ADVANCES IN PROTEIN CHEMISTRY

EDITED BY

C. B. ANFINSEN, JR.
Laboratory of Cellular Physiology
National Heart Institute
Bethesda, Maryland

M. L. ANSON
London, England

KENNETH BAILEY
University of Cambridge
Cambridge, England

JOHN T. EDSALL
Biological Laboratories
Harvard University
Cambridge, Massachusetts

VOLUME 16

1961

ACADEMIC PRESS • *New York and London*

ACADEMIC PRESS INC.
111 FIFTH AVENUE
NEW YORK 3, N. Y.

United Kingdom Edition
Published by
ACADEMIC PRESS INC. (LONDON) LTD.
BERKELEY SQUARE HOUSE
BERKELEY SQUARE, LONDON, W. 1

Library of Congress Catalog Card Number 44-8853

PRINTED IN THE UNITED STATES OF AMERICA

CONTRIBUTORS TO VOLUME 16

P. DESNUELLE, *Laboratoire de Chimie, Biologique, Faculté des Sciences, Marseille, France*

MALCOLM DIXON, *Department of Biochemistry, University of Cambridge, England*

PAUL DOTY, *Department of Chemistry, Harvard University, Cambridge, Massachusetts*

WILLIAM F. HARRINGTON, *McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, Maryland*

M. ROVERY, *Laboratoire de Chimie Biologique, Faculté des Sciences, Marseille, France*

PETER URNES, *Department of Chemistry, Harvard University, Cambridge, Massachusetts*

PETER H. VON HIPPEL, *Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire*

EDWIN C. WEBB, *Department of Biochemistry, University of Cambridge, England*

B. WITKOP, *Laboratory of Chemistry, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland*

JEN TSI YANG, *Cardiovascular Research Institute and Department of Biochemistry, University of California Medical Center, San Francisco, California*

CONTENTS

The Viscosity of Macromolecules in Relation to Molecular Conformation

JEN TSI YANG

Optical Rotation and the Conformation of Polypeptides and Proteins

PETER URNES AND PAUL DOTY

THE STRUCTURE OF COLLAGEN AND GELATIN

WILLIAM F. HARRINGTON

McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, Maryland

and

PETER H. VON HIPPEL

Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire

I. Introduction

The properties of collagen are of interest from many points of view. Since it constitutes the major protein component of skin, bone, tendon, and all the other forms of connective tissue, an understanding of collagen seems to the clinician to be a necessary (though not sufficient) prerequisite to a rational attack on the many and diverse connective tissue disorders currently lumped together as "collagen diseases." The mechanisms of the biosynthesis and incorporation into collagen of the unusual amino acids hydroxylysine and hydroxyproline offer a continuing challenge to the "pathway" biochemist. The leather chemist strives for a better understanding of the interaction of collagen with tanning agents, for such understanding lies directly on the road to improved leather products. The

comparative biologist is intrigued by the direct correlation between the temperature at which the collagen of a given organism undergoes thermal "denaturation" and the temperature of the normal habitat of that organism. And finally, collagen interests the physical biochemist, not only because of its intrinsic importance, but also because the properties of collagen on the molecular level truly belong in a class set apart from those of all other protein molecules.

Clearly, it is impossible to consider, even briefly, all these facets of the collagen problem. So while recognizing both the importance and the intrinsic interest of other areas, limits of space, patience, and competence all have served to confine this review primarily to the last of the points of view mentioned above. And even this area—the molecular properties of collagen—is so large and developing so rapidly that no pretence to completeness of coverage can be made. Thus it will be apparent that this review represents an attempt to summarize the present position as seen by the authors, for although we have tried to take a balanced view throughout, some aspects of certain of the problems discussed are controversial and might have been handled quite differently by others in terms of emphasis, coverage, and perhaps interpretation. However, despite these difficulties, we hope that we have succeeded in assembling a reasonably complete and coherent account of the current situation.

Since "The Structure of Collagen Fibrils" was last reviewed in these pages by Bear (1952), tremendous progress has been made in the field of protein chemistry and structure analysis. And this progress has naturally been accompanied by comparable progress in our understanding of the collagen molecule. The magnitude of this development in the field of collagen might best be indicated by summarizing some of the major developments of the decade which has passed since Bear wrote his review in the spring of 1951.

1. It had been clear for many years prior to 1951 that collagen fibers from all organisms have in common a distinctive wide-angle X-ray diffraction pattern, and thus presumably also a unique polypeptide chain configuration. However the nature of this configuration was not known. Since that time advances in our understanding of the stereochemistry of polypeptide chains and the theory of fiber X-ray diffraction first made possible, and then led to, a generally accepted structural interpretation of the wide-angle X-ray pattern of collagen.

2. The unusual amino acid composition of collagen had also been recognized for some time. One-third of the residues of all collagens seemed to be glycine, while about one-fourth were proline and hydroxyproline. However, the stereochemical consequences of the presence of these residues has only become clear as a result of detailed studies of synthetic homo- and

copolymers of glycine and proline. Consideration of such synthetic polypeptides as simplified models of certain features of collagen and gelatin has been extremely helpful in recent years, and constitutes the rationale for the inclusion of a section dealing specifically with these synthetic polypeptides in this review.

3. Physicochemical studies of soluble collagen, which had just been initiated in 1951, have been pursued vigorously during the last 10 years and have added immeasurably to our understanding of the collagen system. Such studies, in turn, have made it possible to interpret the long-range periodicities which had previously been observed in collagen fibers using electron microscopy and small-angle X-ray diffraction, and have led to a partial understanding of the molecular basis of the patterns observed in the various "reconstituted" collagens.

4. The collagen $\rightleftharpoons$ gelatin transformation in solution has been recognized as a reversible first-order phase transition, subject to the same physical laws which govern the crystalline $\rightleftharpoons$ amorphous phase transitions observed in systems of linear polymers. The direct relationship between the transition in solution and the well-known thermal shrinkage phenomenon exhibited by collagen fibers has also been established.

5. It has become clear that gelatin, long considered to be a degradation product of uncertain composition and properties, can be prepared from collagen under conditions which result in the production of perfectly reproducible protein systems. Thus gelatin itself, and the mechanism of the process by which it undergoes spontaneous reversion to a collagen-like structure on cooling, have been the object of considerable study over the last few years.

Specifically, it is the recent developments in these areas which we propose to discuss in this review. For detailed considerations of earlier work and for coverage of certain peripheral topics the reader is referred to the reviews by Bear (1952) and Kendrew (1954), to the informative monographs by K. H. Gustavson entitled "The Chemistry and Reactivity of Collagen," and "The Chemistry of Tanning Processes" (Academic Press, 1956); and to numerous recent symposia dealing with various aspects of the collagen problem, including: "Nature and Structure of Collagen" (J. T. Randall, ed., Academic Press, 1953a); "Fibrous Proteins and Their Biological Significance" (*Soc. Exptl. Biol.* **9**, 127 (1955); "Connective Tissue" (R. E. Tunbridge, ed., Blackwell, 1957); "Recent Advances in Gelatin and Glue Research" (G. Stainsby, ed., Pergamon Press, 1958); "Calcification in Biological Systems" (R. F. Sognnaes, ed., *Am. Assoc. Advancement Sci.*, 1960); and "Central Leather Research Institute Symposium on Collagen" (N. Ramanathan, ed., Interscience Publishers, 1961).

Before plunging into the detailed discussion, it is important to consider

the problem of the identification of collagen and its distribution through the "animal kingdom," and simultaneously to establish an operational definition of this class of proteins. These problems have been treated at great length by Bear (1952), and since recent work has only served to substantiate most of his conclusions we will confine ourselves to a brief summary of current views.

Collagen fibers may often be recognized histologically on the basis of one or more of the following bulk characteristics: they tend to swell markedly when immersed in acid, alkali, or concentrated solutions of certain neutral salts and nonelectrolytes; they are generally relatively inelastic; they are more resistant than most protein fibers to degradation by proteolytic enzymes, but (in contrast to all other proteins studied) are readily attacked by the enzyme collagenase; they undergo thermal shrinkage to a fraction of their original length at a temperature which is characteristic of the collagen from a given animal (but varies from one species to another); and they are converted in large part to soluble gelatin by prolonged treatment at temperatures above the thermal shrinkage level.

In practice, however, these criteria are not always easy to apply, nor are they infallible. Thus in recent years the presence of the characteristic collagen wide-angle X-ray diffraction pattern has come to be accepted as the fundamental defining criterion for collagen. This pattern, (which will be discussed and interpreted in detail in Section IV) is easily recognized by the strong 2.86 A meridional spacing and by the ~11 A hydration-sensitive reflection on the equator (see Fig. 8). It has served to demonstrate the presence of collagen in the tissues of almost all multicellular animals which have been investigated, ranging from the primitive porifera and coelenterates, through the annelids and echinoderms, and up to the vertebrates (see Marks *et al.*, 1949; Bear, 1952; Gross *et al.*, 1956).

The following physical and chemical features also seem to be generally characteristic of collagen (see Gross *et al.*, 1958; Watson and Silvester, 1959; Piez and Gross, 1959; Piez and Likins, 1960): (1) a content of glycyl residues close to one-third of the total number present; (2) a high content of pyrrolidine-ring-containing residues (relative to the composition of most other proteins); (3) substantial numbers of hydroxyprolyl residues; (4) few or no cystyl, methionyl, valyl, phenylalanyl, tyrosyl, or histidyl residues, and thus minimal absorption at 280 mμ; (5) hydroxylysyl residues; (6) an extensive meridional small-angle X-ray diffraction pattern; (7) fibrils exhibiting periodically-banded (~640 A) structure in the electron microscope; and (8) an infrared absorption band at 3330 cm^{-1}.

In the following sections these criteria will be considered in detail, mostly in terms of certain vertebrate collagens which have been extensively studied. However, the available evidence suggests that in main outline the

results obtained with these collagens apply equally well to most vertebrate (and perhaps invertebrate) collagens. Very recently careful studies of invertebrate collagens have been launched by several groups, and it has become apparent that some of these collagens differ substantially from vertebrate collagen in certain respects. For example, the ~640 A macro-period seen via electron microscopy or small-angle X-ray diffraction in vertebrate fibers has not been detected in certain of the invertebrate materials examined. Also, while in vertebrate collagens the ratio of proline to hydroxyproline is generally in the neighborhood of unity, in the invertebrates this ratio varies widely. Thus gelatin from earthworm (*Lumbricus*) cuticle contains 13 prolyl and 165 hydroxyprolyl residues per 1000, while gelatin from *Ascaris* cuticle contains 280 prolyl and only 24 hydroxyprolyl residues per 1000 (Watson and Silvester, 1959). It is expected that careful physicochemical studies of such invertebrate collagens will help to illuminate further the role of these residues in the collagen structure.

Finally, it must be mentioned that there exist certain classes of proteins which yield the collagen wide-angle X-ray diffraction pattern and thus meet the minimum requirements for membership in the collagen group, but which differ significantly from the better known collagens in other ways. Despite previous controversy, all the following now seem securely established as collagens.

(*a*) Reticulin: This substance is found closely associated with collagen in the connective tissue. It is identified histologically through the presence of branching networks (see Bear, 1952; Kramer and Little, 1953; Kendrew, 1954; Robb-Smith, 1957, 1958).

(*b*) Ichthyocol and elastoidin: These terms simply refer to collagens prepared, respectively, from the swim bladders and skins of fish, and from the fins of the shark. Certain of the former, particularly the collagen derived from the swim bladder of the carp, have been extensively studied and are discussed in detail in subsequent sections. Elastoidin has been examined much less extensively (see particularly Bear, 1952; Gross and Dumsha, 1958).

(*c*) Vitrosin: A fibrous protein obtained from the vitreous humor of the eye. It exhibits the collagen wide-angle X-ray pattern, ~640 A periodicity and a typical collagen amino acid composition (see Gross *et al.*, 1955b).

(*d*) Spongin, gorgonin, and cornein: Fibrous proteins derived from sponges, corals, and coelenterates (see especially Marks *et al.*, 1949; Bear, 1952; Gross *et al.*, 1956; Piez and Gross, 1959).

(*e*) The "secreted" collagens: These may be differentiated from other collagens in being secreted by epithelial cells instead of being mesodermal or mesogleal in origin. The secreted collagens show the typical wide-angle X-ray pattern, but apparently no macro-period. Examples include: earth-

worm and *Ascaris* cuticle, the bivalve byssus threads, and the "ejected filaments" of the sea cucumber (see Bear, 1952; Kendrew, 1954; Watson and Silvester, 1959).

Despite earlier claims to the contrary, it now seems relatively well established that elastin, the third major histological component of vertebrate connective tissue (with collagen and reticulin) is not a member of the collagen class of proteins (Bear, 1952; Kendrew, 1954).

II. Synthetic Polypeptides Related to Collagen

In this section we propose to consider the effects of proline, hydroxyproline, and glycine residues on the molecular architecture of polypeptide chains. The physical chemistry of synthetic polypeptides containing these residues will be examined in detail because of the remarkable similarities which exist between the helical structures generated from polymers of these amino acids and the three-dimensional pattern of the polypeptide chains of the collagen molecule. Moreover, the mechanism of the reversion of the component chains of collagen to form the parent collagen molecule at low temperatures *in vitro* (and possibly in the connective tissue *in vivo*) is now thought to be intimately related to the configurational changes exhibited by these model substances in solution.

Pauling (1940), in considering the general problem of protein configuration, first pointed out that the insertion of a proline residue into a polypeptide chain has marked effects on the chain direction. Indeed, he demonstrated that proline residues, by virtue of their unusual stereochemistry, are ideally suited to serve as hinge points for chain folding in globular proteins. More recently, it has also become apparent that the geometry of the pyrrolidine ring prevents L-proline from being accommodated in a left-handed α-helix. Furthermore, this residue can only fit into an undistorted right-handed α-helix when it occurs as one of the first three residues at the *N*-terminal end of the chain. On the other hand, Lindley (1955) has shown that a change in helical sense (from a left-handed to a right-handed α-helix) can occur without distortion at a proline residue. In all cases the direction of an α-helix is altered at a proline residue and may be completely inverted if the atomic grouping about the imide linkage is disposed in the *cis*-configuration (Edsall, 1954).

In addition to the stereochemical features imposed on a polypeptide chain as a consequence of the insertion of a proline residue, the imino nitrogen atom of the pyrrolidine ring is devoid of a proton and is consequently unable to participate in the formation of inter- or intrachain hydrogen bonds. This means that the regularity of any systematic, hydrogen-bonded peptide structure will be interrupted in proline-containing regions. The effect of such internal breaks on the stability of the α-helix has been considered by Schellman (1955) who demonstrated that usu-

ally at least three hydrogen bonds would be ruptured at such a discontinuity, although little configurational entropy should be acquired by the chain since peptide units on either side of the break are maintained by hydrogen bonds. On the other hand it would be expected that insertion of an appreciable number of randomly distributed proline residues would lead to destabilization of the helix. Some evidence which supports this view comes from the work of Szent-Gyorgyi and Cohen (1957) who compared the α-helical content of a number of the keratin-myosin-epidermin-fibrinogen (KMEF) class proteins, as deduced from optical rotatory dispersion, with their proline content. They found that proteins with less than 30 proline residues per 1000 were in general highly folded, and that as the proline content increased above this level a polypeptide chain without benefit of stabilizing cross-bridges seemed to approach a random-coil-type configuration. At very high concentrations of proline a completely new type of configurational pattern seemed to emerge. This configuration, termed poly-L-proline II, will be the subject of detailed consideration below.

Before embarking upon the discussion of the relatively complex architecture of the synthetic polymers and copolymers related to collagen and gelatin, it is important to examine the structure of glycine, proline, hydroxyproline, and simple peptides composed of these residues. As indicated in the introduction, the stereochemistry of these residues is crucial to all current theories of collagen structure. Their importance may be judged, without reference to a particular structure, when it is recognized that, together, they comprise over 50 % of the amino acid content of most collagens.

A. Amino Acid Structures

1. *Glycine*

The crystal structure of glycine has been redetermined recently by Marsh (1958) from a complete three-dimensional Fourier analysis involving 1867 reflections. In general, the bond lengths and bond angles deduced were essentially identical to those reported earlier by Albrecht and Corey (1939) except that the C—N bond distance is 1.474 A instead of 1.39 A. The former distance is in good agreement with that normally observed in other amino acids. Carbon atoms 1 and 2 and oxygen atoms 1 and 2 (C_1, C_2, O_1, and O_2)* are nearly coplanar (see Fig. 1) whereas the nitrogen atom is 0.44 A out of plane.

2. *L-Proline*

A complete X-ray diffraction study of this imino acid has been prevented as a result of its strongly hygroscopic character. However, Mathieson and

* The subscript numbers with symbols here indicate position of atom in the molecule.

Welsh (1952) have deduced structure parameters from a careful investigation of copper-DL-proline dihydrate. In Fig. 1, atoms C_1 and N, C_2, C_3, and C_5 of the pyrrolidine ring are coplanar. C_4 lies at a distance of about 0.60 A out of plane while the carboxyl group attached to atom C_2 is disposed on the opposite side of the plane with respect to C_4. Bond angles N—C_5—C_4

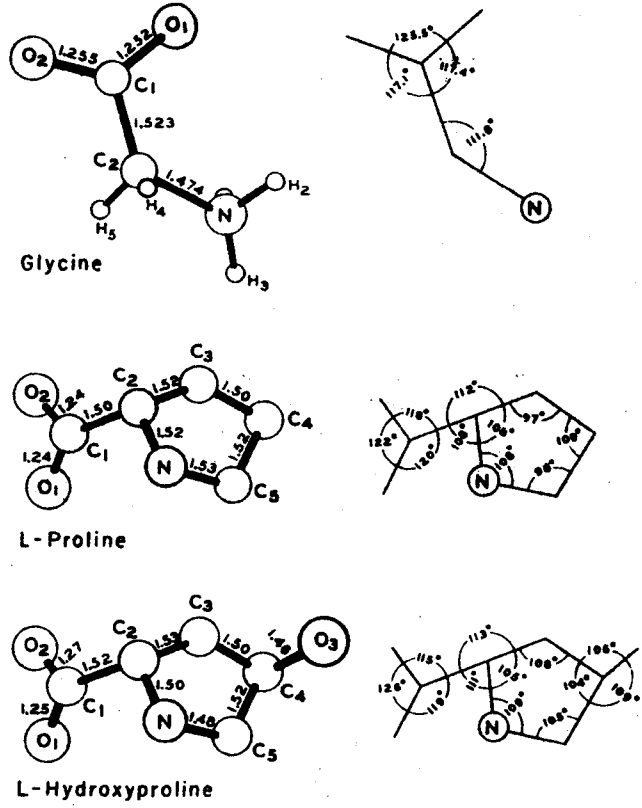

FIG. 1. Interatomic bond lengths and bond angles for glycine (Marsh, 1958) L-proline (Mathieson and Welsh, 1952), and L-hydroxyproline (Donahue and Trueblood, 1952).

and C_2—C_5—C_4 have the unusually low values of 96° and 97°, respectively. These values may be in error since in L-hydroxyproline (Donahue and Trueblood, 1952), in L-leucyl-L-prolylglycine (Leung and Marsh, 1957), and in poly-L-proline II (Cowan and McGavin, 1955; Sasisekharan, 1959a) the corresponding angles vary between 103° and 108°.

3. *L-Hydroxyproline*

The crystal structure of L-hydroxyproline was first reported by Zussman (1951) and a complete three-dimensional Fourier analysis published a year

later by Donahue and Trueblood (1952). As in copper-DL-proline dihydrate, the five-membered ring is appreciably puckered. C_4, the carbon atom bearing the hydroxyl group, is about 0.40 A from a plane defined by the other four atoms of the ring and lies on the opposite side of this plane from the carboxyl group. The angle between the two planes C_5—N—C_2—C_3 and C_3—C_4—C_5 is 17°. Neuberger's (1948) prediction that the carboxyl and hydroxyl groups should be in the *trans*-configuration relative to the ring has been completely confirmed by the structure of Donahue and Trueblood.

B. Simple Peptides of Glycine, Proline, and Hydroxyproline

1. *Glycylglycine*

Glycylglycine can exist in three different crystal forms, α, β, and γ, which were originally observed by Bernal (1931) growing side-by-side in the same mother liquor. The dimensions of this simplest of all linear peptides were first determined by Hughes and Moore (1949) from a two-dimensional Patterson projection analysis of the β-form. In this structure the heavy atoms are all coplanar except for the terminal nitrogen which is 0.64 A out of the molecular plane. More recently Hughes *et al.* (1954) have carried out a three-dimensional analysis of α-glycylglycine involving more than 2000 X-ray reflections. Again the whole molecule is planar except for the terminal nitrogen atom which lies 0.73 A below the plane. In this structure the peptide bond length is 1.32 A, which is closer to the value usually observed for this bond (Corey and Pauling, 1953) than is the 1.29 A distance reported for β-glycylglycine. Both forms of the structure are shown in Fig. 2.

2. *L-Leucyl-L-prolylglycine*

A study of the stereochemistry of this peptide is of great value in gaining an understanding of the configurational effect of a proline residue on the relative arrangement of neighboring amino acid residues. The crystal structure was determined by Leung and Marsh (1957) from a thorough three-dimensional diffraction analysis utilizing about 1700 reflections. Atoms C_4, C_7, C_8, C_9, N_2, and O_4 of the leucyl-prolyl peptide group are coplanar within 0.02 A, with bond lengths and angles close to those reported for other peptides (Corey and Pauling, 1953) except for the bond angle N_2—C_8—C_9, which is about 5° larger than expected (see Fig. 3). Dimensions of the pyrrolidine ring are closely similar to those observed for L-hydroxy-proline (Donahue and Trueblood, 1952) with atoms C_4, C_5, C_7, and N_2 coplanar, and C_6 out of plane by 0.37 A. The dihedral angle between the plane of the leucyl-prolyl amide grouping and that of the proline ring is approximately 7°.

Several noteworthy features emerge from an examination of the structure

which are of special interest to the present review. Most prominent among these is the finding that the

Bond (*i*)

bond length is close to 1.34 A. This value is normal for a peptide bond, demonstrating that the imide linkage between the carbonyl group (C_8) and the imino nitrogen (N_2) of the pyrrolidine ring possesses virtually the same degree of double bond character as do other peptide linkages. Moreover the

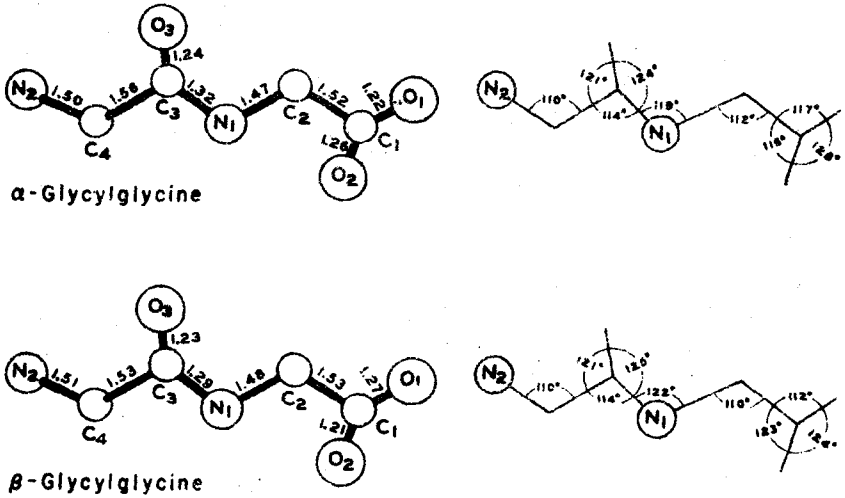

FIG. 2. Interatomic bond lengths and bond angles for α-glycylglycine (Hughes and Moore, 1949) and β-glycylglycine (Hughes, *et al.*, 1954).

angle C_4—N_2—C_7 of 113° is close to the value accepted by Corey and Pauling (1953) for the C_α—NH angle in a peptide group. Another prominent feature is the disposition of the major groups. It appears that the leucyl-prolyl-glycine residues are maximally extended in the crystal structure, with the nine main-chain atoms of the proline and glycine residues (C_1–C_4, N_1, and O_1—O_3) essentially coplanar. Each C_α—N bond is virtually *cis* to the C=O bond of the same residue. At the position of the proline residue the chain undergoes a twist of about 120° from the fully extended configuration, a twist required by the geometry of the pyrrolidine ring.

3. *Tosyl-L-prolyl-L-hydroxyproline Monohydrate*

Only a partial three-dimensional analysis of this peptide has been reported (Beecham *et al.*, 1958) although the known occurrence of proline-

hydroxyproline sequence in the chains of collagen (see Section III) endow it with special significance. The molecule appears to be composed of four approximately planar groups: *p*-methyl-thiophenyl, prolyl, peptide + pyrrolidine, and carbonyl. These are disposed at right angles to each other. Evidence of the flexibility of the pyrrolidine ring comes from the position of the carbon atom opposite the C_α—N bond of the proline residue. In this structure, unlike that of proline (Mathieson and Welsh, 1952) and hy-

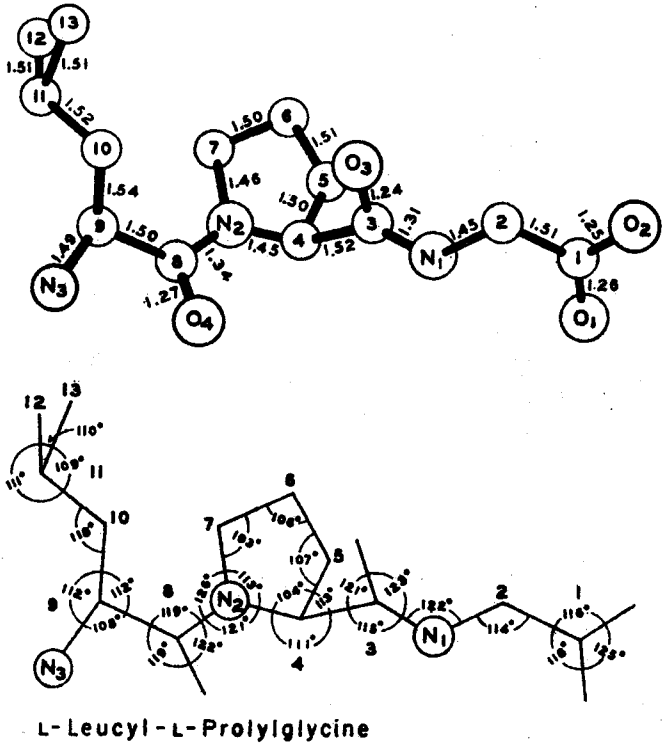

FIG. 3. Interatomic bond lengths and bond angles for L-leucyl-L-prolylglycine (Leung and Marsh, 1957).

droxyproline (Donahue and Trueblood, 1952) this carbon atom appears to be on the same side of the ring plane as the peptide C—O. Another example of the apparent flexibility of the pyrrolidine ring will be given in the discussion of the poly-L-proline II structure.

C. Physical Chemical Properties of Synthetic Polypeptides

1. *Polyglycine II*

a. Studies on Polyglycine II in the Solid State. In 1934, Meyer and Go reported an X-ray diffraction study of various polyglycine preparations, in