INFORMATION TO USERS

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

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

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

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

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

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

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

A Xerox Education Company

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE CRYSTAL AND MOLECULAR STRUCTURES OF A COPPER(II) COMPLEX WITH L-TYROSINE AND N-ACETYL-L-PROLINE MONOHYDRATE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

CLINTON EUGENE TATSCH

Norman, Oklahoma

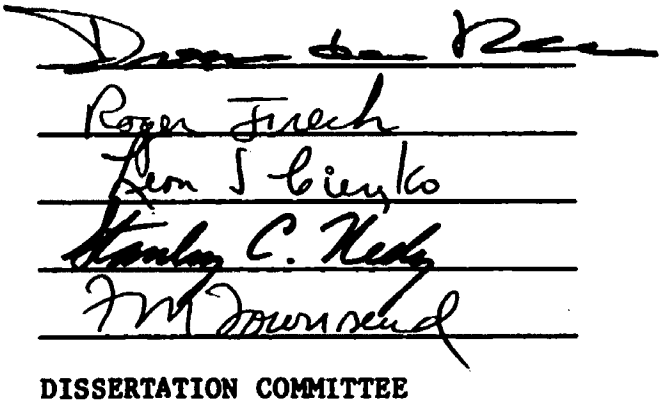
1972

THE CRYSTAL AND MOLECULAR STRUCTURES OF

A COPPER(II) COMPLEX WITH L-TYROSINE AND

N-ACETYL-L-PROLINE MONOHYDRATE

APPROVED BY



DISSERTATION COMMITTEE

To Carolyn

ACKNOWLEDGMENT

"No man is an island, entire of itself . . . "
John Donne, 1623

I welcome this opportunity of expressing my sincere appreciation for the support and conscientious guidance of Dr. Dick van der Helm during the course of these investigations. From him I have learned not only of the 'doing' of chemistry, but also of a proper attitude towards it. His insistence on a personal scientific integrity has been adamant; it has been taken to heart.

The guidance, encouragement and support of Dr. Leon Ciereszko are thankfully acknowledged. His investment may yet pan out. The many discussions with Dr. Eric Enwall have been profitable.

The perseverance and support of Carolyn - the "graduate wife" - has not gone unnoticed, nor unappreciated. Without her continued support and encouragement, this work would not have begun, let alone be completed. The constant encouragement received from our parents has been an important factor through these years.

Many hours of computing time have been expended in obtaining the results presented in the following pages. Thanks are due the personnel of the O.U. Computing Centers for their aid in this regard.

Finally, the contribution of the many individuals who have participated, directly or indirectly, in this research is acknowledged. Although too numerous to mention, without their timely aid, these studies would yet be incomplete.

Financial support by PHS grant HE 05675 and GM-10514 is thankfully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PART I

THE CRYSTAL AND MOLECULAR STRUCTURE OF A COPPER(II) COMPLEX WITH L-TYROSINE

INTRODUCTION

As part of a study of the interactions between transition metal ions and amino acids and peptides, the crystal structure of the copper(II) chelate of L-tyrosine (CUTY) has been determined. Previously it was reported (1, 2) that a possible weak interaction had been observed between chelated copper(II) ions and the phenolic portion of tyrosyl residues in the structure of the copper(II) chelate of glycyl-L-leucyl-L-tyrosine, CUGLT. This behavior could have had at least two possible causes: (1) a weak interaction between the copper ion and the aromatic π-electron system, or (2) structure-specific effects, such as packing requirements. The implications of a weak interaction, if verified, could prove useful in elucidating the mechanism of certain oxidase enzymes. In order to further examine the possibility of an interaction, the crystal structures of several related compounds are under investigation in this laboratory. The copper(II) chelate of L-phenylalanine, CUPA, has been reported (3); the structure determination of CUTY is the subject of this investigation.

EXPERIMENTAL

Single crystals of CUTY were obtained by slow aqueous diffusion of cupric acetate and L-tyrosine at room temperature. Deep blue, thin plates [plate face is (010) plane] appeared at the surface of the tyrosine after 2-3 days. Crystals thus obtained were quite stable; the solution, however, showed slow decomposition after 10-20 days. As the crystals were not appreciably soluble in common solvents, no recrystallization was accomplished.

Crystallographic data (Table 1) and integrated X-ray intensity data were collected, at room temperature, on a G.E. XRD-5 diffraction unit equipped with a SPG single crystal orienter, scintillation counter and pulse-height analyzer. The 2012 data, comprising all unique reflections with $2\theta \leq 140^{\circ}$, were collected by using Ni-filtered Cu K_{α} radiation (λ = 1.5418 Å) and θ-2θ scans which corresponds to a resolution of 0.82 Å. For 193 scans, the intensity was not visibly distinguishable from the background; these were tagged as unobserved and assigned an intensity equal to 1/5 of the background at the location of the expected reflection. Lorentz, polarization and absorption corrections (μ = 21.4 cm^{-1}) were applied to the data. For the absorption correction, the program of Coppens, Leiserowitz & Rabinovich (1965) (4) was used with 216 sampling points. This program uses the numerical integration method of Gauss.

Table 1

Crystallographic Data, CUTY

Formula: $Cu(C_9H_{10}NO_3)_2$

F.W. = 423.91

Systematic absences: h00, h = 2n + 1
0k0, k = 2n + 1
001, l = 2n + 1

Space group = $P_{2_1}2_12_1$

a = 13.049 ± 0.007 Å

b = 22.227 ± 0.008 Å

c = 6.078 ± 0.003 Å

(determined by least squares fit to the 2θ values of 28 reflections)

Z = 4

ρ_{calc} = 1.60 g·cm^{-3}

ρ_{obs} = 1.58 g·cm^{-3}

(measured by flotation, at 24° C, in a CCl_4-CH_3I mixture)

Crystal dimensions: 0.6 x 0.1 x 0.03 mm.

F(000) = 876.

STRUCTURE DETERMINATION AND REFINEMENT

A sharpened Patterson synthesis was calculated. The positions of the copper atom were determined from the Harker sections, while 11 non-hydrogen atoms were also located. These atoms were used in a structure factor calculation. The subsequent Fourier synthesis yielded the positions of the remaining non-hydrogen atoms. Block diagonal least-squares refinement of these atoms using isotropic thermal parameters converged to an $R = (\Sigma \big| |kF_o| - |F_c| \big| / \Sigma |kF_o|)$ of 0.11. All hydrogen atoms, excepting the two hydroxyl hydrogens, were located from a difference Fourier. The hydrogen atom coordinates used in the further refinement were calculated from geometrical considerations and were not refined. They were given isotropic thermal parameters, which were $\frac{1}{2}$ Å^2 larger than those of the atoms to which they are attached. All non-hydrogen atoms were given anisotropic thermal parameters. The observed structure factors were corrected for the anomalous dispersion of copper: $\Delta f' = -1.862$, $\Delta f'' = 0.604$ (5). Least-squares refinement of the non-hydrogen atoms, using all the data, was terminated when all parameter shifts were less than 0.1 of the corresponding calculated standard deviations. The final R, based on the final parameters (Tables 2 and 3), is 0.042 for all data, and 0.034 when the unobserved reflections are excluded. A final difference Fourier showed a number of peaks between -0.3 and +0.3 e.Å^{-3}.

The atomic scattering factors for Cu^{2+}, N, C and O were taken from International Tables for X-ray Crystallography (1962) (6). The scattering factors for the hydrogen atoms were those of Stewart, Davidson & Simpson (1965) (7). The quantity minimized in the least-squares refinement was

Table 2

Atomic Coordinates (x 10^4) and Thermal Parameters

The temperature factor is expressed in the form exp $\left[-(h^2b_{11}+k^2b_{22}+l^2b_{33}+hkb_{12}+hlb_{13}+klb_{23})\times10^{-4}\right]$.

Standard deviations for the last digit given in parentheses.

	x	y	z	b_{11}	b_{22}	b_{33}	b_{23}	b_{13}	b_{12}
Cu	3886.3(4)	-326.6(2)	-3358.6(9)	38.0(3)	11.4(1)	93(1)	-6.2(8)	-7(1)	5.1(4)
N(1)	3712(3)	512(1)	-2265(5)	41(3)	10(1)	69(8)	-9(4)	1(8)	-2(2)
O(1)	3425(2)	- 5(1)	-6128(5)	42(2)	11(1)	72(8)	-6(4)	-16(7)	7(2)
O(3)	2747(3)	786(1)	-7754(5)	48(2)	14(1)	104(8)	10(4)	-43(7)	-2(2)
C(1)	3107(3)	532(2)	-6108(7)	26(2)	12(1)	70(10)	-8(4)	15(8)	-9(2)
C(3)	3116(3)	870(2)	-3901(7)	28(2)	9(1)	88(10)	-9(5)	2(9)	-1(2)
C(5)	3507(4)	1513(2)	-4191(8)	70(4)	10(1)	94(10)	-6(6)	19(11)	-4(3)
C(7)	3392(4)	1897(2)	-2130(7)	45(3)	8(1)	153(14)	8(6)	35(10)	-8(3)
C(9)	2447(4)	2165(2)	-1653(9)	51(3)	14(1)	169(12)	-4(8)	-36(14)	-1(3)
C(11)	4193(4)	1995(2)	- 724(8)	39(3)	14(1)	134(12)	7(6)	13(10)	-2(3)
C(13)	2340(4)	2522(2)	184(10)	41(3)	15(1)	244(16)	-22(7)	26(13)	3(3)
C(15)	4104(4)	2356(2)	1149(8)	39(3)	11(1)	195(14)	-1(6)	22(11)	-11(3)
C(17)	3463(4)	2623(2)	1570(9)	50(3)	9(1)	145(12)	4(7)	4(13)	-8(2)
O(5)	3026(3)	2985(1)	3425(6)	58(2)	13(1)	186(10)	-38(6)	37(10)	-1(2)
N(2)	4373(3)	-1125(2)	-4531(6)	40(2)	12(1)	84(9)	-4(5)	4(8)	-4(2)
O(2)	4374(3)	-0651(1)	- 523(5)	48(2)	13(1)	74(8)	-10(4)	1(7)	16(2)
O(4)	5308(3)	-1384(2)	900(5)	66(3)	16(1)	112(8)	11(4)	-1(8)	18(2)
C(2)	4941(4)	-1108(2)	- 702(7)	44(3)	12(1)	97(12)	12(6)	2(10)	-3(3)
C(4)	5228(4)	-1306(2)	-3061(7)	42(3)	11(1)	95(12)	-1(6)	-8(10)	8(2)
C(6)	6239(4)	-1002(2)	-3808(7)	41(3)	15(1)	157(13)	6(6)	5(12)	6(3)
C(8)	6191(3)	- 316(2)	-4089(6)	26(2)	15(1)	150(10)	-1(7)	26(10)	4(3)
C(10)	6546(4)	71(2)	-2422(8)	38(3)	16(1)	132(11)	-10(6)	6(11)	-3(3)
C(12)	5868(4)	- 67(2)	-6051(7)	50(3)	15(1)	103(11)	-27(6)	8(10)	-1(3)
C(14)	6571(4)	694(2)	-2797(8)	39(3)	16(1)	172(15)	-24(7)	-16(11)	-9(3)
C(16)	5879(3)	549(2)	-6410(9)	40(3)	16(1)	172(14)	7(7)	-18(12)	-12(3)
C(18)	6263(4)	927(2)	-4790(9)	28(3)	16(1)	252(15)	11(7)	1(12)	-2(3)
O(6)	6306(3)	1541(2)	-5222(7)	50(3)	15(1)	422(15)	21(6)	-50(12)	-8(2)

Table 3

Hydrogen Atom Parameters, CUTY

	x	y	z	B*
H(1)$_1$	.332	.055	-.096	2.4
H(1)$_2$	.433	.073	-.212	2.4
H(3)	.240	.090	-.332	2.2
H(5)$_1$	.429	.149	-.461	2.2
H(5)$_2$	.309	.172	-.547	2.2
H(9)	.181	.210	-.268	3.4
H(11)	.489	.179	-.109	3.0
H(13)	.163	.272	.053	3.5
H(15)	.469	.236	.222	3.1
H(2)$_1$	.389	-.145	-.443	2.6
H(2)$_2$	.467	-.111	-.596	2.6
H(4)	.532	-.175	-.310	2.6
H(6)$_1$	.680	-.110	-.263	3.2
H(6)$_2$	.645	-.119	-.533	3.2
H(10)	.679	-.010	-.091	3.1
H(12)	.561	-.035	-.730	3.2
H(14)	.682	.099	-.157	3.3
H(16)	.560	.073	-.739	3.3

*Isotropic Thermal Parameter

$\sum w(|F_o| - |F_c|)^2$, where $\sqrt{w} = |kF_o| / P$ for $|kF_o| \leq P$, and $\sqrt{w} = P / |kF_o|$ for $|kF_o| > P$, with P = 30 electrons, giving maximum weight to those reflections which were determined most accurately. A previously described logical routine (8) was also used to optimize the refinement. The list of observed and calculated structure factors is shown in Table 4).

DESCRIPTION AND DISCUSSION OF THE STRUCTURE

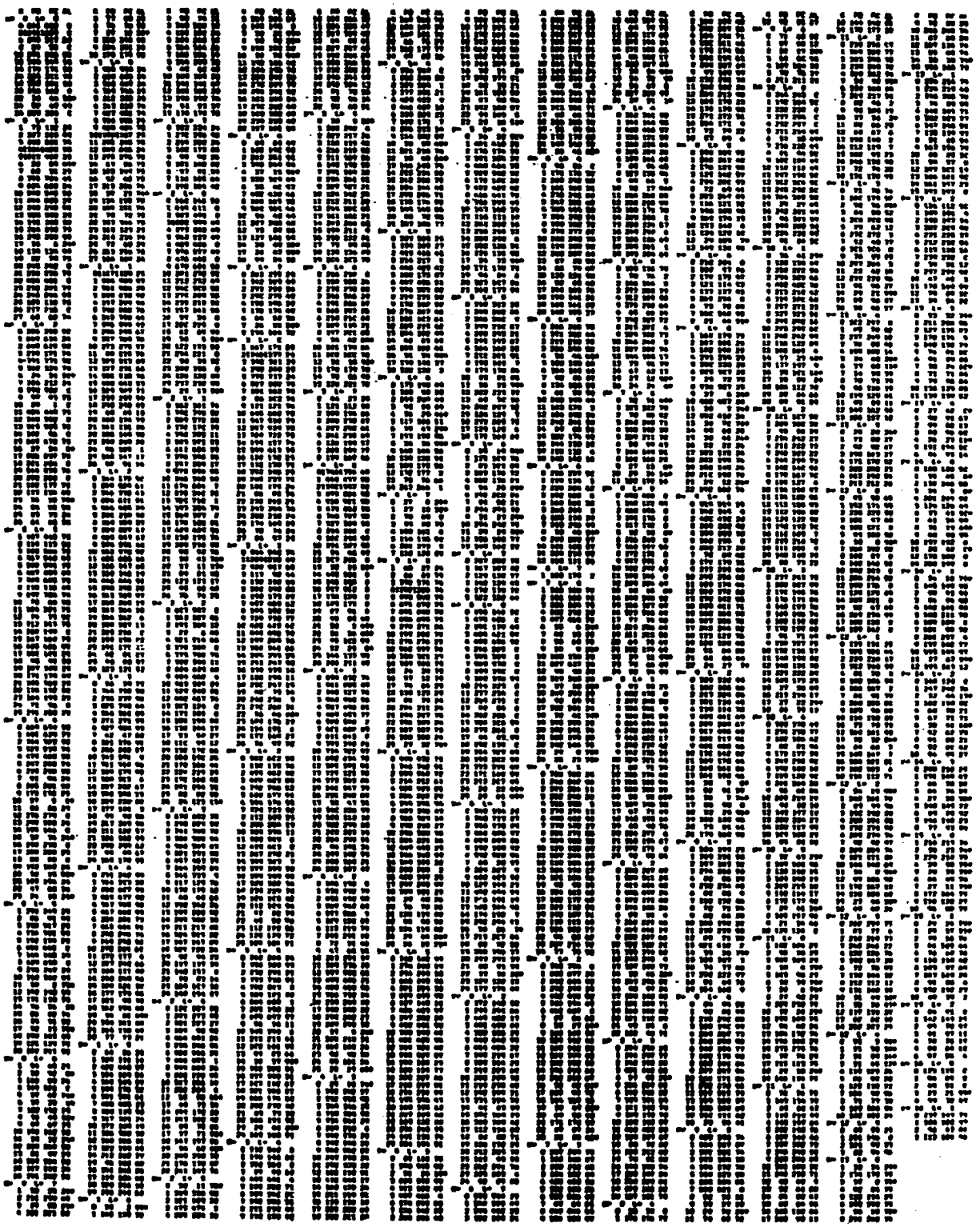
A projection of the structure down the c axis is shown in Fig. 1. The chelate molecules form infinite chains around screw axes parallel to the c axis. Perpendicular to these chains the crystal is stabilized by two strong hydrogen bonds (2.62 and 2.71 Å). No other hydrogen bonding occurs in the structure.

The tyrosine residues form two five-membered chelate rings with the copper atom. The residues are trans with respect to each other. The copper coordination is square pyramidal. The bond distances and bond angles for the copper coordination are given in Figs. 1, 2 and 3. While the Cu-N distances are similar and compare well with the literature value (2.00 Å), the Cu-O distances in the basal plane differ considerably (1.925 and 1.973 Å), and the Cu-O(1) distance is significantly shorter than the average value (1.98 $\pm$ 0.012 Å) given by Freeman (1967) (9). The basal plane of the square pyramid is tetrahedrally distorted (Table 5, Figs. 4, 5), while the copper ion is displaced 0.11 Å from this plane toward the top of the pyramid. The two chelate rings are not planar (Planes 2 and 3, Table 5). The nitrogen atoms show the largest deviations (0.22 and 0.65 Å) from the least-squares

Table 4

Observed and Calculated Structure Factors, CUTY

The values of $|10kF_o|$, $|10F_c|$ and calculated phase angles, in centicycles, are given. Unobserved reflections are indicated by a star.



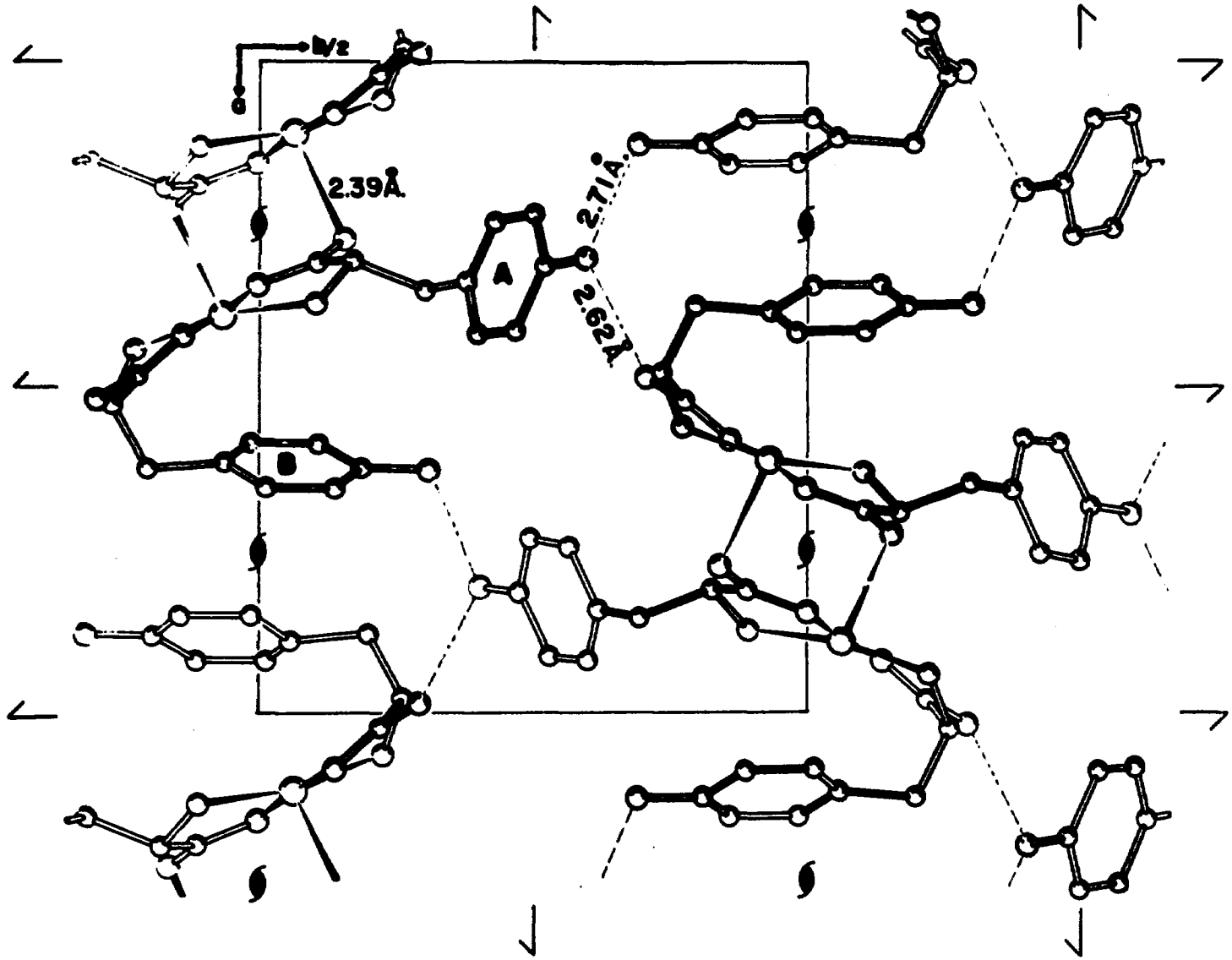


Figure 1. Projection of the structure down the c axis, CUTY

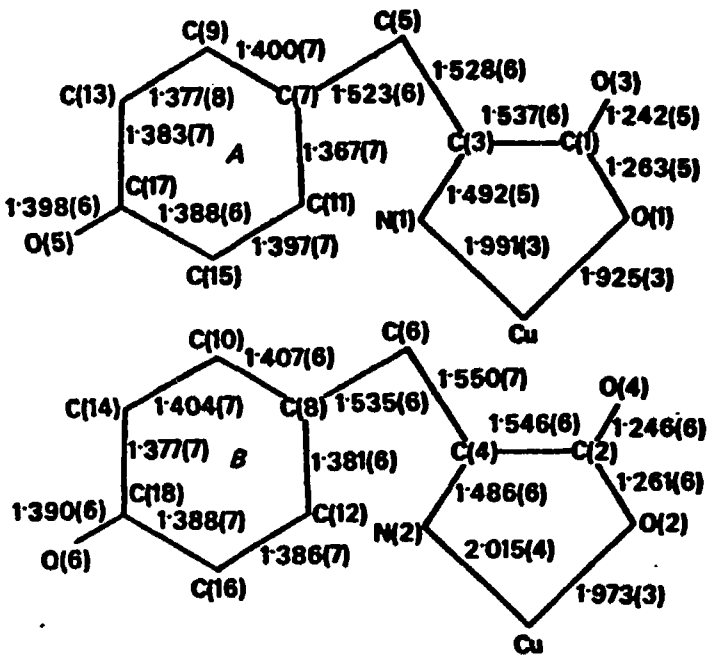


Figure 2. Intramolecular Distances, CUTY.

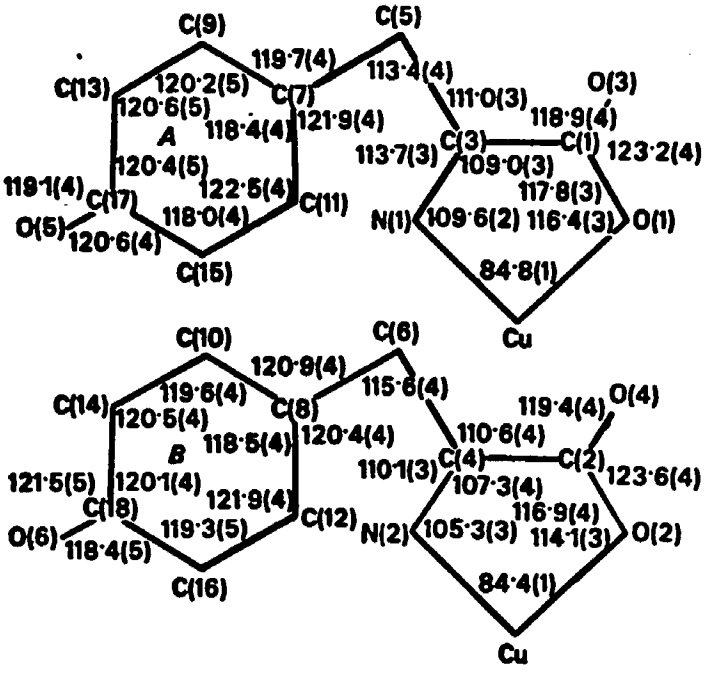


Figure 3. Bond Angles, CUTY

Other angles at Cu: O(3')-O(1)=90.8(1); N(1)O(2)=95.0(1);
O(3')-N(1)=104.3(1); N(1)-N(2)=168.0(1);
O(3')-O(2)=89.8(1); O(1)-N(2)=96.7(1);
O(3')-N(2)=87.6(1); O(1)-O(2)=179.3(1).

Table 5

Least-Squares Planes

The equations of the planes are expressed in the form Ax + By + Cz = D, where x, y and z are fractional coordinates and D is the distance from the origin in Å. The method of Schomaker, Waser, Marsh & Bergman (1959) (16) was used to calculate the least-squares planes.

Plane	Atoms	A	B	C	D
1	N(1), O(1), N(2), O(2)	12.135	6.686	-1.284	5.039
2	O(1), O(3), C(1), C(3)	11.946	7.563	-1.306	4.894
3	O(2), O(4), C(2), C(4)	10.418	13.383	0.054	3.677
4	C(7), C(9), C(11), C(13), C(15), C(17)	3.470	17.783	-3.268	5.241
5	C(8), C(10), C(12), C(14), C(16), C(18)	12.102	-1.854	-2.215	8.450

	Δ(1)		Δ(2)		Δ(3)		Δ(4)		Δ(5)
N(1)	0.099 A	O(1)	-0.006 A	O(2)	0.006 A	C(5)	0.035 A	C(6)	0.130 A
O(1)	-0.098	O(3)	-0.006	O(4)	0.006	C(7)	0.005	C(8)	0.007
N(2)	0.098	C(1)	0.017	C(2)	-0.016	C(9)	-0.002	C(10)	-0.005
O(2)	-0.099	C(3)	-0.005	C(4)	0.004	C(11)	-0.003	C(12)	0.004
Cu	-0.110	N(1)	0.223	N(2)	-0.652	C(13)	-0.005	C(14)	-0.007
O(3)	-2.477	Cu	-0.060	Cu	-0.084	C(15)	-0.003	C(16)	-0.017
						C(17)	0.007	C(18)	0.019
						O(5)	-0.002	O(6)	0.053
								Cu	-2.942

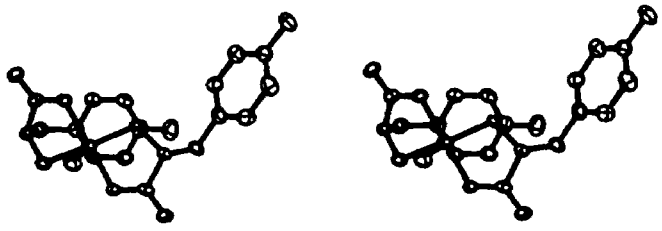


Figure 4. Stereodiagram (17) of CUTY.
Looking along the normal of the phenyl group of residue B.

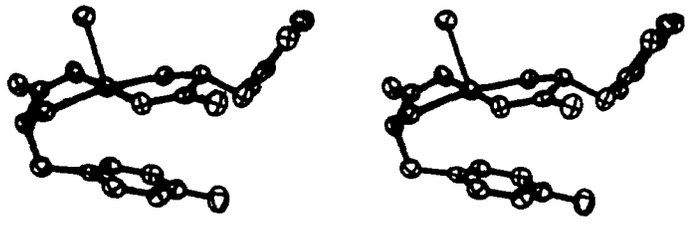


Figure 5. Stereodiagram (17) of CUTY.
The view is rotated 80° about the interocular line from that of Figure 4.

planes through the carboxylic acid groups. The bond angles in the rings are given in Figure 3. The non-planarity of the rings is the result of a rotation around the C^{α}-C' bond (Figure 6), by 11.2° for the C(3)-C(1) bond and -29.3° for the C(4)-C(2) bond. The larger rotation around the C(4)-C(2) bond is reflected in the observation that all bond angles in the chelate ring formed by molecule B are smaller than those in the one formed by molecule A.

The thermal motion of the atoms in the amino acid residues, observed in the present structure (Figs. 4 and 5), is considerably less than the motion found in the copper chelate of L-phenylalanine. The largest thermal motion is observed for O(6) with 6.57 and 2.66 $\AA^2$ for the temperature factors along the major and minor axes. The bond distances and bond angles in two tyrosine residues are shown in Figures 2 and 3. The dimensions of the two residues are quite similar. The valence angles of the C^{β} atoms are significantly greater (113.4 and 115.6°) than the tetrahedral value. This was also observed for one of the phenylalanine residues in CUPA, and in the structure of the potassium salt of L-tyrosine-O-sulfate (10). The other distances and angles are normal.

The conformation of N and C^{γ} about the C^{α}-C^{β} bond is described by the torsional angle X_1 (11). It is found that C^{γ} occurs only for values of X_1 close to 60, 180 and 300°, conformations I, II and III respectively (12).

For amino acids with aromatic side chains, conformation II seems to be the most common (13). Conformation III has been observed in 3,4-dihydroxy-L-phenylalanine (14) and in the two phenylalanine residues in

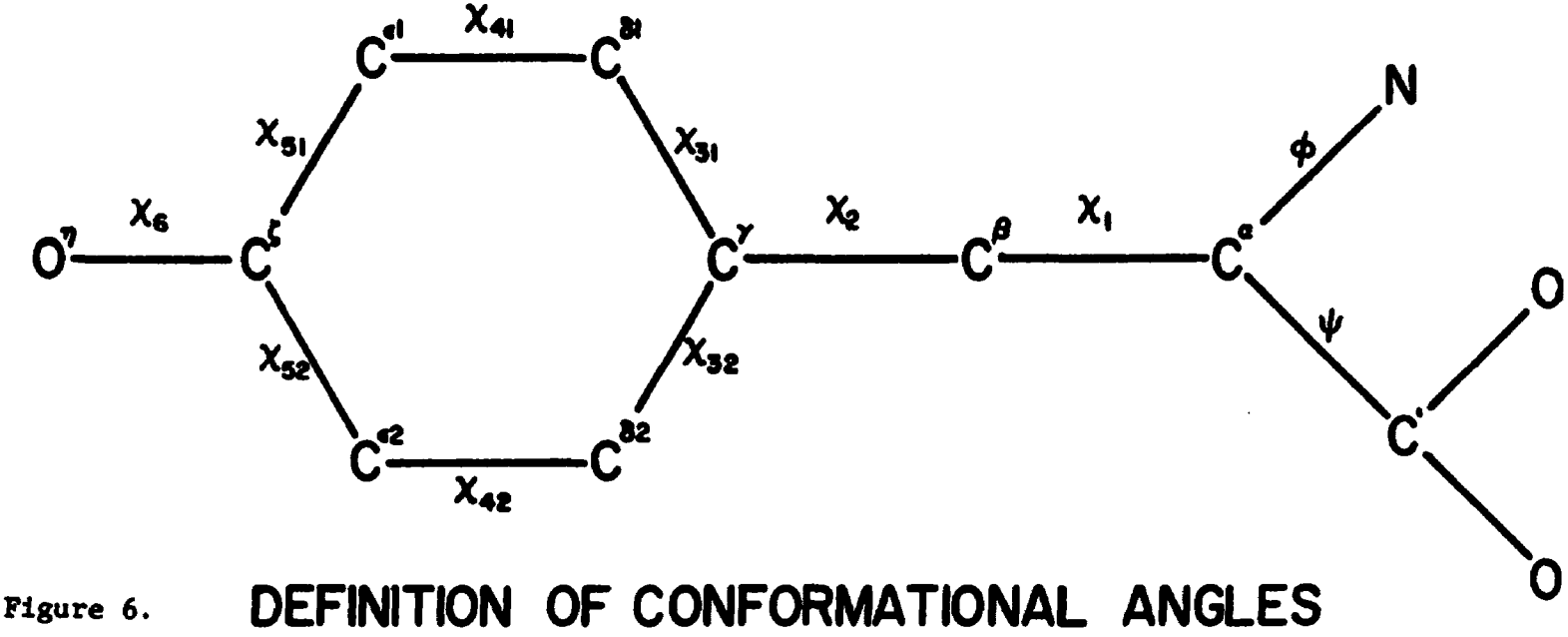


Figure 6. **DEFINITION OF CONFORMATIONAL ANGLES OF TYROSINE**

(Edsall, Flory, Kendrew, & Liquori; BIOPOLYMERS, (1966), 4, 121-130)

CUPA. Residue A in the present structure also has this conformation, with the result that the phenyl group points away from the chelation side. Position I has been observed in L-phenylalanine-HCl (15), in L-tyrosine-O-sulfate (10) and in CUGLT (2). This conformation is observed for residue B in the present structure and allows the phenyl group to be positioned below the base of the square pyramidal copper coordination, as was similarly observed in the CUGLT structure.

The phenyl group of residue B is approximately parallel to the basal plane of the metal coordination (Figure 5, Table 6). There are two close approaches between the Cu^{2+} ion and carbon atoms of the phenyl group [C(8): 3.04 Å and C(12): 3.11 Å] (Figure 4, Table 6). Similar observations were made for both tyrosine residues in the CUGLT structure. The close approaches, as observed, are believed to constitute interactions between the Cu^{2+} ion and the phenyl group. It is interesting to note that these interactions do not occur in the CUPA structure. The possible inferences, regarding enzyme mechanisms, is being considered for publication elsewhere.

The basal plane of the square pyramidal copper coordination is tetrahedrally distorted (Figure 5, Table 5). These distortions (0.10 Å) are similar to those observed in CUGLT, while more than eight times those in the structure of CUPA. It may well be that this distortion is related to the postulated interaction between the Cu^{2+} ion and the phenyl group.

Both aromatic rings are planar (Table 5, Planes 4 and 5). The exocyclic atoms C(6) and O(6) in residue B, however, show large displacements (0.13 and 0.05 Å respectively) from the least squares plane through the six

Table 6

Comparison of CUTY and CUGLT

The notations follow the conventions of Edsall, Flory, Kendrew, Liquori, Nemethy, Ramachandran & Scheraga (1966).

	CUTY		CUGLT	
	Mol. A	Mol. B	Mol. A	Mol. B
Angle between basal plane and phenyl group	--	24°	21°	18°
Cu-C$^{\gamma}$ distance	--	3.04 Å	3.34 Å	3.21 Å
Cu-C$^{\delta}$ distance	--	3.11	3.27	3.17
Average deviation of the four atoms from basal plane	0.098 Å		0.155 Å	0.125 Å
X_1 angle	295.2°	52.5°	54°	59°
X_2 angle	99°/278°	99°/274°	90°/281°	84°/274°

atoms of the aromatic ring system; the convex side of the aromatic ring is presented to the basal plane of the metal coordination (Plane 5).

All intermolecular distances less than 3.5 Å are listed in Table 7.

An interesting feature of the final difference Fourier was the intensity and location of the residual peaks. On placing these peaks in the three-dimencional model, one of the present limitations of x-ray structure determination was emphasized.

According to the present description of the copper(II) ion, the electron distribution is spatially anisotropic. Electron distributions used in calculation of scattering factors are, at this time, for spherical atoms.

The distribution of peaks found in the final difference Fourier - using $\left[|F_o| - |F_c|\right]$ as Fourier coefficients - is just what one would expect if a similar calculation were done, using instead of $|F_o|$, a scattering function representative of the anisotropic quantum mechanical ion.

Table 7

Intermolecular Distances Less Than 3.5 Å, CUTY

N(1)	O(1^{i})	3.086 Å	N(2)	O(3^{i})	3.063 Å	(C)
N(1)	O(3^{i})	3.468 (C)	O(2)	O(3^{i})	3.096	
N(1)	C(1^{i})	3.392	O(2)	C(1^{i})	3.268	
N(1)	O(3^{iv})	3.079	O(2)	C(3^{i})	3.430	
O(1)	O(3^{i})	3.092 (C)	O(2)	O(1^{iv})	3.276	
C(3)	O(1^{i})	3.253	O(4)	N(2^{iv})	3.087	
C(15)	O(4^{ii})	3.415	C(10)	C(12^{v})	3.477	
C(15)	C(5^{iv})	3.484	Cu	O(1^{i})	3.388	
C(17)	O(4^{ii})	3.351	Cu	O(3^{i})	2.391	(C)
O(5)	C(18^{iii})	3.439	Cu	C(1^{i})	2.974	
O(5)	O(6^{iii})	2.709 (H)				
O(5)	O(4^{ii})	2.620 (H)				

The small letter in parentheses indicates that one of the following operations has to be applied to the coordinates given in Table 2.

i) $\frac{1}{2} - x, \quad - y, \frac{1}{2} + z$

ii) $1 - x, \frac{1}{2} + y, \frac{1}{2} - z$

iii) $-\frac{1}{2} + x, \frac{1}{2} - y, \quad - z$

iv) $x, \quad y, 1 + z$

v) $1\frac{1}{2} - x, - y, \frac{1}{2} + z$

The letter H indicates hydrogen bonds, and the letter C, distances occurring in the metal coordination.

PART II

THE CRYSTAL AND MOLECULAR STRUCTURE OF N-ACETYL-L-PROLINE MONOHYDRATE

INTRODUCTION

Molecular rationalization of biological processes has long been difficult, if not impossible. As new probes - optical spectroscopy, nuclear magnetic resonance, optical rotatory dispersion, fluorescence tagging, crystal structure analysis, to name a few - have become available, additional clues have been provided to those speculating on these topics. Of particular interest is the growing body of knowledge regarding the structure and function of amino acids, peptides and proteins. Interactions are being observed which were not anticipated from study of the component molecules in their more simple forms.

Initial data was for molecules such as could survive the comparatively brutal attack of mineral acids, inorganic bases and sustained high temperatures. Such data was, and is, essential. Yet interpretation of this data is valid only for molecular events energetically equivalent to the strength of 'chemical bonds' - ca. 100 kcal/mole. The postulation (18), and immediate verification (19) of hydrogen-bond stabilized helices of amino acid residues in proteins identified the so-called 'weak interactions' as forces to be reckoned with in the chemical models of biological interactions.

As the necessary computing power - theoretical framework, "software" programming, and machinery - was developed, x-ray crystallographers were enabled to process the masses of data associated with crystal structures of proteins (20, 21). This provided objective molecular models by which inferences made using other techniques could be tested. Thus, while protein crystal structures were being refined, so were the complementary experimental and mathematical techniques. In this manner, the foundation was laid for predicting, as well as observing, the molecular structure and behavior of large, biologically important molecules. Contributions by Ramachandran and co-workers (12) along these lines are many. In recent years additional groups have undertaken similar studies using various noncrystallographic techniques (22, 23, 24).

When incorporated in a peptide chain, the amino acid L-proline, (Pro), has effects which are distinctive from other - except hydroxy-L-proline, (HyPro) - amino acids. These effects are related to the fact that the C^{δ} -N bond (see Figure 7a) replaces the N-H amide bond found in other peptide linkages. This has two very important effects. The nominally free rotation about the C^{α} -N bond is restricted, fixing the angle between the C^{α} -C" and N-C' (see Figure 7b) within narrow limits. In addition, there is no amide proton for participation in hydrogen-bonding.

Thus proline has acquired the title of "helix-breaker". Dickerson and Geis state with reference to the extensive α-helix structure of myoglobin and hemoglobin, "...every bend does not have a Pro, but every Pro produces a bend" (25). Similarly, proline disrupts the extended pleated

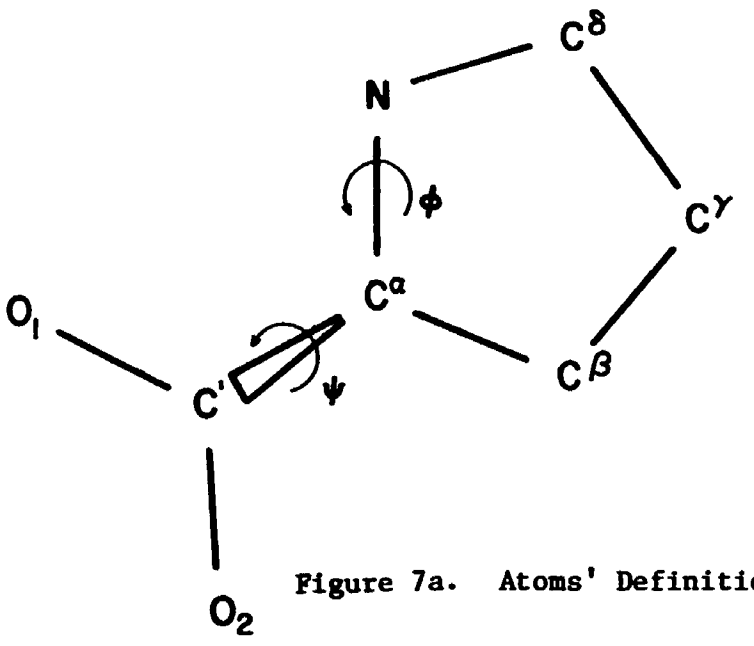


Figure 7a. Atoms' Definition, L-Proline

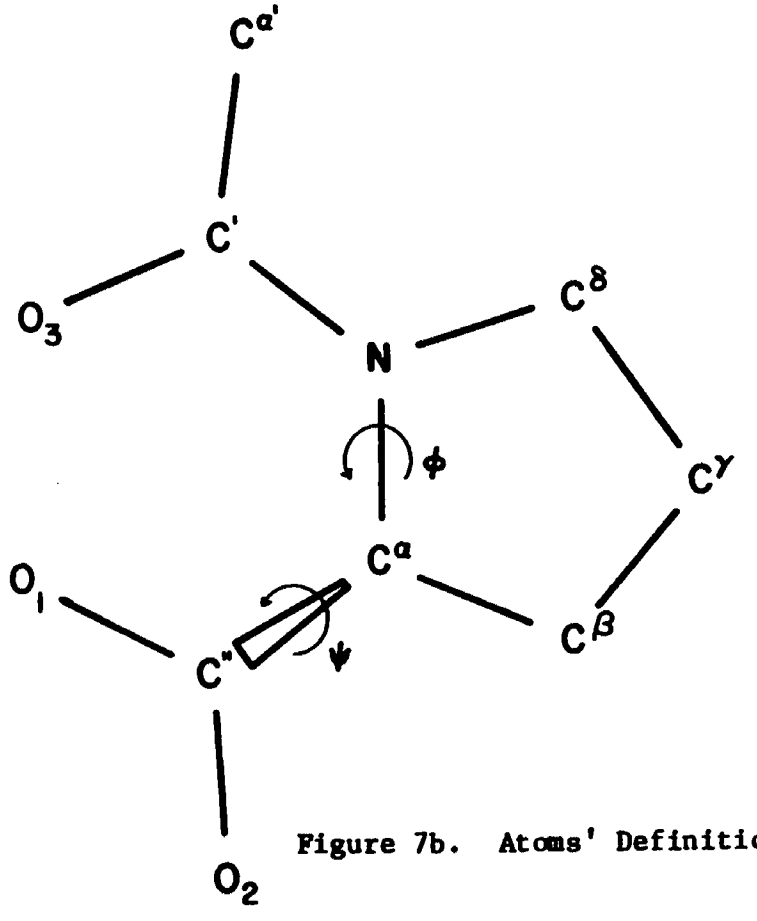


Figure 7b. Atoms' Definition, NAP

sheet structure frequently found in polypeptides and proteins. Experimental and theoretical investigations of the structures of cyclic polypeptides (22, 23, 24) frequently show proline residues at corners of an intramolecularly bonded pleated sheet structure. It appears that one of the functions of proline is the redirecting of major structural components of larger molecules which might otherwise extend indefinitely.

Proline is also one of the major components of collagen which maintains the structural integrity of many non-rigid biological structures. A triple-helical model has been fit to x-ray data using molecular models of the component amino acids (26).

Programs are under way in several laboratories (for example, see 22, 23, 24) to calculate the conformations and structures of polypeptide molecules based on structural data available for the component amino acids. Results (22, 27) of such calculations indicate several configurations may be equivalent to within a few kilocalories per mole. The results of such calculations have also been shown (28, 29, 30) to be sensitive to the structural parameters assumed for the component molecules.

Especially for non-crystallographic techniques, it is important that extrapolation be made from reliable data based on allowed and/or preferred molecular parameters (e.g. bond distances and angles, conformation, intermolecular interactions).

In order to obtain such data on peptide-bonded proline, with a minimum of extraneous influences, it was decided to determine the structure of N-acetyl-L-proline (NAP) using the technique of single-crystal x-ray crystallography.

EXPERIMENTAL

Single crystals of NAP were grown by slow evaporation of an aqueous solution of the compound as received from Cyclo Chemical Corporation, Los Angeles, California. Thick, transparent, colorless plates developed in the viscous solution in a few days. The major faces, the $\bar{1}01$ and $10\bar{1}$ planes, were bounded by the 121, $\bar{1}\bar{2}\bar{1}$, $1\bar{2}1$ and $\bar{1}2\bar{1}$ planes.

Space group determination of NAP was carried out on a G.E. XRD-5 diffraction unit, equipped with an SPG single crystal orienter. Initial cell constants and systematic absences indicated space group P_{2_1} (No. 4) (31) as previously determined (32). The data crystal was transferred to an Enraf-Nonius CAD-4 automatic diffractometer for further study. Cell constants (Table 8) were determined by measurement of the diffraction angle θ for ten general crystallographically independent general reflections. Data for each of these reflections was measured at each of the four equivalent (for space group P_{2_1}) reciprocal lattice points. Thus 40 measurements were used in computing least squares cell parameters.

Integrated intensity measurements were made on the same crystal used in determination of the final cell constants. Data were collected for the $(h\bar{k}l)$ reciprocal hemisphere to the instrumental limit of $\theta = 77^{\circ}$. This corresponds to a resolution of 0.79 Å. In this way, two separate but equivalent intensity measurements were made for each crystallographically independent reflection, since one quadrant of reciprocal space is unique for monoclinic space groups. (The axial data - of the type (0k0) - is present

Table 8

Crystallographic Data for N-acetyl-L-proline Monohydrate (NAP)

Formula: $H_{13}C_7NO_4$

F.W. = 179.19

Systematic absences: 0k0, k = 2n + 1

Space group = P_{2_1}

a = 6.6144(1) Å*

b = 10.6897(2) Å

c = 6.6636(2) Å

$\alpha = \gamma = 90^{\circ}$

$\beta = 108.814(2)^{\circ}$

*(determined by least squares fit to the 2θ values of 40 reflections, between 41° and 79° 2θ)

Z = 2

ρ_{obs} = 1.301 gm cm^{-3} (by flotation, at 29°C, in CH_2Cl_2-n-Butylbromide mixture)

ρ_{calc} = 1.306 gm cm^{-3}

F(000) = 188.

Crystal dimensions: 0.12 mm. x 0.28 mm. x 0.28 mm.

Crystal volume: 0.0095 mm^3 (as calculated by DATAP2).

μ (CuK_α) = 9.177 cm^{-1}

only once in the ($h\bar{k}l$) hemisphere). Two such data sets were collected using Ni-filtered copper radiation; this made a total of four intensity measurements for each unique non-axial reciprocal lattice point (two measurements for the axial lattice points). ω-2θ scans (also called θ-2θ scans) were used.

Data collection parameters used are:

ω-scan angle:	1.0 deg. + (0.15 deg.)tan θ
counter aperture:	4.0 mm + (0.7 mm) tanθ
maximum scan time:	200 seconds
desired peak-counts:	50000
time between measurement of monitor ($\bar{2}03$) reflection:	60 minutes

After the prescans for each reflection, a suitable scan rate was calculated so the desired peak-counts would be obtained during the final intensity scan; the lowest scan rate was 3/20 that of the highest. Scan time was calculated; if this exceeded the specified maximum scan time, the scan rate was appropriately increased so the maximum scan time was not exceeded. If the counting rate, with attenuator inserted, exceeded 50,000 counts/sec. the reflection was skipped for later measurement using additional attenuators.

Integrated peak intensities were computed and scaled for variations in monitor intensity by the relationship

$$\text{Intensity} = (\text{Monitor Scale Factor/NPI})(\text{Filter Scale Factor})(N_T - N_{B1} - N_{B2}).$$

where $3 \leq \text{NPI} \leq 20$ is the relative rate at which the intensity scan was observed,

N_T, N_{B1}, N_{B2} are total integrated counts for peak scan, background 1, and background 2 respectively.

Raw data estimates of the relative weights (33; vide infra) were also calculated. Reflections for which the net intensity = 1.2x $\sqrt{\text{total counts}}$ were assigned a net intensity of 0.6x $\sqrt{\text{total counts}}$ and flagged unobserved. Absorption corrections were made using a locally modified version of the program DATAP2 (6, 34) with 288 sampling points. This program uses the numerical integration method of Gauss. Although the linear absorption coefficient (μ CuK$_\alpha$ = 9.177 cm^{-1}) is small, calculated transmission factors varied from a high of 0.899 for the ($\bar{7}06$) reflection to a low of 0.792 for the (1 $\overline{13}$ 1) reflection.

At this point we had two data sets, each containing complete intensity data for the two equivalent quadrants of the (h$\bar{k}$l) hemisphere. This redundancy allowed calculation of the limit of accuracy imposed by the data collection procedure, as well as a check of the internal consistency of absorption correction. Assessment of the accuracy of the data was now more than an academic problem. For each unique reflection, there were four intensity values. Each was an objective measurement, therefore should not be included in, or excluded from, the final data set arbitrarily.

A function (33, 35, 36) which is used to estimate the standard deviation, σ_I, of an intensity measurement for crystallographic data is of the form

$$\sigma_I^2 = \sigma_{pk}^2 + (B \cdot N_{pk})^2$$

where it is assumed that

$$\sigma_{counts}^2 = N,$$

for

$$N = \text{no. of counts},$$

$$\sigma_{pk}^2 = N_T + N_{B1} + N_{B2},$$

$$N_{pk} = N_T - N_{B1} - N_{B2},$$

where B is an input parameter described below.

It can be shown (33) that for the reflection having indices (hkl) the estimate of the standard deviation of the observed structure factor, $|F^o_{hkl}|$, may be written in the form

$$\sigma_{F^o_{hkl}} = (K/\sqrt{N_{pk}})(\sigma_I) ,$$

where

K is a combination of various correction factors which depend only on the geometry involved, not on intensity (excepting extinction effects).

Thus, the error estimate in each of the $|F^o_{hkl}|$ is directly related to the raw data and B. This method allows estimation of the error of a single measurement and is frequently used in crystal structure analyses.

This "raw data" error estimate is then available for direct use in least squares refinement. For a particular model - coordinates, thermal parameters, etc. of a collection of atoms - the structure factor for each reflection (hkl), F^c_{hkl}, may be calculated. Using standard least squares techniques, the sum of the Δ^2_{hkl} ($= \left[|F^o_{hkl}| - |F^c_{hkl}| \right]^2$) is minimized for all (hkl) in the data set. In this form, all data is treated equally, regardless of accuracy. In practice, the sum of $\left[w_{hkl} \cdot \Delta^2_{hkl}\right]$ is minimized, where w_{hkl} denotes a 'weight' associated with $|F^o_{hkl}|$. The methods of choosing a value for the weights are many: when estimates of standard deviations are available as in this determination, the weight may be calculated (34) as

$$w_{hkl} = 1/(\sigma_{F^o_{hkl}})^2$$

This method of calculating weights has the distinct advantage that all (except the F^{c}_{hkl}'s, which are being refined) quantities in the least squares sums

$$\sum_{i=1}^{m} w_{hkl} \cdot \Delta^{2}_{hkl}$$

for m reflections (hkl), are derived from the data. The exception is the value of B.

The value chosen for B, although apparently arbitrary does appear to reflect the short-term stability of a particular instrument. The validity of the value is tested at the end of a refinement by plotting ($w_{hkl} \cdot \Delta_{hkl}$) vs. pertinent parameters, such as $|F^{o}_{hkl}|$, $\sin\theta_{hkl}$, etc. When such a curve is constant, i.e. slope = 0, the weighting function is considered satisfactory (37). Thus, when the weighting function described above is used, and the curve is a constant, the value of B is considered proper.

Using a new instrument, we were not certain of what accuracy to expect, or what would be a proper choice of B. The ex post facto method is not particularly satisfying. Since a complete unique intensity data set for NAP required approximately three days for collection, we obtained redundant intensity data as described above. This would allow an independent evaluation of the estimated standard deviations, σ_F, as calculated above. For "small" data sets, one can estimate the standard deviation, s, where

$$\sigma^{2} = 1/(q-1) \sum d_i^{2}$$

where q = no. of independent measurements, four for this experiment

d_i = deviation of intensity i from the mean of the q equivalent intensity measurements.

In earlier data reduction, the value of B had been specified as 0.01 for NAP. A sampling of all data was divided into ten intensity ranges. For each range the average σ_I and average $\underline{s}$ was plotted with results similar to Figure (i).

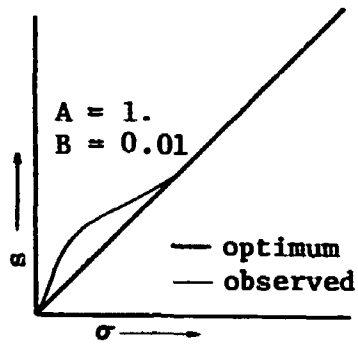


Figure i

It was noted that the equation

$$\sigma_I^2 = \sigma_{pk}^2 + (B \cdot N_{pk})^2$$

is actually of the form

$$\sigma_I^2 = (A \cdot \sigma_{pk})^2 + (B \cdot N_{pk})^2$$

where A = 1.

The data was then plotted for A = 2.0 and A = 3.0 giving curves similar to those shown in Figures (ii), (iii). On the basis of these results, revised σ_I's and relative weights were calculated using

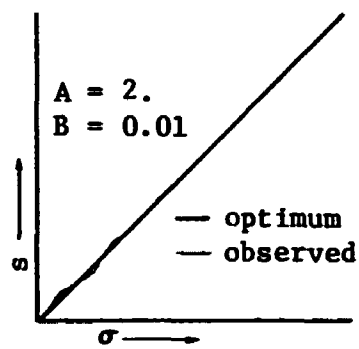


Figure ii

$$\sigma_I^2 = (A \cdot \sigma_{pk})^2 + (B \cdot N_{pk})^2$$

where A = 2.0,

B = 0.01

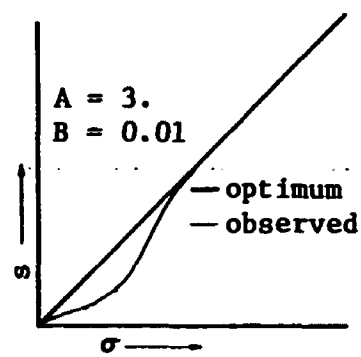


Figure iii

Using the error estimates thus obtained, the ratio $(\Delta_{hkl_i} / \sigma_{hkl_i}, i=1,..,4)$ was calculated for each of the four equivalent data sets. There were ten intensities for which $(\Delta_{hkl_i} / \sigma_{hkl_i}) > 5$. For eight of these there was a

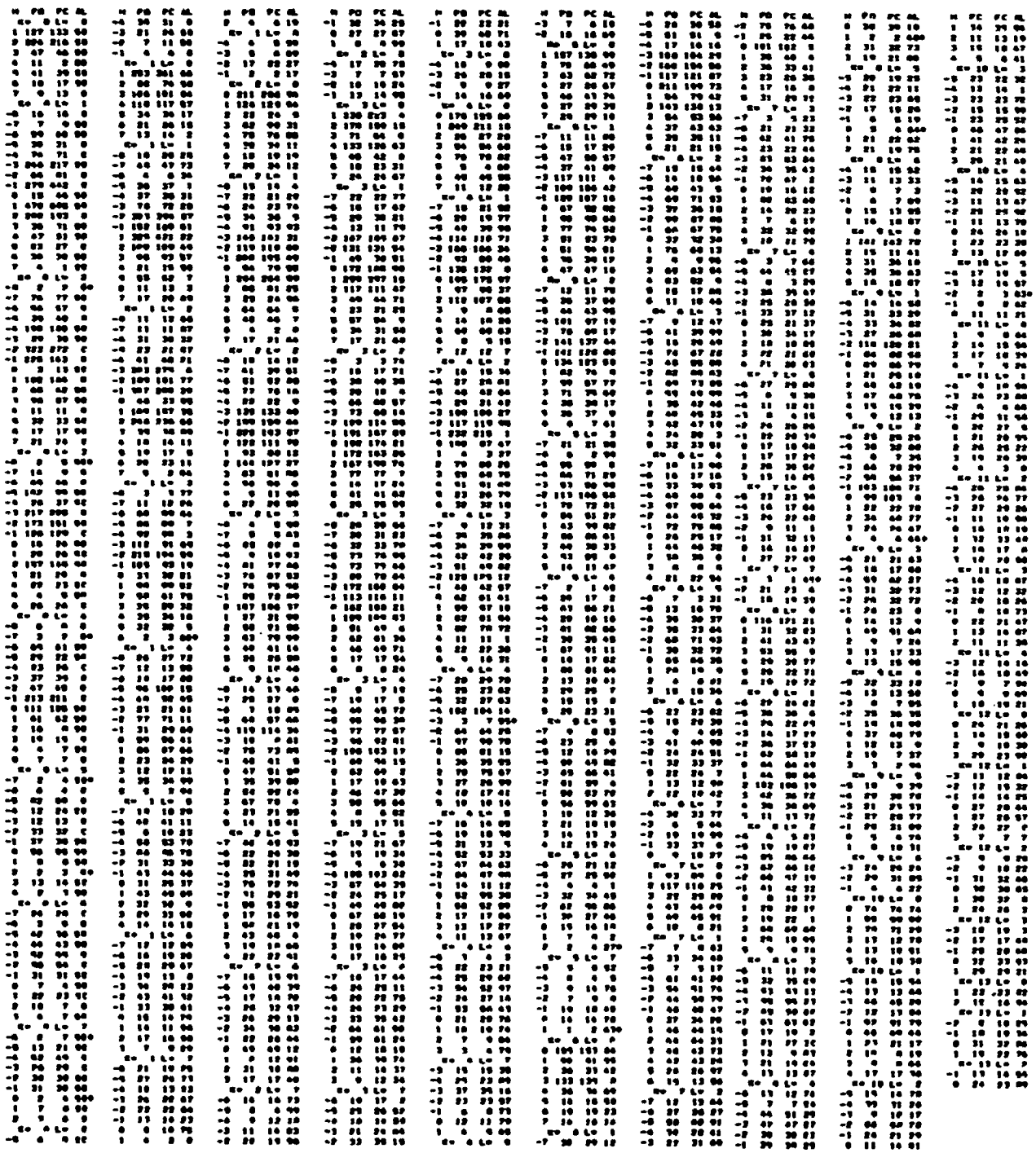
record of instrument malfunction during data collection. The two values obtained for the $(10\bar{1})$ reflection were 40% of those obtained for the $(\bar{1}01)$, possibly due to extinction effects, and were also removed from the data. (In the final structure factor calculation, $\left[|F^{o}_{hkl}| - |F^{c}_{hkl}|\right]$ = -16 electrons; it appears that both the $(\bar{1}01)$ and the $(10\bar{1})$ reflections were strongly affected). Omitting these points, the equivalent intensity data and the relative weights were averaged, producing the final data set containing only crystallographically unique reflections. These weights were used in the rest of the refinement. An indication of the limiting accuracy of the data was also obtained at this time. R_D $\left[= \Sigma|d_i| \; / \; \Sigma I_{pk}\right.$, where I_{pk} is the peak intensity (N_{pk}) scaled and corrected for absorption effects; sums over all intensity data$\left.\right]$ was calculated. For the data described above, R_D = 0.034. This may be compared with the value of the 'conventional R' $\left[=\Sigma\left||F^{o}_{hkl}| - |F^{c}_{hkl}|\right| \; / \; \Sigma \; |F^{o}_{hkl}|\right]$ obtained at the end of refinement of the structure. Reports of $R<R_D/2$ apparently indicate overoptimism for R, and most likely, in the other estimates of errors.

Anticipating a function of one of the impending data reduction programs, in order to maintain control of the signs of the (hkl) indices, the indices were transformed from the $(h\bar{k}l)$ to the (hkl) reciprocal hemisphere, and further transformed so that all data were of the set $\left[hkl, \bar{h}kl\right]$. Such data are completely equivalent (except for anomalous dispersion effects, which are very small for light atom compounds) to the original data, and are those shown in the list of calculated and observed structure factors (Table 9).

TABLE 9

Observed and Calculated Structure Factors, NAP

The values of $|10kF_o|$, $|10F_c|$ and calculated phase angles, in centicycles, are given. Unobserved reflections are indicated by an asterisk.



STRUCTURE SOLUTION AND REFINEMENT

As described previously (32), several unsuccessful attempts had been made to obtain a trial structure which could be refined. We recently obtained a tested set of non-centric direct phasing programs (38). It was decided that another attempt should be made, using these programs, to obtain a trial structure for NAP. Since the newly collected data were not yet ready for use, we used the original data to recalculate a three-dimensional Patterson map, which showed features seen in the earlier work. Preliminary calculations were made, in which overall scale (.1048) and temperature (4.33) factors were obtained. Several unsuccessful sets of origin-defining phases were tried. During the search for possible sources of error, it was noticed that the programs being used were limited - in P_{2_1} - to data of either (hkl) or (h$\bar{k}$l). Similarly, mixture of crystallographically equivalent indices of ($\bar{h}$kl) and (hk$\bar{l}$) is not allowed. This restriction was not being met for the data in use. Transformation of the data to $\left[hkl, \bar{h}kl\right]$ form was carried out.

In addition, it was felt that the search for starting reflections should be made more systematic. Accordingly, a matrix was set up in which previously generated Σ_2 interactions (39) were tabulated. This allowed a trial set of starting phases to be chosen and their propagation via the Σ_2 relationship to be checked quickly by hand. By this means it was found that the set chosen led most rapidly to initial phase estimates for most of the reflections with $E_{hkl} \leq 1.7$. ($E_{hkl} = k|F^o_{hkl}|$, where k is a

function which essentially corrects for thermal motion). Choice of a symbolic phase by the same method indicated that all reflections with $E_{hkl} \geq 1.7$ should have initial phase estimates within the first ten iterations of the phasing program. Use of the $\sum_1$ relationship (40) indicated that phases for the structure invariants $\bar{6}04$ and $\bar{4}02$ should both have a value of π. Hand calculations using the $\sum_2$ relationship indicated that the value of the symbolic phase should be either zero or π. In addition, it appeared that the phase of the $\bar{3}12$ reflection should be $\pi/2$. Tangent refinement using this starting set of phases failed to give reasonable results. Since a complete set of these calculations required about 4 minutes of computing time (on an IBM 360/50, 192 reflections included) the approach suggested by Germain and Woolfson (41) was followed, letting the phase of the $\bar{3}12$ reflection take the initial value of $\pi/4$, $\pi/2$, $3\pi/4$, and π. Results of the calculation starting with $\phi_{\bar{3}12} = \pi$ appeared quite good: starting phases were rapidly extended and the value of $R_K \left[= \sum \left| |E^o_{hkl}| - |E^c_{hkl}| \right| / \sum |E^o_{hkl}| \right.$, summed over the $E \geq Emin$, in this case $Emin = 1.7$] was below 0.20.

The phase of the $\bar{3}12$ reflection gradually shifted to a value of $\pi/2$ during tangent refinement. Although no apparent reason could be found for this behavior, it has been reported previously in a related structure (42). An E-map (Fourier synthesis using E_{hkl} rather than F_{hkl}) using the phases thus obtained was calculated and a model built. The NAP molecule and the oxygen of the water of hydration were clearly visible except for atom C(3). A structure-factor ($R = 0.36$) - Fourier calculation was used to locate this atom. Inclusion of this atom in a structure factor calculation yielded

R = 0.26. Further refinement by structure factor block-diagonal-least-squares converged to R = 0.10.

When the newly collected data, described above, were ready for use, refinement was continued using it in an attempt to obtain accuracy approaching the indicated limit of R_D = .03 - .04. This produced no significant changes. It was then noticed that the thermal ellipsoids for the C^{β} and C^{γ}, atoms C(2), C(3), were so large as to be physically meaningless (i.e. larger than 8-10 $\AA^2$ which is usually considered quite high). The parameters for these atoms were deleted from three cycles of refinement. A difference Fourier (coefficients of $\left[|F_o| - |F_c|\right]$) was then calculated. These two atoms appeared disordered in the direction perpendicular to the plane of the pyrrolidine ring. The fractionally occupied sites were visually located from this synthesis.

Further refinement proceeded via two routes for the following reasons. Previous crystal structure determinations involving proline have consistently described the C^{γ} atom as having extremely high thermal motion in a direction approximately normal to the plane of the ring (Table 10). Leung and Marsh (43) reported the C^{γ} atom of the prolyl residue in the structure of leucyl-L-prolyl-glycine is disordered, and that the 'half-atoms' are separated by about 0.6 $\AA$. No other refinement of proline structures is known in which the pyrrolidine ring atoms are treated as disordered. On the other hand, increasing the number of parameters without a proportional increase in the amount of data should lead to improved fit of the model to the data, but the improvement is of doubtful significance.

Table 10

Major Axes of Anisotropic Thermal Ellipsoids and Comparison of Anisotropic Thermal Parameters of Pyrrolidine Rings.

Compound	Atom	Magnitude of Major Axis of Thermal Ellipsoid	Angle between Major Axis and Normal to Ring Plane*
NAP	C^{β}	16.5 $Å^2$	16°
APNMA	C^{α}	3.8	70
	C^{β}	6.0	40
	C^{γ}	6.1	13
	C^{δ}	4.6	20
Pro	C^{α}	2.3	61
	C^{β}	7.4	16
	C^{γ}	18.1	14
	C^{δ}	6.6	27
CBZ GPLGP	C^{α}	8.2	36
	C^{β}	7.2	39
	C^{γ}	13.4	26
	C^{δ}	11.3	27
CBZ GPLG	C^{α}	5.4	83
	C^{β}	11.0	33
	C^{γ}	13.0	25
	C^{δ}	9.2	44

* Ring planes calculated using atoms C^{α}, C^{β}, C^{δ}, N, except for the ring in Pro, in which atoms C^{β}, C^{γ}, C^{δ}, N were used.

It was intended that the two-pronged refinement - allowing disorder of the C^{γ} atom in both cases, but not allowing C^{β} disorder in one - would provide means for deciding whether disorder of the C^{β} atom was physically meaningful. (There was no doubting the C^{γ} was disordered). Fixing the y-coordinate of the nitrogen atom at $\frac{1}{2}$, the two models were each subjected to three cycles of structure factor full-matrix least-squares refinement. Coordinate shifts for atoms in the last cycle of each refinement were less than $\frac{1}{2}$ of the estimated error. Final parameters for both refinements are given in Tables 11 and 12. Table 9 lists observed and calculated structure factors for the refinement with C^{β} and C^{γ} disordered. Some other structures containing the proline residue are listed in Table 13, with the abbreviations used in the following discussion.

Although several diffuse peaks occurred in the final difference Fouriers which may be interpreted as hydrogen atoms, only that peak corresponding to the hydrogen (C^{α}-H) disappeared when it was included in a structure factor difference Fourier calculation. Inclusion of coordinates for other hydrogen atoms did not improve the situation. Several diffuse peaks from 0.2 - 0.5 electrons/Å^3 occur, as well as two regions of negative electron density, indicating appreciable discrepancy between the model and data. The quantity $(w_{hkl} \cdot F^{o}_{hkl})$ is approximately 1 for all ranges of F^{o}_{hkl} except for the very highest values. The large discrepancies for the F^{o}_{hkl}'s are a feature common to proline-containing structures. A stereoscopic view of the unit cell, looking in the $\bar{1}0\bar{1}$ direction is shown in Figure 8.

Table 11

Final Parameters, N-acetyl-L-proline Monohydrate. C^{γ} Disordered.

Atom	x/a	y/b	z/c	B_{iso} or B_{11}	B_{22}	B_{33}	B_{12}	B_{13}	B_{23}
C(1)	.3165(11)	.3780(7)	.4073(9)	353 (19)*	69(5)	243(14)	7(8)	138(14)	1(7)
C(2)	.1804(19)	.3927(10)	.5535(16)	897 (49)	121(9)	601(34)	43(17)	594(37)	1(15)
C(31)	.0493(32)	.4990(19)	.4769(32)	5.4(5)					
C(32)	.1228(43)	.5341(24)	.5510(43)	6.1(6)					
C(4)	.1648(9)	.5887(8)	.3677(10)	269 (15)	95(6)	300(18)	38(9)	109(14)	-36(9)
C(5)	.2380(8)	.2738(7)	.2496(9)	199 (12)	66(4)	287(15)	- 1(7)	120(12)	- 3(7)
C(6)	.4231(8)	.5242(6)	.1828(9)	247 (14)	67(5)	272(15)	- 4(7)	109(12)	- 9(7)
C(7)	.4083(13)	.6514(8)	.0859(12)	499 (27)	66(6)	459(26)	- 7(11)	253(22)	7(10)
N(1)	.3004(7)	.5000(0)	.3028(7)	239 (11)	60(4)	249(12)	14(6)	111(10)	-14(6)
O(1)	.3200(6)	.1651(5)	.3320(6)	307 (12)	68(4)	277(12)	14(6)	99(10)	13(6)
O(2)	.1089(6)	.2846(6)	.0724(7)	263 (10)	78(4)	332(13)	14(6)	20(9)	-19(6)
O(3)	.5405(7)	.4403(6)	.1512(7)	326 (12)	85(4)	383(13)	31(6)	207(11)	10(6)
O(4)	.2223(7)	-.0374(6)	.1084(7)	321 (12)	80(4)	344(12)	-18(6)	162(10)	-43(6)
H(10)	.487	.352	.491	5.0					
H(41)	.260	.664	.470	5.5					
H(42)	.056	.625	.248	5.5					

*Anisotropic temperature factor of the form: $\exp\left[-(h^2B_{11}+k^2B_{22}+l^2B_{33}+2hkB_{12}+2hlB_{13}+2klB_{23})\right] \times 10^4$

Atom	Occupancy
C(31)	0.54(5)
C(32)	0.46(5)
All others:	1.0

Table 12

Final Parameters for N-acetyl-L-proline Monohydrate. Cβ, Cγ Disordered.

Atom	x/a	y/b	z/c	B_{iso} or B_{11}	B_{22}	B_{33}	B_{12}	B_{13}	B_{23}
C(1)	.3175(11)	.3782(7)	.4075(10)	346 (19)*	70(5)	234(14)	7(8)	135(14)	3(7)
C(21)	.1180(43)	.3916(17)	.5062(36)	4.1(6)					
C(22)	.2124(31)	.3940(12)	.5783(26)	5.2(3)					
C(31)	.0487(32)	.4980(19)	.4771(32)	5.3(5)					
C(32)	.1227(41)	.5335(24)	.5492(41)	5.3(6)					
C(4)	.1643(9)	.5887(7)	.3673(10)	269 (15)	92(6)	299(17)	37(9)	112(14)	-33(9)
C(5)	.2380(8)	.2739(7)	.2497(9)	204 (12)	66(5)	289(16)	- 3(7)	122(12)	- 5(7)
C(6)	.4226(8)	.5242(7)	.1821(9)	239 (14)	67(5)	271(15)	- 4(7)	108(12)	- 9(7)
C(7)	.4076(12)	.6515(8)	.0852(12)	472 (26)	68(6)	455(26)	- 3(11)	244(22)	8(10)
N	.3007(7)	.5000(0)	.3031(7)	240 (12)	59(4)	246(12)	12(6)	109(10)	-14(6)
O(1)	.3205(6)	.1653(5)	.3324(6)	308 (12)	67(4)	277(12)	13(6)	99(10)	13(6)
O(2)	.1088(6)	.2846(6)	.0720(7)	258 (10)	79(4)	329(13)	14(6)	20(10)	-18(6)
O(3)	.5406(7)	.4404(6)	.1510(7)	337 (13)	84(4)	386(14)	31(7)	219(11)	9(6)
O(4)	.2224(7)	-.0372(6)	.1083(7)	329 (12)	79(4)	344(12)	-22(6)	167(10)	-45(6)
H(1)	.487	.352	.491	5.0					
H(41)	.260	.664	.470	5.5					
H(42)	.056	.625	.248	5.5					

*Anisotropic temperature factor of the form: $\exp\left[-(h^2B_{11}+k^2B_{22}+l^2B_{33}+2hkB_{12}+2hlB_{13}+2klB_{23})\right] \times 10^4$

Atom	Occupancy
C(21)	0.38(5)
C(22)	0.62(5)
C(31)	0.55(5)
C(32)	0.45(5)

All others: 1.0

Table 13

Crystal Structures Containing the Proline Residue

Abbreviation	Structure	Reference(s)
CuPro	Copper Proline dihydrate	44
PdPro	Bis-(L-prolinato) palladium(II)	45
HyPro	Hydroxy-L-proline	46[1], 47[2]
Pro	DL·proline hydrochloride	48
APNMA	Acetyl-L-proline-N-methylamide	42
NAP	N-acetyl-L-proline monohydrate	
LeuProGly	L-leucyl-L-prolyl-glycine	43
CBZ GPLG	p-Bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycine	49
CBZ GPLGP	o-Bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline ethyl acetate monohydrate	50
ProHyPro	Tosyl-L-proline-L-hydroxy proline monohydrate	51,52[3]
	L-proline	53[4]

[1] three-dimensional data

[2] two-dimensional data

[3] full-matrix refinement of the data from (51).

[4] although this structure will not be directly discussed, this reference is included to complete the list of known proline-containing crystal structures to date.

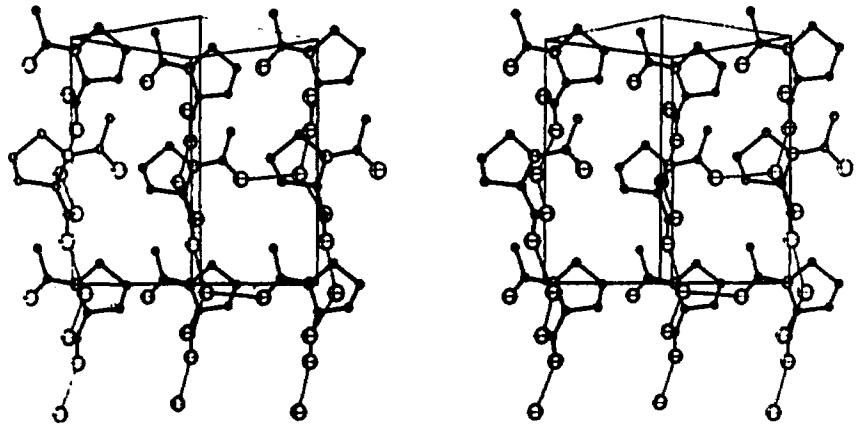


Figure 8. Stereodiagram of the Unit Cell, NAP

DISCUSSION OF THE STRUCTURE

Differences in the results of the two refinements are generally what one would expect. Bond angles and distances in the portion of the molecule unaffected by the splitting of C^{β}, C^{γ} are identical within the limits of error. Pertinent measures of the "goodness" of refinement are given below:

Result for	C^{γ} disordered	C^{β}, C^{γ} disordered
R	0.084	0.083
$\sum w_{hkl} \cdot \Delta^2$	1379.	1406.
$\sum w \cdot \Delta^2$ / (NO-NV)	1.02	1.04

(where NO = no. of data; NV = no. of variables).

While these results may seem to indicate that only C^{γ} should be treated as disordered, the magnitude of the anisotropic thermal ellipsoid (Table 14) for C^{β} (i.e. atom C(2)) is so large as to be physically meaningless. Isotropic thermal parameters resulting from refinement for C^{β} and C^{γ} disordered are reasonable (Table 14), indicating that C^{β} is disordered in the data crystal. Inspection of bond angles and distances (Figures 9 and 10) indicates no especially favored geometrical arrangements involving the disordered C^{β}, C^{γ} atoms. Although the results are not as definite as desired, it appears that C^{γ} is definitely disordered, and that C^{β} should be treated similarly.

The separation of the disordered C^{γ} half-atoms is similar to that observed in leucyl-L-prolyl-glycine (43) and the sites are equally occupied within experimental error. When C^{β} is refined as two disordered

Table 14

Principal Axes and Direction Cosines of Anisotropic Thermal Ellipsoids, NAP. C Disordered (B, l_i) and C , C Disordered (B', l'_i).

Atom	B	l_1	l_2	l_3	B'	l'_1	l'_2	l'_3
C(1)	5.62	.875	.072	.170	5.51	.872	.081	.176
	3.22	-.316	-.664	.743	3.19	.240	.790	-.612
	3.08	-.366	.744	.647	3.16	-.427	.608	.771
C(2)	16.54	.721	.081	.419				
	5.72	.154	.950	-.307				
	2.89	-.676	.302	.854				
C(4)	5.96	.471	.718	-.637	5.75	.471	.712	-.645
	4.70	.590	.177	.556	4.74	.591	.184	.553
	2.48	-.657	.673	.554	2.46	-.655	.678	.527
C(5)	4.68	.224	-.082	.847	4.72	.229	-.084	.844
	3.10	-.359	.919	.268	3.01	-.031	.995	.098
	2.45	-.906	-.385	.459	2.61	-.973	-.051	.527
C(6)	4.50	.351	-.163	.759	4.46	.314	-.161	.784
	3.46	-.934	-.132	.616	3.35	-.947	-.123	.586
	3.02	-.068	.978	.210	3.02	-.067	.979	.203
C(7)	8.76	.644	-.004	.516	8.45	.601	.014	.562
	5.29	-.757	.138	.849	5.15	-.794	.125	.820
	2.93	.108	.990	-.116	3.07	.091	.992	-.111
N(1)	4.26	.471	-.056	.681	4.22	.489	-.066	.666
	3.58	-.717	-.605	.558	3.54	-.732	-.573	.584
	2.25	-.513	.794	.473	2.27	-.474	.817	.464
O(1)	4.90	.879	.219	.118	4.90	.889	.199	.104
	4.39	-.459	.141	.978	4.39	-.443	.150	.979
	2.99	-.132	.966	-.170	2.95	-.114	.968	-.173
O(2)	6.79	-.657	-.230	.891	6.70	-.653	-.228	.894
	3.68	.675	.243	.441	3.65	.606	.414	.447
	3.36	-.335	.942	.104	3.39	-.454	.881	.022
O(3)	6.91	.470	.199	.662	7.10	.494	.183	.645
	4.17	.432	.798	-.537	4.12	.411	.812	-.524
	2.96	-.770	.569	.522	2.88	-.766	.553	.556
O(4)	6.37	.383	-.376	.675	6.50	.419	-.377	.646
	4.14	-.897	-.370	.518	4.11	-.891	-.345	.566
	3.03	-.221	.849	.525	2.97	-.174	.859	.511

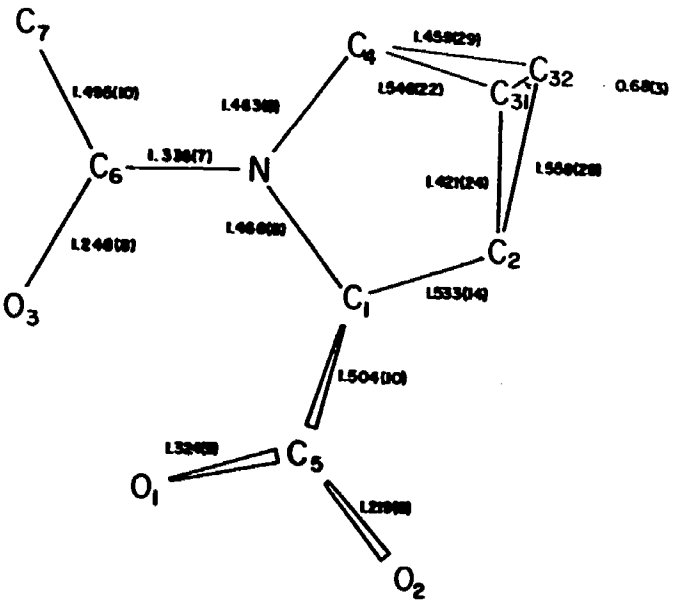


Figure 9a. Bond Distances, C^{γ} Disordered, NAP.

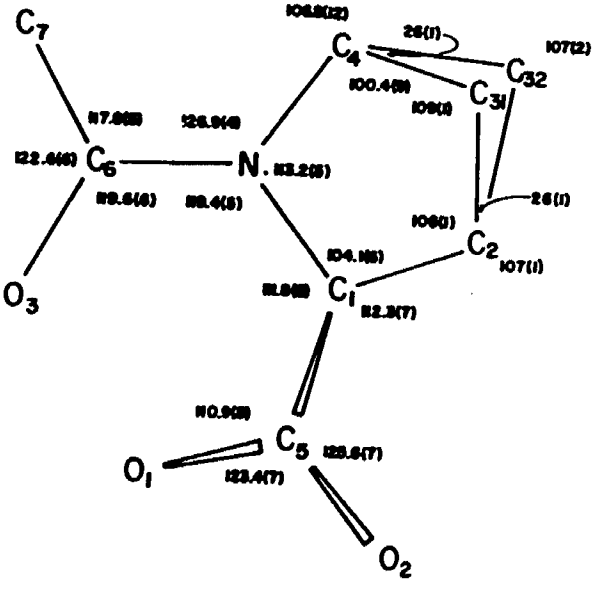


Figure 9b. Bond Angles, C^{γ} Disordered, NAP.

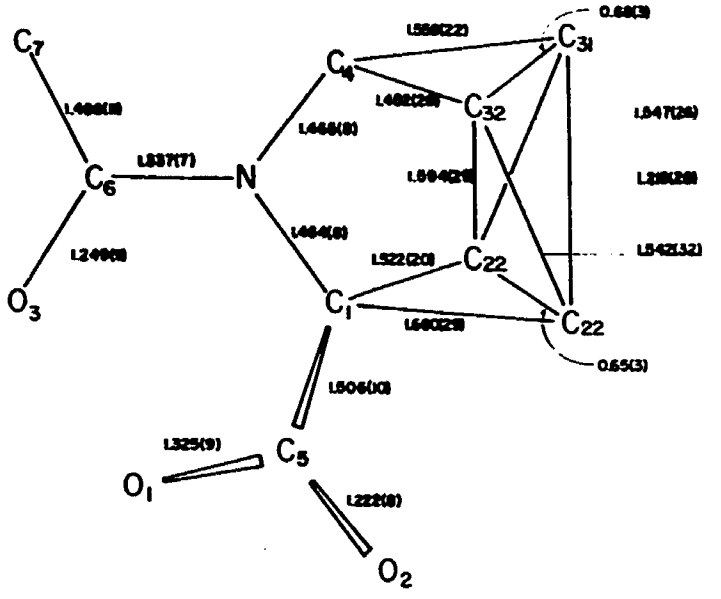


Figure 10a. Bond Distances, C^{β}, C^{γ} Disordered, NAP.

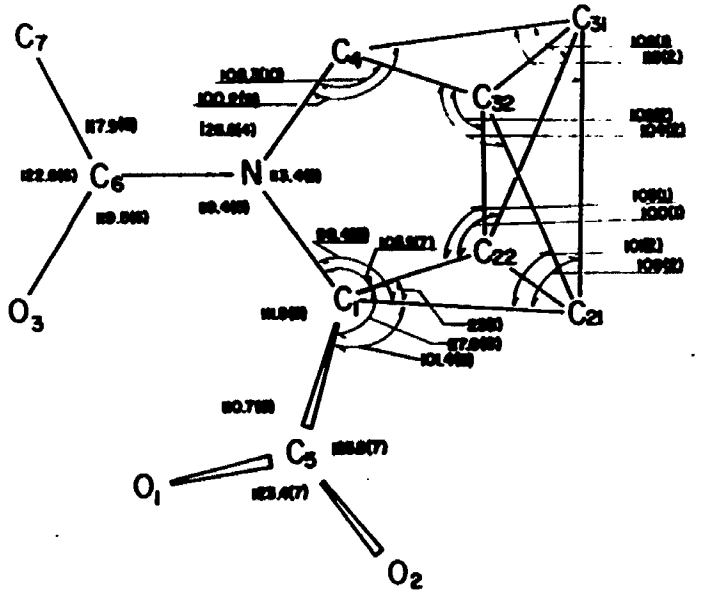


Figure 10b. Bond Angles, C^{β}, C^{γ} Disordered, NAP.

atoms, the two fractional atoms also separate 0.6 - 0.7 Å, but with occupancy of site C(22) more probable by two-to-one than site C(21). Projections of the molecule across the approximate plane of the pyrrolidine ring have been drawn. Figures 11a and 12a are drawn looking down the C^{α}-N bond.

Estimated positional deviations are about 0.005 Å for all atoms except C^{β} and C^{γ} - the disordered atoms - and $C^{\alpha'}$ - the acetyl carbon atom. For these, errors are estimated to be 0.01 - 0.03 Å. Bond distances and angles (Figures 9 and 10) are similar to those found for other peptide-bonded proline residues (Table 15). Excepting the C^{β}, and C^{γ} atoms, the corresponding structural parameters of NAP and APNMA (N-methylamidified NAP) are identical but for the apparent distortion of the carboxyl group, which will be discussed later. The disorder and resulting positional errors of C^{β}, and C^{γ} effect correspondingly the bond distances and angles involving these atoms. As a result, comparison of these parameters with those found in other proline-containing structures is of doubtful significance.

The equations for, and deviations from, several planes in NAP are shown in Tables 16 and 17. The acetyl group (Planes I, II) is flat within experimental error. Deviation of the nitrogen from the plane of the surrounding atoms (Planes III, IV) is about three times the expected positional errors, therefore of significance. The deviation is in a direction towards the carboxyl group, and is apparent in the projection of the molecule along the C^{α}-N bond (Figures 11a and 12a). Deviation of the carboxyl

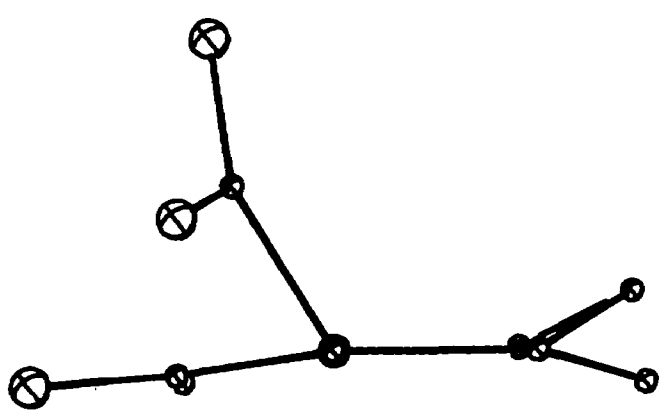


Figure 11a. Projection of NAP (C^{γ} disordered) along the C^{α}-N bond.

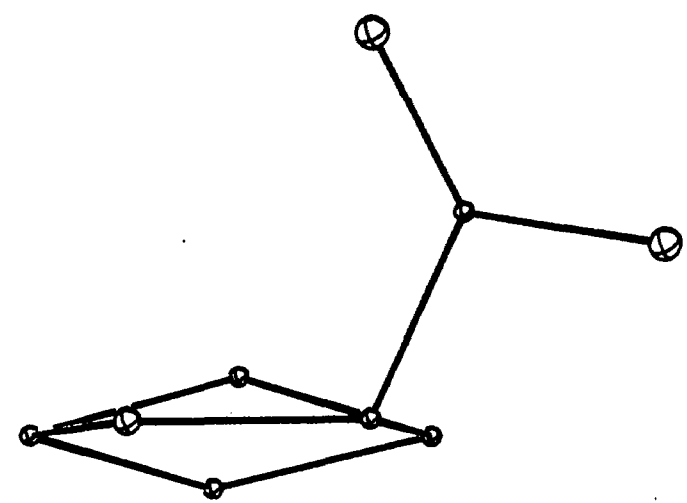


Figure 11b. Projection of NAP (C^{γ} disordered) perpendicular to the C^{α}-N bond, approximately in the plane of the pyrrolidine ring.

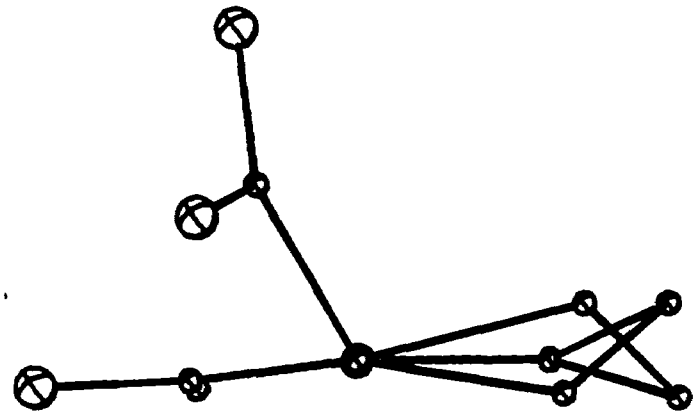


Figure 12a. Projection of NAP (C^{β}, C^{γ} disordered) along the C^{α}-N bond.

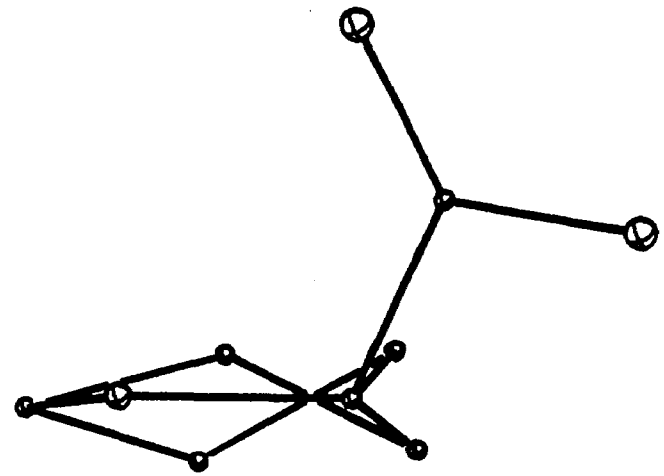


Figure 12b. Projection of NAP (C^{β}, C^{γ} disordered) perpendicular to the C^{α}-N bond, approximately in the plane of the pyrrolidine ring.

Table 15

Tabulation of Published Bond Angles and Distances for Proline Residues in Amino Acid-Peptide Structures

Cmpd.	CuPro	PdPro	HyPro	Pro	APNMA	NAP
C' - X'					1.490 Å	1.496 Å
C' - O'					1.245	1.249
C' - N					1.337	1.337
C^{α} - C^{β}	1.52 Å	1.48 Å	1.53 Å	1.54 Å	1.530	1.52, 1.66
C^{β} - C^{γ}	1.50	1.54	1.50	1.51	1.503	1.54 - 1.59
C^{γ} - C^{δ}	1.52	1.52	1.52	1.50	1.530	1.45, 1.55
C^{δ} - N	1.53	1.49	1.48	1.52	1.476	1.465
N - C^{α}	1.52	1.52	1.50	1.47	1.472	1.464
C^{α} - C"	1.50	1.52	1.52	1.52	1.530	1.506
C" - O"	1.24	1.19	1.25	1.24	1.231	1.222
C" - X"	1.24	1.33	1.27	1.32	1.316	1.325
X' C' N'					117.0°	117.9°
X' C' O'					122.9	122.6
O' C' N					120.2	119.5
C' N C^{α}	112°	108.5°			121.4	119.4
C' N C^{δ}	113	114.8			125.6	126.8
C^{δ} N C^{α}	108	105.0	109	104.6	112.2	113.4
N C^{α} C^{β}	108	106.2	105	106.8	103.4	99 - 106
C^{α} C^{β} C^{γ}	97	106.3	108	100.8	104.7	100 - 109
C^{β} C^{γ} C^{δ}	109	104.9	104	109.2	104.2	104 - 116
C^{γ} C^{δ} N	96	102.6	105	106.0	102.8	100 - 106
C" C^{α} N_{β}	108	110.2	111	111.3	114.3	111.8
C" C^{α} C^{β}	112	114.7	113	113.0	111.5	101 - 118
C^{α} C" O	118	120.8	119	122.7	117.6	125.8
C^{α} C" X"	120	116.8	115	112.9	117.9	110.7
O" C" X"	122	122.3	126	124.5	124.4	123.4
X'					C_1	C_7
X"	0	0	O_2	O_2	N_2	O_1
σ(distance)	.04 Å	.02 Å	.01 Å	.01 Å	.004 Å	.01 Å
σ(angles)	5°	1°	1°	0.8°	0.3°	0.7°

Table 15 Continued

Cmpd.	LeuProGly	CBZ GPLG	CBZ GPLGP		ProHyPro	
C' - X'	1.50 Å	1.51 Å	1.47 Å	1.52 Å		1.52 Å
C' - O'	1.27	1.26	1.26	1.27		1.22
C' - N	1.34	1.37	1.33	1.32		1.33
C^{α} - C^{β}	1.50	1.55	1.62	1.54	1.54 Å	1.54
C^{β} - C^{γ}	1.51	1.41	1.54	1.46	1.50	1.55
C^{γ} - C^{δ}	1.50	1.49	1.54	1.58	1.54	1.48
C^{δ} - N	1.46	1.46	1.46	1.55	1.47	1.47
N - C^{α}	1.45	1.46	1.50	1.46	1.48	1.47
C^{α} - C"	1.52	1.49	1.57	1.51	1.52	1.50
C" - O"	1.24	1.22	1.22	1.26	1.22	1.20
C" - X"	1.31	1.36	1.37	1.12	1.33	1.33
X' C N'	119°	119°	118°	119°		116°
X' C' O'	119	122	123	122		123
O' C' N'	122	120	120	119		121
C' N C^{α}	121	122	112	125	117°	120
C' N C^{δ}	126	124	126	121	121	130
C^{δ} N C^{α}	113	114	114	114	112	110
N C^{α} C^{β}	104	102	103	103	102	103
C^{α} C^{β} C^{γ}	107	108	103	105	108	104
C^{β} C^{γ} C^{δ}	106	109	108	108	102	101
C^{γ} C^{δ} N	103	104	105	96	104	107
C" C N	111	117	113	110	111	110
C" C^{α} C^{β}	113	115	106	111	109	110
C^{α} C" O	121	122	119	122	123	122
C^{α} C" X"	115	116	116	111	116	118
O" C" X"	123	122	125	127	121	120
X	C	C	C	C_{22}		C_9
X	N	N	N	O_7	N_2	O_6(-H)
σ(distance)	.015 Å	.025 Å	.02 Å		.015 Å	
σ(angles)	1.0°	1.4°				

Table 16

Least-Squares Planes for NAP, C^{γ} Disordered

Equations of planes of the form

$$Ax + By + Cz = D,$$

where x, y, z are fractional coordinates and D is the distance of the plane from the cell origin, in Å. The method of Shomaker, Waser, Marsh and Bergman (16) was used to calculate the planes. The distances (Å) for atoms used to define the plane are underscored.

	Distances from Planes (Å)								
Atoms	I	II	III	IV	V	VI	VII	VIII	IX
C(1)	.0611	.0645	<u>.0126</u>	<u>0</u>	<u>-.0046</u>	<u>0</u>	<u>-.0043</u>	<u>0</u>	<u>.0084</u>
C(2)	.2379	.2411	.1339	.1229	<u>-1.4293</u>	<u>-1.4251</u>	<u>.0026</u>	.0213	.1433
C(31)	-.1520	-.1491	-.2747	-.2864	-1.7363	-1.7316	-.4016	-.3844	-.2580
C(32)	.5063	.5091	.3882	.3714	-1.5446	-1.5404	.2584	.2755	.4004
C(4)	.0977	.1004	<u>.0138</u>	<u>0</u>	-.4269	-.4221	<u>-.0028</u>	<u>0</u>	<u>.0231</u>
C(5)	-1.1774	-1.1738	-1.2075	-1.2200	<u>.0160</u>	.0214	-1.2048	-1.2029	-1.2179
C(6)	<u>.0094</u>	.0125	<u>.0160</u>	<u>0</u>	1.8277	1.8328	.1681	.1500	<u>.0021</u>
C(7)	<u>-.0028</u>	<u>0</u>	.0115	-.0061	2.3497	2.3551	.2257	.1989	<u>-.0012</u>
N	<u>-.0031</u>	<u>0</u>	<u>-.0424</u>	-.0566	.5333	.5382	<u>.0046</u>	<u>0</u>	<u>-.0458</u>
O(1)	-.9366	-.9327	-.9591	-.9705	<u>-.0051</u>	<u>0</u>	-.9826	-.9765	-.9740
O(2)	-2.2799	-2.2764	-2.3047	-2.3179	<u>-.0063</u>	<u>0</u>	-2.2707	-2.2734	-2.3155
O(3)	<u>-.0035</u>	<u>0</u>	.0376	.0215	2.5438	2.5489	.2325	.2096	<u>.0134</u>
O(4)	-2.8296	-2.8252	-2.8203	-2.8312	.0197	.0260	-2.8213	-2.8177	-2.8466
A	3.344	3.345	3.550	3.545	6.104	6.102	3.934	3.887	3.499
B	3.331	3.329	3.277	3.269	1.596	1.596	3.471	3.441	3.312
C	3.988	3.988	3.805	3.812	-4.224	-4.227	3.361	3.419	3.847
D	3.881	3.877	3.899	3.910	0.820	0.813	3.930	3.924	3.918

Table 17

Least-Squares Planes for NAP, C^{β}, C^{γ} Disordered

Equations of planes of the form

Ax + By + Cz = D,

where x, y, z are fractional coordinates and D is the distance of the plane from the cell origin, in Å. The method of Shomaker, Waser, Marsh and Bergman (16) was used to calculate the planes. The distances (Å) for atoms used to define the plane are underscored.

	Distances from Planes (Å)									
Atoms	I	II	III	IV	V	VI	VII	VIII	IX	X
C(1)	.0607	.0631	.0119	0	-.0040	0	.0736	-.0479	0	.0090
C(21)	-.1708	-.1685	-.2785	-.2889	-1.6178	-1.6138	-.0432	-.5926	-.3875	-.2657
C(22)	.4404	.4426	.3383	.3280	-1.3431	-1.3396	.5614	.0277	.2335	.3495
C(31)	-.1620	-.1600	-.2824	-.2934	-1.7492	-1.7452	-.0682	-.5725	-.3861	-.2645
C(32)	.4918	.4938	.3717	.3606	-1.5467	-1.4531	.5720	.0841	.2708	.3899
C(4)	.0932	.0951	.0129	0	-.4359	-.4317	.0530	-.0302	0	.0217
C(5)	-1.1828	-1.1803	-1.2149	-1.2266	.0139	.0186	-1.1606	-1.2264	-1.2109	-1.2234
C(6)	.0066	.0088	.0150	0	1.8220	1.8264	-.1857	.3375	.1412	-.0005
C(7)	-.0020	0	.0163	-.0001	2.3418	2.3465	-.2883	.4851	.1930	.0003
N	-.0022	0	-.0398	-.0532	.5279	.5321	-.0834	.0504	0	-.0436
O(1)	-.9417	-.9389	-.9682	-.9790	-.0044	0	-.8664	-1.0486	-.9851	-.9800
O(2)	-2.2865	-2.2841	-2.3124	-2.3248	-.0055	0	-2.2961	-2.2415	-2.2831	-2.3220
O(3)	-.0024	0	.0386	.0235	2.5411	2.5456	-.2164	.4482	.2004	.0132
O(4)	-2.8405	-2.8374	-2.8381	-2.8484	.0253	.0308	-2.7330	-2.8637	-2.8363	-2.8602
A	3.352	3.353	3.554	3.550	6.102	6.100	2.963	4.356	3.872	3.500
B	3.353	3.352	3.316	3.309	1.581	1.580	2.901	3.784	3.472	3.341
C	3.975	3.975	3.792	3.798	-4.232	-4.235	4.411	2.773	3.428	3.838
D	3.892	3.889	3.916	3.926	.815	.808	3.762	3.992	3.939	3.930

group from planarity is approximately the same as the estimated positional errors, therefore of questionable significance. Deviation from a plane by atoms of the peptide group is significant only for atoms N, C^{δ} . Deviations from pyrrolidine ring planarity compares closely with those observed in LeuProGly (45) - (Plane VII, Table 16; Table 18) if only CY is considered as disordered.

Peptide conformational angles (see Figure 7) for NAP are: $\phi = 120.0^{o}$, $\psi = -24.9^{o}$, 157.8^{o}. Two values for ψ are due to the lack of amide nitrogen attached to C" , hence with coordinates (ϕ,ψ) attachment is equally probable at the location of either O(1) or O(2). For peptide linkages, the interrelation of ϕ and ψ may be described using a two-dimensional Ramachandran plot (Figure 13). The coordinates for NAP are shown with those of some other proline-containing structures. There are two important conclusions which may be drawn by inspection of this conformational map. (1) NAP has approximately the same conformation as other proline structures. Since ϕ is restricted to values near 120^{o}, this indicates that rotation about the C^{α} - C' bond (4) may also be restricted, in Pro, for reasons which are not clear at this point. (2) The conformation of NAP is equally suitable for inclusion in the collagen structure (120^{o}, -25^{o}) as in, say, an α- helical structure (120^{o}, 160^{o}) but does not display suitable conformation for inclusion into β- sheet structures. The 'helix-breaker' reputation of Pro appears dependent on its inability to participate in hydrogen-bonding stabilizing interactions.

Table 18

Comparison of Deviations From Planarity of Some Pyrrolidine Rings

Structure	Ring Plane Deviations						
	N	C^{α}	C^{β_1}	C^{β_2}	C^{γ_1}	C^{γ_2}	C^{δ}
HyPro	<.03*	<.03	<.03		≃ 0.4		<.03
CuPro	<.06	<.06	<.06		0.60		<.06
LeuProGly	<.06	<.06	<.06		.44	-.29	<.06
Tos Pro HyPro	•	•	•			.51	
Tos Pro HyPro	•	•	•			.54	
CBZGlyProLeuGly	-.020	.036	-.035		.257		.021
DL-Pro	<.02	≃.05	<.02		<.02		<.02
APNMA	.043	-.041	.025		-.527		-.027
CBZGlyProLeuGlyPro	.096	-.053	.014		-.407		-.050
CBZGlyProLeuGlyPro	.084	-.079	.048		-.493		-.049
PdPro	.04	-.03	.04		-.04		-.56
NAP; C , C disordered	-.083	.074	-.043	-.561	-.068	.572	.053
NAP; C , C disordered, alternate configuration	.050	-.048	.593	.028	-.572	.084	-.030

*Mean plane defined by atoms whose deviations are underscored; where two residues are found in one structure; the one underscored is listed.

• Atoms described simply as coplanar.

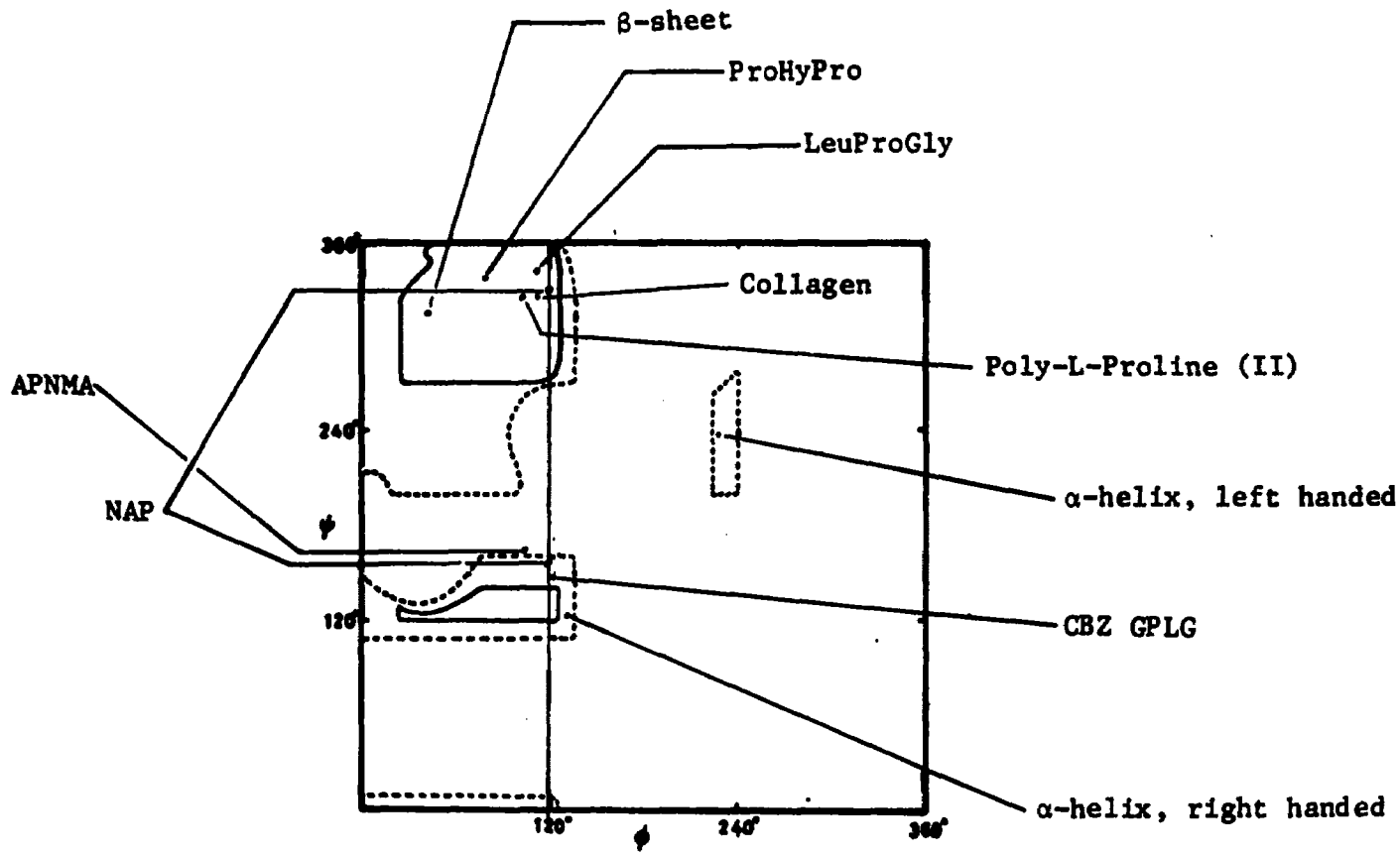


Figure 13. Ramachandran Plot for NAP and Some Related Compounds. —bounds fully allowed regions, ···· bounds partially allowed regions. Adapted from (42).

It is anticipated that three protons will be involved in a hydrogen-bonding scheme. They are the proton from the carboxyl group, and the two from the water molecule. As mentioned previously, these atoms could not be located with certainty. A hydrogen-bonding scheme consistent with expectations and with calculated intermolecular distances (Table 19) may be described. The distance O(4) - O(3)b is similar to hydrogen bond distances observed in other Pro-containing structures; since O(3) is a carbonyl oxygen, it is reasonable that the proton should be donated by the water molecule, O(4). The distance O(4) - O(2)c is also similar to previous observations; the origin of the proton assumed to participate in the hydrogen-bond is not certain. The C^{α} - O(2) bond has more double bond character than the C^{α} - O(1) bond; for this reason it is expected that O(2) shares the second proton from the water molecule. The other short contact - O(4) - O(1)a - is significantly shorter than observed in other Pro-containing structures (Table 20), indicating a very strong hydrogen-bond - that the C^{α} - O(1) bond length is longer than the C^{α} - O(2) bond is an indication that O(1) donates its proton to the water molecule.

Other intermolecular contacts less than 3.5 Å are also listed in Table 20. The distance O(1) - N(1)d is the shortest contact distance, with O(1) directly above the nitrogen atom (perpendicular to the plane of pyrrolidine ring). This distance is greater than the sum of the van der Waal radii for these atoms (54). Nevertheless, distortion of the carboxyl group is in a manner that O(1) appears to be pulled from a location behind the ring - which is just the location of N(1)d. Excepting the C^{β}, C^{γ} disorder,

Table 19

Intermolecular Distances Less Than 3.5 Å, NAP

Atoms	Sym.	For C Disordered	For C , C Disordered	
O(4) - O(1)	a	2.589 Å	2.591 Å	(H)
O(4) - O(3)	b	2.695	2.692	(H)
O(4) - O(2)	c	2.858	2.859	(H)
O(4) - C(5)	a	3.450	3.449	
O(1) - C(4)	d	3.448	3.448	
O(1) - C(6)	d	3.474	3.476	
O(1) - N(1)	d	3.376	3.373	
O(3) - C(31)	e	3.423	3.418	

The small letter in the column headed "Sym." indicates which of the following symmetry operations should be applied to the coordinates, as shown in Table 12, of the second atom:

(a)	x	y	z
(b)	1 - x	$-\frac{1}{2}$ + y	- z
(c)	- x	$-\frac{1}{2}$ + y	- z
(d)	1 - x	$-\frac{1}{2}$ + y	1 - z
(e)	1 + x	y	z

The letter (H) indicates hydrogen bonds.

Table 20

Comparison of Hydrogen Bond Distances*

Structure	Atoms	Dist.	Atoms	Dist.	Atoms	Dist.	Atoms	Dist.
HyPro	O(3)-O(1)	2.80 Å	N-O(2)	2.69 Å	N-O(2)	3.17 Å		
CuPro	N(A)-O_2(E)	2.86	O_3'(A)-O_1(E)	2.94	O_3'(A)-O_2'(B)	3.00		
LPG	O(4)-N(1)	3.18	N(3)-O(2)	2.84	N(3)-O(2)	2.87	O(3)-O(1)	2.83 Å
	O(5)-O(1)	2.81	O(1)-O(2)	2.94				
TosProHyPro	O(7)-O(3)	2.81	O(7)-O(5)	2.93	O(7)-O(4)	2.78	O(5)-O(6)	2.66
GPLG	O(5)-N(1)	2.82	O(3)-N(4)	2.97	O(7)-N(3)	2.99		
DL-Pro	O(2)-Cl	2.96	N-Cl	3.14	NH-Cl	3.18		
APNMA	O(1)-N(2)	2.88						
CBZGPLGP	Gly(1)N - Gly(2)O			3.03	Gly(1)O - Gly(2)N			3.00
PdPro	N-O(2)	2.85						
NAP	O(4)-O(1)	2.59	O(4)-O(3)	2.69	O(4)-O(2)	2.86		

*The numbering given in the original report has been retained.

this distortion of the carboxyl group is the only discrepancy between the structural parameters of NAP and APNMA. For this reason, the distortion appears significant, and may indicate a weak interaction between O(1) and N(1)d.

CONCLUSIONS

In the crystal structure of N-acetyl-L-proline several features of interest have been observed.

Although the refinement is less than satisfactory, final difference Fouriers show no recognizable remnant structure. Analysis of multiple intensity-data sets indicates a limiting accuracy much better than obtained; photographic analysis of the data crystal shows no x-ray-detectable disorder. The only conclusion possible at present is that there is a sufficient amount of low-level disorder in the structure to make further refinement impractical.

The C^{γ} atom is disordered, the second occasion of such a refinement (43), although mentioned in connection with another structure (48). C^{β} is also disordered, or has very high thermal motion (the former appears more reasonable). Consideration of the calculated bond distances and angles indicates that the disorder may be concerted, i.e. that the C^{β}, C^{γ} atoms occupy positions on opposite sides of the ring in any one molecule (see Figure 14). This interpretation is consistent with other studies of the flexibility of cyclopentane rings (55, 56), and with crystal-structure analysis involving the cyclopentene ring (57). The importance of this possible disorder in calculating protein tertiary structure was noted earlier (22). The rigidly planar nature of the peptide group is well-known (12); one may hypothesize that this rigidity stabilizes the C^{α} and C^{δ} atoms of the proline pyrrolidine ring with respect to high

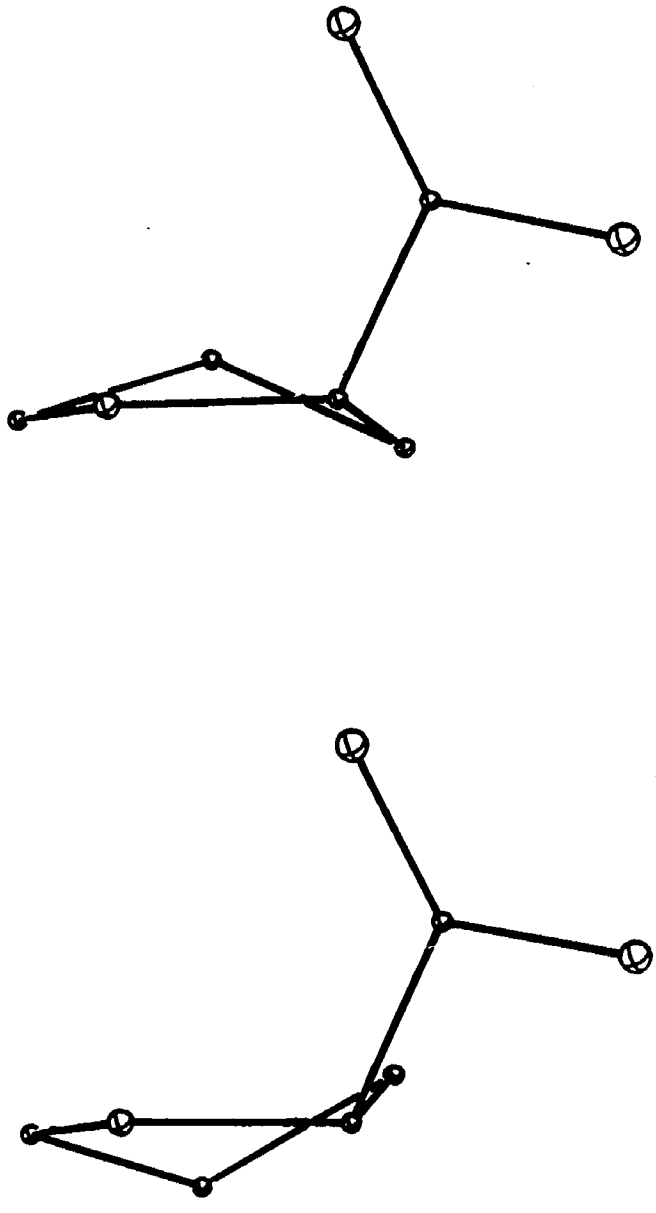


Figure 14. Suggested Conformations of the Pyrrolidine Ring, NAP. The view is the same as that of Figure 12b.

thermal motion/disorder. It might be anticipated that future observations of high thermal motion/disorder in this system would occur mainly for the C^{α} and/or C^{δ} atoms in peptide-bonded residues.

The inclusion of a proline residue in an α-helix appears to be disruptive not so much by geometrical considerations, as by the elimination of hydrogen-bond stabilization.

A very short hydrogen bond has been observed, shorter than in previously reported proline-containing structures. Although not unequivocal, a possible very weak interaction between the amide nitrogen and one of the carboxyl oxygens has been observed. The location of this interaction is near the nitrogen p_z non-bonding orbital, and may be significant in understanding the tertiary structure of proteins.

LIST OF REFERENCES

1. D. van der Helm and W. A. Franks, J. Amer. Chem. Soc., 90, 5627 (1968).

2. W. A. Franks and D. van der Helm, Acta Cryst., B27, 1299 (1971).

3. D. van der Helm, M. B. Lawson and E. L. Enwall, Acta Cryst., B27, 2411 (1971).

4. P. Coppens, L. Leiserowitz and D. Rabinovich, Acta Cryst., 18, 1035 (1965).

5. D. T. Cromer and D. Lieberman, Private Communication (1969).

6. "International Tables for X-Ray Crystallography," Vol. III, Kynoch Press, Birmingham, 1962, p. 202.

7. R. F. Stewart, E. R. Davidson and W. T. Simpson, J. Chem. Phys., 42, 3175 (1965).

8. D. van der Helm and H. B. Nicholas, Acta Cryst., B26, 1858 (1970).

9. H. C. Freeman, Adv. Protein Chem., 22, 257 (1967).

10. D. C. Fries and M. Sundaralingam, Acta Cryst., B27, 401 (1971).

11. J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran and H. A. Scheraga, Biopolymers, 4, 121 (1966).

12. G. N. Ramachandran and V. Sasisekharan, Adv. Protein Chem., 23, 283 (1968).

13. A. F. Pieret, F. Durant, M. Griffe, G. Germain and T. Debaerdemaeker, Acta Cryst., B26, 2117 (1970).

14. J. W. Becker, V. T. Thathachari and P. G. Simpson, Biochem. Biophys. Res. Commun., 41, 444 (1970).

15. G. Gurskaya, Kristallografiya, 9, 839 (1964).

16. V. Schomaker, J. Waser, R. E. Marsh and G. Bergman, Acta Cryst., 12, 600 (1959).

17. C. K. Johnson, ORTEP. Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1965.

18. L. Pauling, R. B. Corey and H. R. Branson, Proc. Natl. Acad. Sci. U.S., 37, 205 (1951).

19. M. F. Perutz, Nature, 167, 1053 (1951).

20. D. C. North, Scientific American, 216, 78 (1966).

21. M. J. Perutz, Eur. J. Biochem., 8, 455 (1969).

22. H. A. Scheraga, Chem. Revs., 71, 195 (1971).

23. F. A. Bovey, A. I. Brewster, D. J. Patel, A. E. Tonelli and D. A. Torchia, Accts. Chem. Res., 5, 193 (1972).

24. C. H. Hassall, W. A. Thomas, Chem. Brit., 7, 145 (1971).

25. R. E. Dickerson and I. Geis, "The Structure and Action of Proteins," Harper and Row, New York, N.Y., 1969, p. 21.

26. G. N. Ramachandran, in "Aspects of Protein Structure," G. N. Ramachandran, ed., Academic Press, New York, N.Y., 1963, p. 39ff.

27. N. Gō, P. N. Lewis and H. A. Scheraga, Macromolecules, 3, 628 (1970).

28. P. R. Schimmel and P. J. Flory, Proc. Nat. Acad. Sci. U.S., 58, 52 (1967).

29. N. Gō and H. A. Scheraga, Macromolecules, 3, 188 (1970).

30. S. J. Leach, G. Nemethy and H. A. Scheraga, Biopolymers, 4, 369 (1966).

31. "International Tables for X-Ray Crystallography," Vol. I, Kynoch Press, Birmingham, 1962.

32. H. B. Nicholas, Ph.D. Dissertation, The University of Oklahoma, Norman, Oklahoma, 1969.

33. G. H. Stout and L. H. Jensen, "X-Ray Structure Determination," The MacMillan Company, Collier-MacMillan Limited, London, 1968, p. 454ff.

34. P. Coppens, in "Crystallographic Computing," F. R. Ahmed, ed., Munksgaard, Copenhagen, 1970, p. 255.

35. W. R. Busing and H. A. Levy, J. Chem. Phys., 26, 563 (1957).

36. P. W. R. Corfield, R. J. Doedens and J. A. Ibers, Inorg. Chem., 6, 197 (1967).

37. D. W. J. Cruickshank, in "Crystallographic Computing," F. R. Ahmed, ed., Munksgaard, Copenhagen, 1970, p. 197.

38. F. R. Ahmed, NRC Crystallographic Programs for IBM/360 System, Program Nos. NRC-5-I, -II, -III, -IV, -V (Direct Phasing Methods).

39. J. Karle and I. L. Karle, Acta Cryst., 21, 849 (1966).

40. H. Hauptman and J. Karle, "ACA Monograph No. 3," Edwards Brothers, Inc., Ann Arbor, Michigan, 1953.

41. G. Germain and M. W. Woolfson, Acta Cryst., B24, 91 (1968).

42. T. Matsuzaki and Y. Iitaka, Acta Cryst., B27, 507 (1971).

43. Y. C. Leung and R. E. Marsh, Acta Cryst., 11, 17 (1958).

44. A. McL. Mathieson and H. K. Welsh, Acta Cryst., 5, 599 (1952).

45. T. Ito, F. Marumo and Y. Saito, Acta Cryst., B27, 1062 (1971).

46. J. Donohue and K. N. Trueblood, Acta Cryst., 5, 419 (1952).

47. J. Zussman, Acta Cryst., 4, 493 (1951).

48. Y. Mitsui, M. Tsuboi and Y. Iitaka, Acta Cryst., B25, 2182 (1969).

49. T. Ueki, T. Ashida, M. Kakudo, Y. Sasada and Y. Katsube, Acta Cryst., B25, 1840 (1969).

50. T. Ueki, S. Bando, T. Ashida and M. Kakudo, Acta Cryst., B27, 2219 (1971).

51. J. Fridrichsons and A. McL. Mathieson, Acta Cryst., 15, 569 (1962).

52. M. N. Sabesan and K. Venkatesan, Acta Cryst., B27, 1879 (1971).

53. R. L. Kayushina and B. K. Vainshtein, Kristallografiya, 10, 833 (1965).

54. L. Pauling, "The Nature of the Chemical Bond," 3rd ed., Cornell University Press, 1960, p. 260.

55. K. S. Pitzer and W. E. Donath, J. Amer. Chem. Soc., 81, 914 (1959).

56. J. B. Hendrickson, J. Amer. Chem. Soc., 83, 4537 (1963).

57. H. Shenk, Acta Cryst., B27, 667 (1971).