated from the next three quantities, as follows:

UTOR = (AAN) [1 + (NSSN)cos(NBBN)θ]  

AAN  \( U_o/2 \), where \( U_o \) is the torsional barrier height in kcal/mole.

NBBN Torsional symmetry number; has values 2 or 3.

NSSN Defines the position of the torsional minimum; has values +1 or -1.

NANGN Number of the rotatable bond to which this term applies.

In the present example, this card conveys the information that a torsional term is to be computed for rotatable bond no.2 (C'-N), as follows:

UTOR = (10.0)(1 - cos 2θ)
(Note that data for only one intrinsic torsional term can be read for each end group. In the C-terminal ethyl ester end group there are two such terms, and the second term is dealt with in special coding in sub-routines INITOR and TORTOT.)

**Card Type 4** (one card):

Variables: \((LR(L,I),L=1,NB)\)

Format: \((14I5)\)

\(LR\)

For each rotatable bond, the lowest-numbered atom following that bond in the chain (not used in the present program).

**Card Type 5** (three cards):

Variables: \((MSX15(L,I),L=1,ND)\)

Variables: \((MSX14(L,I),L=1,ND)\)

Variables: \((MFX14(L,I),L=1,ND)\)

Format: \((14I5)\)

These integer variables define for each of the atoms in the end group itself (not including N, HN and CA) the limiting numbers for 1-4 and 1-5 interactions with the atoms in the adjacent full residue. The limiting numbers correspond to the internal numbering system of the adjacent residue.

\(MKS15\)

The number of the first atom which interacts 1-5.

\(NSX14\)

The number of the first atom which interacts 1-4.

\(MFX14\)

The number of the last atom which interacts 1-4.

In the present examples, the 6th field of cards 5, 6 and 7 contains the integers 7, 4 and 6, respectively. This indicates that atom no. 6 of the end group (the carbonyl carbon) interacts 1-5 with atoms no. 7 (the carbonyl oxygen) and higher of the adjacent full residue, and interacts 1-4 with atoms no. 4, 5 and 6 (the \(H^\alpha\), \(C^\beta\) and \(C^\prime\)).
Card Type 6 (one card for each atom, including N, HN and CA of the following full residue):

Variables: (ENOORD(L,J,I),L=1,3), EALPHA(J,I),
NDTYPE(J,I), ECHG(J,I), MSN15(J,I),
MFN14(J,I)

Format: (3F10.3, 7X, A3, I2, F10.6, I5, 2X, 4I3)

ENOORD Coordinates of the atom in the local coordinate system.
EALPHA Characters giving the chemical type of the atom.
NDTYPE Atom type in nonbonded interactions.
ECHG Partial charge as $q(332/2)^{1/2}$, where $q$ is in electronic charge units.

The following three integer variables give the limiting numbers for the intraresidue 1-4 and 1-5 interactions, not including the attached N, HN and CA:

MSN15 The number of the first atom which interacts 1-5.
MSN14 The number of the first atom which interacts 1-4.
MFN14 The number of the last atom which interacts 1-4.

In the present example, the last three fields in cards 8-16 indicate that there are no intraresidue 1-5 interactions, and the three methyl hydrogens interact 1-4 with atom no.5 (the carbonyl oxygen).

4. Carboxyl End Group Example: Carboxyl-NHCH$_3$

The standard residue data for carboxyl end groups are grouped with the amino end group data and are read by the same statements. Unless otherwise stated here, the data are organized and interpreted in the same way. This section will describe only those points which differ for the carboxyl end groups.
In these cases, the atoms are numbered in two ways:
(1) as parts of the data set, taking the C' of the preceding full residue as atom no. 0 and the attached oxygen as atom no. 1, and (2) as parts of the residue itself, ignoring the preceding C' and oxygen atom.

**Card Type 2:**

**NDATOM** The total number (7 in the example) of atoms for which data will be given, including the atoms in the residue itself (6 in the example) and the carbonyl oxygen of the preceding full residue.

**NBOND** The number (2 in the example) of rotatable bonds, including the bond to the C' of the preceding full residue.

**NNDPT1** For each rotatable bond, the number of the first atom forming the bond, using the "data set" numbering system.

**NNDPT2** The number of the second atom forming the bond, as in the preceding definition.

In the present example, the sequence 0, 2, 2, 4, 0, etc., indicates that the rotatable bonds are between atoms no. 0 and 2 (C'-N), and between atoms no. 2 and 4 (N-C), in the "data set" numbering system.

**Card Type 3:**

**NANGN** Number of the rotatable bond to which the torsional term (if any) applies, in the numbering system of the preceding full residue.

In the present example, the value 3 indicates that the torsional term applies to the ω dihedral angle (around the C'-N) of the preceding residue.
Card Type 4:

LR Defined as in the preceding section on the amino end groups, but using the "residue" numbering system.

Card Type 5:

The integer arrays MSX15, etc., here contain limiting numbers for each of the 7 atoms in the preceding full residue which are capable of participating in a 1-4 interaction with end group atoms, i.e. atoms no.1-7 in the full residue numbering system. The limiting numbers designate atoms in the end group, using the "residue" numbering system. Note that the method here is the reverse of that used for the amino end groups.

In the present example, the third field of cards 5, 6 and 7 contains the integers 4, 2 and 3, respectively. These integers are limiting numbers for atom no.3 (Ca) of the preceding full residue, and indicate that this atom has 1-5 interactions with atoms no.4, 5 and 6 (the methyl H atoms), and 1-4 interactions with atom no.2 (the amino H) and atom no.3 (the methyl C) of the end group.

Card Type 6 (one card for each atom, including the carbonyl oxygen of the preceding full residue):

The limiting numbers in arrays MSN15, MSN14 and MFN14 are in the "residue" numbering system. However, the values given for the elements of these arrays (the last three fields of each card) are not for the atom named and described on the same card, but for the atom named on the following card. In the present example, the sequence 0,4,6
appears in the last three fields of card 9. This indicates that the atom named on card 10 (the amino H) has no intramolecular 1-5 interactions, but has 1-4 interactions with atoms no.4, 5 and 6 in the residue numbering system, i.e. the three methyl hydrogens.
APPENDIX II

MATHEMATICAL DERIVATION OF STATISTICAL WEIGHTING
PROCEDURE USED IN CHAPTER II, SECTION 3.B

\begin{equation}
\langle \mathcal{T} \rangle^{g^+g^-} = p_t \langle \mathcal{T} \rangle^t + p_{g^+} \langle \mathcal{T} \rangle^{g^+} + p_{g^-} \langle \mathcal{T} \rangle^{g^-}
\end{equation}

where \( p_t + p_{g^+} + p_{g^-} = 1 \), and \( t, g^+, g^- \) are abbreviations for the trans, gauche\(^+\), gauche\(^-\) rotameric states defined by \( \chi^1 \) the first side chain torsion angle.

let \( r = t \) or \( g^+ \) or \( g^- \), then \( E^r = E(\phi, \psi, \chi^r) \)

Putting \( p^r = \frac{z^r}{z_t + z_{g^+} + z_{g^-}} \)

Now

\begin{equation}
\frac{z^r}{z_t + z_{g^+} + z_{g^-}} \langle \mathcal{T} \rangle^r = \frac{z^r}{(z_t + z_{g^+} + z_{g^-})} (z^r)^{-1} \sum_T(\phi, \psi) \exp(-E^r/RT) \begin{pmatrix} \phi \psi \end{pmatrix}
\end{equation}

Therefore

\begin{equation}
\frac{z^r}{z_t + z_{g^+} + z_{g^-}} \langle \mathcal{T} \rangle^r = (z_t + z_{g^+} + z_{g^-})^{-1} \sum_T(\phi, \psi) \exp(-E^r/RT) \begin{pmatrix} \phi \psi \end{pmatrix}
\end{equation}

Substituting eq.4 in eq.1 for \( r = t, g^+ \) and \( g^- \)

\begin{equation}
\begin{split}
\langle \mathcal{T} \rangle^{g^+g^-} &= (z_t + z_{g^+} + z_{g^-})^{-1} [\sum_T(\phi, \psi) \exp(-E^g^-/RT) \\
&+ \sum_T(\phi, \psi) \exp(-E^g^-/RT) + \sum_T(\phi, \psi) \exp(-E^g^-/RT)]
\end{split}
\end{equation}

Therefore

\begin{equation}
\begin{split}
\langle \mathcal{T} \rangle^{g^+g^-} &= (z_t + z_{g^+} + z_{g^-})^{-1} \sum_T(\phi, \psi) [\exp(-E^t/RT) \\
&+ \exp(-E^g^+/RT) + \exp(-E^g^-/RT)]
\end{split}
\end{equation}
APPENDIX III  LISTING OF PROGRAM FLORY CR REFERRED TO IN
CHAPTER II, SECTION 3.D

PROGRAM FLORY CR (INPUT,OUTPUT,TAPE10,TAPES=INPUT,TAPES=OUTPUT)

PROGRAM FOR CALCULATING CHARACTERISTIC RATIOS AND STATISTICAL
WEIGHTS FOR TRANS POLYAMINO-ACIDS AND COPOLYMERS COMPOSED OF
REPEATING SEQUENCES OF TRANS PEPTIDES.
THE PROGRAM IS DIVIDED INTO THREE MAIN PARTS. THE MAIN PROGRAM
CALCULATES THE AVERAGE TRANSFORMATION MATRIX (TAVE) AND THE
PARTITION COEFFICIENT (Z) FOR THE SPECIFIED AMINO ACID RESIDUES.
THE GEOMETRY OF THE PEPTIDE UNIT IS DEFINED BY ANGLES ALPHA,
BETA AND GAMMA (=XI,THETA,-ETA RESPECTIVELY IN FLORY'S NOTATION)
AND ARE SET BY A DATA STATEMENT IN SUB-Routine COORD. FLAGS ARE
SET AS PART OF THE DATA READ TO ALLOW THE USER TO CALL EITHER
THE SUBROUTINE THAT CALCULATES THE CR FOR THE POLYAMINACID
(SUBROUTINE CRMONO) OR THE SUBROUTINE THAT CALCULATES THE CR
OF THE COPOLYMER (SUBROUTINE CRGROUP). BOTH SUBROUTINES MAY
BE CALLED SO THAT THE CR OF THE MONOMER UNITS CAN BE CALCULATED
IN ADDITION TO THAT OF THE COPOLYMER.

COMMON TITLE(20,6)
DIMENSION TNC(3,3),TCC(3,3),TA(1,3),TAUXA(3,3),
1 TA(3,3),TG(3,3),TSCOS(37),TSSIN(37),TAUXC(3,3),TAUJZ(2,3,3)
DIMENSION ETOT(37,37),TAVE(3,3)
COMMON/TAVE/TAVE(1,3),TAVE(3,3),TAVE(3,3)
DATA TNC/ 1.0,1.0*0.0/, TCC / 1.0, 1.0*0.0 /
DATA TSCOS/ 1.0, 1.0*0.0/, TCC / 1.0, 1.0*0.0 /
DATA TSSIN/ 0.0, 0.0*0.0/, TCC / 1.0, 1.0*0.0 /

109 FORMAT(2I5)
110 FORMAT(2X,E12.5)
115 FORMAT(5X,I3,9X,I3)
116 FORMAT(4X,7HNC ROT.,4X,7HCC ROT.)
117 FORMAT(13A6)
118 FORMAT(/,1X,13A6,/) 200 FORMAT(3H0 ENERGY VALUES FOR ALL ROT ANGLES,/) 202 FORMAT (1H ,3X,19F6.2) 203 FORMAT(4X,18F6.2) 323 FORMAT (1HU,20X,4H2 = ,F10.0) 324 FORMAT (25X,3(F12.3),/25X,3(F12.3),/25X,3(F12.3)) 335 FORMAT(1HU,23H TAVE FOR TRANS PEPTIDE) 336 FORMAT(1HU,37H PARTITION FUNCTION FOR TRANS PEPTIDE) 340 FORMAT(1HU,32H WEIGHTED TAVE FOR TRANS PEPTIDE) 342 FORMAT(1HU,1X,13H JOB NUMBER = )

C... READ NUMBER OF JOBS I.E. NO OF MONOMER UNITS
C
READ(5,109) NJ
DO 1 IN = 1,NJ
WRITE(6,342) IN
III = IIN
C
C... READ HEADING FOR JOB
C
READ (5,117) (TITLE(I,IIN),I=1,13)
WRITE (6,118) (TITLE(I,IIN),I=1,13)
C
C... READ FLAG FOR CALCULATING (1) C.R. OF POLYMER COMPOSED ONLY OF
C... THIS MONOMER UNIT I.E. IMONO = 1, OR (2) C.R. OF
C... COPOLYMER I.E. ICOPOL = 1.
C
READ (5,109) IMONO, ICOPOL
C
C... READ ACTUAL CONFORMATION ENERGIES FOR ESTIMATE OF C.R.
C... TRUNCATE ENERGIES ABOVE 20K.CAL/MOLE
C
WRITE (6,200)
DO 36 L=1,37
DO 36 K=1,37
READ (10,110) ETOT(L,K)
36 IF(ETOT(L,K) .GE. 20.0) ETOT(L,K) = 20.0
C
C... CALCULATE ROTATION MATRICES TA,TJ,TG
C
CALL COORD(TA,TB,TG)
C
C... SET T-AVE MATRIX ELEMENTS EQUAL TO ZERO
C
DO 12 K=1,3
DO 12 L=1,3
12 TAVE(K,L) = 0.0
C
C... CALCULATE T-AVE MATRIX ELEMENTS
C
K=1.967/1000.0
T = 233.
WRITE (6,116)
DO 2 KNC = 1,36
DO 2 KCC = 1,36
DO 11 K =2,3
11 CONTINUE
TGC(K,K) = TSCOS(KCC)
TNC(K,K) = -TSCOS(KNC)
11 CONTINUE
TGC(3,2) = TSSIN(KCC)
TNC(3,2) = -TSSIN(KNC)
TGC(2,3) = TSSIN(KCC)
TNC(2,3) = -TSSIN(KNC)
CALL MATMUL(TA,TNC,3,TAUXA)
CALL MATMUL(TAUXA,TB,3,TAUXB)
CALL MATMUL(TAUXB,TCC,3,TAUXC)
CALL MATMUL(TAUXC,TG,3,TADJ2)
WEXP = EXP(-ETOT(KNC,KCC)/(R*T))
DO 14 K=1,3
DO 14 L =1,3
14 TAVE(K,L) = TAVE(K,L) + TADJ2(K,L)*WEXP
2 CONTINUE
WRITE (6,335)
WRITE (6,324)((TAVE(K,L),K=1,3,L=1,3)
C
C... DETERMINE THE PARTITION FUNCTION FOR THE DATA SET
C
Z = 0.0
DO 5 KNC=1,36
DO 5 KCC=1,36
5 Z = Z + EXP(-ETOT(KCC,KNC)/(R*T))
WRITE (6,336)
WRITE (6, 323) Z
C
C WRITE MATRIX ELEMENTS USING A BOLTZMANN METHOD
C
DO 14 K = 1, 3
    DO 14 L = 1, 3
14   TAVE(K, L) = TAVE(K, L) / Z
WRITE (6, 324) (( TAVE(K, L), K = 1, 3), L = 1, 3)
DO 10 (30, 31, 32), III

C
C IF THE C.R. OF THE MONOMER UNIT IS REQUIRED (IMONO = 1) THE
C CALL TO SUBROUTINE CRMONO IS MADE HERE.
C
IF (IMONO .EQ. 1) CALL CRMONO(TAVE)
1 CONTINUE
C
C IF THE C.R. OF THE COPOLYMER IS REQUIRED (ICOPOL = 1) THE
C CALL TO SUBROUTINE CRCPOL IS MADE HERE.
C IF (ICOPOL .EQ. 1) CALL CRCPOL
END

C
C SUBROUTINE CRMONO (TAVE)
C
C THIS SUBROUTINE CALCULATES THE CHARACTERISTIC RATIO OF THE
C POLYMER COMPOSED SOLELY OF THE SPECIFIED POLYMER UNIT =
C TFIM = (L + TAVE)(E - TAVE) ** -1 - ((2/XRES) * TAVE * (E - TAVE)** NRES)** -1
C
C FOR THE THEORETICAL DEVELOPMENT OF THIS EXPRESSION SEE REFERENCE:
C
DIMENSION TPOW(3, 3), TEMPI(3, 3), TEMPl(3, 3), TEMPIP(3, 3), TEMPl(3, 3),
   1TEMPI(3, 3), TEMPl(3, 3), TEMPl(3, 3), IMP(3, 3), TFIM(3, 3), TAVE(3, 3)
330 FORMAT (1H0, 25X, F10.4)
339 FORMAT (1H0, 33H CHARACTERISTIC RATIO FOR MONOMER, 1X, I5, 1X
   114H RESIDUES LONG)
334 FORMAT (25X, 3F10.2, /, 25X, 3F10.2, /, 25X, 3F10.2)
337 FORMAT (1H0, 17H TFIM FOR PEPTIDE)
111 FORMAT (215)
112 FORMAT (1H0, 37H TEMP3 IS SINGULAR WITH DETERMINANT = , F4.1)
C
C READ THE NUMBER OF RESIDUES IN THE POLYPEPTIDE AND INCREMENT
C THE CHAIN.
C
READ (5, 110) NRES, NSTEP
DO 2050 L = 1, 3
    DO 2050 K = 1, 3
2050   TEMPl(K, L) = TAVE(K, L)
    DO 140 K = 1, 3
140   TPOW(K, L) = TAVE(K, L)
    DO 1 I = NSTEP, NRES, NSTEP
        XRES = FLOAT(I)
        WRITE (6, 323) Z
DO 2048 K = 1,3
DO 2049 L = 1,3
2040 TAVE(K,L) = TEMP6(K,L)
IRES = NSTEP
IF(I-NSTEP .EQ. 0 ) IRES = NSTEP - 1
DO 150 J = 1, IRES
CALL MATMUL(TAVE, TEMP7, 3, TEMP7)
DO 4020 K = 1,3
DO 4020 L = 1,3
TPOW(K,L) = TEMP7(K,L)
4020 TEMP7(K,L) = 0.0
150 CONTINUE
DO 160 K = 1,3
DO 160 L = 1,3
TEMP1(K,L) = -TAVE(K,L)
TEMP2(K,L) = -TPOW(K,L)
IF(K .EQ. L ) TEMP1(K,L) = 1.0 - TAVE(K,L)
IF(K .EQ. L ) TEMP2(K,L) = 1.0 - TPOW(K,L)
160 CONTINUE
CALL MATMUL(TEMP1, TEMP1, 3, TEMP3)
CALL MINV(TEMP3, 3, J)
IF(J .NE. 0,0 ) GO TO 550
WRITE(6,111) D
550 CALL MATMUL(TAVE, TEMP2, 3, TEMP4)
CALL MATMUL(TEMP4, TEMP3, 3, TEMP5)
DO 170 K = 1,3
DO 170 L = 1,3
170 TEMP5(K,L) = (2./XRES)*TEMP5(K,L)
DO 15 K=1,3
DO 15 L=1,3
TPLUS(K,L) = TAVE(K,L)
IF (L .EQ. K) TPLUS(K,L) = TAVE(K,L) + 1.0
TAVE(K,L) = - TAVE(K,L)
15 IF (L .EQ. K) TAVE(K,L) = TAVE(K,L) + 1.0
CALL MINV(TAVE, 3, D)
IF(J .NE. 0,0 ) GO TO 550
WRITE(6,112) D
550 CALL MATMUL(TPLUS, TAVE, 3, TF1N)
IF (I .NE. NSTEP) GO TO 3
WRITE (6,337)
WRITE (6,338) ((TFIN(K,L), K=1,3),L=1,3)
3 WRITE (6,339) I
C
C
CHARACTERISTIC RATIO = TF1N(1,1)
C
TF1N(1,1) = TF1N(1,1) - TEMP5(1,1)
WRITE (6,336) TF1N(1,1)
1 CONTINUE
RETURN
END
C
C
SUBROUTINE MATMUL(A,B,L,C)
C
DIMENSION A(3,3),B(3,3),C(3,3)
DO 1 I=1,3
DO 1 J=1,L
C(I,J)=0.0
DO 1 K=1,3
1 C(I,J)=C(I,J)+A(I,K)*B(K,J)
RETURN
END
C
SUBROUTINE MINV

PURPOSE
INVERT A MATRIX

USAGE
CALL MINV(A,N,L)

DESCRIPTION OF PARAMETERS
A - INPUT MATRIX, DESTROYED IN COMPUTATION AND REPLACED BY
RESULTANT INVERSE.
N - ORDER OF MATRIX A
L - RESULTANT DETERMINANT
M - WORK VECTOR OF LENGTH N

REMARKS
MATRIX A MUST BE A GENERAL MATRIX

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
NONE

METHOD
THE STANDARD GAUSS-JORDAN METHOD IS USED. THE DETERMINANT
IS ALSO CALCULATED. A DETERMINANT OF ZERO INDICATES THAT
THE MATRIX IS SINGULAR.

SUBROUTINE MINV(A,N,L)
DIMENSION A(9),L(3),M(3)

DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
STATEMENT WHICH FOLLOWS.

DOUBLE PRECISION A,N,L,BIGA,HOLD
THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
ROUTINE.
THE DOUBLE PRECISION VERSION OF THIS ROUTINE MUST ALSO
CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. ABS IN STATEMENT
10 MUST BE CHANGED TO DABS.

SEARCH FOR LARGEST ELEMENT

D=1.0
NK=N
DO 30 K=1,N
NK=NK+N
L(K)=K
M(K)=K
KK=NK+K
BIGA=A(KK)
DO 20 J=K,N
IZ=N*(J-1)
DO 20 I=K,N
IJ=IZ+I
10 IF( ABS(BIGA) .GT. ABS(A(IJ))) 15,20,20
15 BIGA=A(IJ)
L(K) = I
H(K) = J
20 CONTINUE

C INTERCHANGE ROWS
C
J = L(K)
IF (J = K) 35, 35, 25
25 KI = K - N
DO 30 I = 1, N
KI = KI + N
HOLD = A(KI)
JI = KI - K + J
A(KI) = A(JI)
30 A(JI) = HOLD

C INTERCHANGE COLUMNS
C
35 I = M(K)
IF (I = K) 45, 45, 38
38 JP = N * (I - 1)
DO 40 J = 1, N
JK = NK + J
JI = JP + J
HOLD = A(JK)
A(JK) = A(JI)
40 A(JI) = HOLD

C DIVIDE COLUMN BY MINUS PIVOT (VALUE OF PIVOT ELEMENT IS
C CONTAINED IN BIGA)
C
45 IF (BIGA) 48, 48, 48
46 D = 0.0
RETURN
48 DO 55 I = 1, N
IF (I = K) 50, 50, 50
50 IK = NK + I
A(IK) = A(IK) / (-BIGA)
55 CONTINUE

C REDUCE MATRIX
C
DO 65 I = 1, N
IK = NK + I
HOLD = A(IK)
IJ = I - N
DO 65 J = 1, N
IJ = IJ + N
IF (I = K) 60, 60, 60
60 IF (J = K) 62, 62, 62
62 KJ = IJ - I + K
A(IJ) = HOLD * A(KJ) + A(IJ)
65 CONTINUE

C DIVIDE ROW BY PIVOT
C
KJ = K - N
DO 75 J = 1, N
KJ = KJ + N
IF (J = K) 70, 70, 70
70 A(KJ) = A(KJ) / BIGA
75 CONTINUE

C PRODUCT OF PIVOTS
C
D = J * BIGA
REPLACE PIVOT BY RECIPROCAL

A(KK) = 1.0/BIGA
80 CONTINUE

FINAL ROW AND COLUMN INTERCHANGE

K = N
160 K = (K - 1)
165 I = L(K)
170 J = N * (I - 1)
175 DO 110 J = 1, N
180 JR = J + J
185 HOLD = A(JK)
190 J = JR + J
195 A(JK) = -A(JI)
200 CONTINUE

IF(J - K) 100, 100, 125
210 J = M(K)
215 IF(J - K) 100, 100, 125
220 KI = K - N
225 DO 130 I = 1, N
230 KI = KI - K + J
235 HOLD = A(KI)
240 J = J - K + J
245 A(KI) = -A(JI)
250 CONTINUE

150 RETURN
END

***********************************************************************

SUBROUTINE COORD(TA, TB, TG)

FOR AN EXPLANATION OF THIS METHOD OF TRANSLATING PEPTIDE COORDINATES
SYSTEMS SEE REFERENCE:-

DIMENSION TA(3,3), TB(3,3), TG(3,3)
DATA ALPHA/14.94/, BETA/70.7/, GAMMA/-29.94/
10 FORMAT(1H0, 9H ALPHA =, F10.2, 8H BETA =, F10.2, 9H GAMMA =, F10.2)
12 FORMAT(1H0, 49H THE PEPTIDE GEOMETRY IS DETERMINED BY THE ANGLES)

WRITE(6, 12) ALPHA, BETA, GAMMA
WRITE(6, 10) ALPHA, BETA, GAMMA

THIS SUBROUTINE CALCULATES THE CONSTANT ROTATIONAL MATRICES TA, TB, TG.

DO 1 J = 1, 3
DO 1 K = 1, 3
TA(J,K) = 0.0
TB(J,K) = 0.0
TG(J,K) = 0.0
1 CONTINUE

NOTE. ALPHA, BETA AND GAMMA ARE FIXED VALUE ANGLES FOR EACH
PEPTIDE GEOMETRY.

ALPHA = ALPHA/57.2958
BETA = BETA/57.2958
GAMMA = GAMMA/57.2958
C GENERATE TA WHICH ROTATES THE NEW X(I) AXIS INTO THE X(I-1) AXIS.
C
TA(1,1) = COS(ALPHA)
TA(1,2) = SIN(ALPHA)
TA(2,1) = -SIN(ALPHA)
TA(2,2) = COS(ALPHA)
TA(3,3) = 1.0
C
C GENERATE TB WHICH ROTATES THE NEW X(I) AXIS INTO THE N - C ALPHA BOND.
C
TB(1,1) = COS(BETA)
TB(1,2) = SIN(BETA)
TB(2,1) = -SIN(BETA)
TB(2,2) = COS(BETA)
TB(3,3) = 1.0
C
C GENERATE TG WHICH ROTATES THE NEW X(I) AXIS INTO THE C ALPHA - C' BOND.
C
TG(1,1) = COS(GAMMA)
TG(1,2) = SIN(GAMMA)
TG(2,1) = -SIN(GAMMA)
TG(2,2) = COS(GAMMA)
TG(3,3) = 1.0
C
RETURN
END
SUBROUTINE CRCopol
C
THIS SUBROUTINE CALCULATES THE CHARACTERISTIC RATIO OF THE COPOLYMER.
C
IT DOES THIS BY FIRST CALCULATING THE COMPOSITE MATRICES GM FOR EACH
C
MONOMER FROM THE RESPECTIVE TAVE AND FROM V3OND - THE VIRTUAL
C
BOND VECTOR OF THE AMINO ACID RESIDUE.
C
COMMON TITLE(20,6)
DIMENSION NSTEP(4)
COMMON TAVE1(3,3),TAVE2(3,3),TAVE3(3,3)
DIMENSION GM(5,5),GM1(5,5),GM2(5,5),GM3(5,5),TEMP2(5)
DIMENSION TEMP5(5,5),TEMP6(5,5)
DIMENSION V3OND(3),UVECT(5),LVECT(5),TEMP3(5)
DATA V3OND/3.001,0.0,0,0,0/
DATA GM/1.0,501,0,0,0,0,0,1.0/
DATA UVect/1.0,0,4*0.0,0,0,1.0/
DATA LVECT/3.0,0.801,2*0.0,1.0/
 10 FORMAT(19I5)
 12 FORMAT(25X,5F10.2,/,25X,5F10.2,/,25X,5F10.2)
 14 FORMAT(1H0,45H NUMBER OF AMINO ACID RESIDUES IN THE CHAIN =,1X,I)
 16 FORMAT(1H0,23H CHARACTERISTIC RATIO =,1X,F19.3)
 17 FORMAT(1H0,23H CHARACTERISTIC RATIO =,1X,F19.3)
 18 FORMAT(1X,5F8.3)
 19 FORMAT(1X,5F8.3)

C

READ THE LENGTH OF POLYMER YOU WISH TO CONSIDER AND THE INTERMEDIATE
C
LENGTHS.  NOTE THAT NRES AND NSTEP DO NOT REFER TO THE SAME LENGTHS
C
AS IN SUBROUTINE CRMnO.  IN THIS CASE NRES REFERS TO THE LOG TO
C
THE BASE OF TWO OF THE NUMBER OF RESIDUES IN THE CHAIN, AND NSTEP
C
TO THE FOUR INCREMENTS OF NRES FOR WHICH YOU WISH THE C.R. OF
C
THE COPOLYMER TO BE PRINTED.
C
C
READ(5,10) NRES,(NSTEP(I),I=1,4)
C
DO 1 I=1,5
DO 1 J=1,5
C INITIALISE GM1-GM3

C
G'M1(I,J) = GM(I,J)
G'M2(I,J) = GM(I,J)
G'M3(I,J) = GM(I,J)
1 GM(I,J) = 0.0
C CONSTRUCT GM1-GM3
C
DO 2 I = 2, 4
DO 2 J = 2, 4
II = I - 1
JJ = J - 1
GM1(I,J) = TAVE1(II, JJ)
GM2(I,J) = TAVE2(II, JJ)
GM3(I,J) = TAVE3(II, JJ)
2 CONTINUE
C
CALL MATPROD(VBOND, TAVE1, 1, 3, 3, TEMP1)
DO 3 J = 2, 4
I = J - 1
GM1(I,J) = TEMP1(I)
3 TEMP1(I) = 0.0
CALL MATPROD(VBOND, TAVE2, 1, 3, 3, TEMP1)
DO 4 J = 2, 4
I = J - 1
GM2(I,J) = TEMP1(I)
4 TEMP1(I) = 0.0
CALL MATPROD(VBOND, TAVE3, 1, 3, 3, TEMP1)
DO 5 J = 2, 4
I = J - 1
GM3(I,J) = TEMP1(I)
5 TEMP1(I) = 0.0
C
WRITE OUT THE COMPOSITE MATRICES, GM1, GM2, GM3
C
WRITE((6,11))(TITLE(J,1), J = 1, 13)
WRITE((6,12))((GM1(I,J), J = 1, 5), I = 1, 3)
WRITE((6,11))(TITLE(J,2), J = 1, 13)
WRITE((6,12))((GM2(I,J), J = 1, 5), I = 1, 5)
WRITE((6,11))(TITLE(J,3), J = 1, 13)
WRITE((6,12))((GM3(I,J), J = 1, 5), I = 1, 5)
C
NOW THE CHARACTERISTIC RATIO OF THE COPOLYMER IS CALCULATED
C TO THE LENGTH NRES FROM GM1, GM2, AND GM3.
C
CALL MATPROD(GM1, LVEC1, 5, 5, 1, TEMP3)
C
CONSTRUCT THE MATRIX PRODUCT GM1*GM2*GM3
C
CALL MATPROD(GM1, GM2, 5, 5, 5, GM)
CALL MATPROD(GM, GM3, 5, 5, TEMP5)
DO 70 I = 1, 5
DO 70 J = 1, 5
70 GM(I,J) = TEMP5(I,J)
C
SUCCESIVELY SQUARE THE MATRIX PRODUCT
C
DO 60 II = 1, NRES
CALL MATPROD(GM, TEMP5, 5, 5, 5, TEMP6)
DO 80 I = 1, 5
DO 80 J = 1, 5
GM(I,J) = TEMP6(I,J)
80 TEMP5(I,J) = TEMP6(I,J)
DO 90 I = 1, 4
90 IF((II.EQ.NSTEP(I)).OR.(II.EQ.NRES)) GO TO 65
GO TO 60
CONTINUE

CALCULATE THE FIRST ELEMENT OF THE MATRIX PRODUCT

CALL MATPROD(GM,TEMP3,5,5,1,TEMP2)
CALL MATPROD(UVECT,TEMP2+1,5,1,G11)

WRITE(6,17) G11

CALCULATE THE NUMBER OF TRIPLETTINES IN THE CHAIN FROM NSTEP

NP=3*(2**(II+1))-1

THE FINAL EXPRESSION FOR THE CHARACTERISTIC RATIO.

GUOND=GUOND(1)
CR=1.9*(2.0/(NP*GUOND**2))*G11

WRITE(6,14) NP
WRITE(6,16) CR

CONTINUE
STOP
END

SUBROUTINE MATPROD(A,B,N,M,L,C)

THIS SUBROUTINE MULTIPLIES NON-SQUARE MATRICES OF SPECIFIED SIZES
NOTE THAT THE OUTPUT MATRIX CAN NOT SHARE THE LOCATION
OF EITHER OF THE INPUT MATRICES

A LEFT HAND INPUT MATRIX
B RIGHT HAND INPUT MATRIX
C OUTPUT MATRIX = A*B
N = NUMBER OF ROWS IN A AND C
M = NUMBER OF COLUMNS IN A AND ROWS IN B
L = NUMBER OF COLUMNS IN B AND C

DIMENSION A(N,M),B(M,L),C(N,L)

DO 1 I=1,N
    DO 1 J=1,L
        C(I,J)=0.0
    1   DO 1 K=1,M

        C(I,J)=C(I,J)+A(I,K)*B(K,J)

RETURN
END
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Paterson, Yvonne

Title:
The effect of side chain complexity on the dimensions of polypeptide chains

Date:
1978

Citation:

Publication Status:
Unpublished

Persistent Link:
http://hdl.handle.net/11343/36822

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only, download, print, and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.