

The Crystal Structure of Hydroxy-L-Proline. II. Determination and Description of the Structure*

BY JERRY DONOHUE

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California, U.S.A.

AND KENNETH N. TRUEBLOOD

Department of Chemistry, University of California at Los Angeles, Los Angeles, California, U.S.A.

(Received 20 November 1951)

The determination of the crystal structure of hydroxy-L-proline, based on nearly complete data accessible with Cu $K\alpha$ radiation, confirms the chemical assignment of the relative configurations about the two asymmetric carbon atoms, and leads to the following intramolecular bond distances: carboxyl group: $C_1-O_1=1.25$, $C_1-O_2=1.27$, $C_1-C_2=1.52$ Å; pyrrolidine ring: $C_2-C_3=1.53$, $C_3-C_4=1.50$, $C_4-C_5=1.52$, $C_5-N=1.48$, $N-C_2=1.50$ Å; hydroxyl group: $C_4-O_3=1.46$ Å. The crystal is held together by a system of hydrogen bonds, with $N \dots O_2=2.69$, $N \dots O'_2=3.17$, $O_3 \dots O_1=2.80$ Å. The molecule is in the *Zwitterion* form. These and various other structural features are discussed.

Introduction

The determination of the crystal structures of individual amino acids and their derivatives is part of a long-range program of investigation of the constitution and configuration of proteins at the California Institute of Technology. The structures of diketopiperazine (Corey, 1938), α -glycine (Albrecht & Corey, 1939), DL-alanine (Levy & Corey, 1941; Donohue, 1950), β -glycylglycine (Hughes & Moore, 1949), *N*-acetylglycine (Carpenter & Donohue, 1950), and L-threonine (Shoemaker, Donohue, Schomaker & Corey, 1950) have already been reported; the structures of serine (Shoemaker, Barieau, Donohue & Lu, private communication), α -glycylglycine (Hughes, Biswas & Wilson, private communication), and *N, N'*-diglycylcystine (Hughes & Yakel, private communication) have also been determined. Much of the information about molecular geometry and forces provided thereby has been applied to considerations of the structures and folding of protein chains (Corey & Donohue, 1950; Pauling & Corey, 1950; Pauling, Corey & Branson,

1951; Pauling & Corey, 1951). The present study of hydroxy-L-proline is a part of this program. The structure of this amino acid (Zussman, 1951)[†] is of special interest for several reasons. First, it has been pointed out (Pauling, 1940) that hydroxyproline and proline may be of particular significance in the architecture of protein chains since it appears from models that they interfere with parallel configuration of these chains and thus tend to force them to fold back upon themselves. Secondly, the relative configurations of the carbon which carries the hydroxyl group and the α -carbon have not been established with complete certainty by chemical methods (Neuberger, 1948); an unambiguous determination of this configurational relationship would be desirable since it provides a means of relating the configurations of the sugars and the amino acids.

Experimental

Unit cell and space group

Crystals of hydroxy-L-proline (Mann Fine Chemicals Co.) were grown from 95% ethanol. Goniometric examination indicated that they were orthorhombic and the Laue symmetry (D_{2h}) confirmed this. Accurate axial lengths were obtained from a large number of indexed equatorial reflections on rotation photographs taken about each of the axes with a 5 cm. camera, and Ni-filtered Cu $K\alpha$ radiation; based on a value $\lambda = 1.542$ Å for Cu $K\alpha$, the unit-cell dimensions were found to be

$$a_0 = 5.00, b_0 = 8.31, c_0 = 14.20 \text{ \AA},$$

with a maximum error of about 0.2%. Zussman (1951) reports $a_0 = 5.01$, $b_0 = 8.35$, $c_0 = 14.1$ Å, all $\pm 1\%$. The axial ratios calculated from our X-ray data are 0.6017:1:1.7088; by goniometric examination of a

* Contribution No. 1627 from the Gates and Crellin Laboratories of Chemistry, California Institute of Technology, and from the Department of Chemistry, University of California at Los Angeles. A preliminary report was presented at the Second Congress of the International Union of Crystallography at Stockholm in June 1951, and a more detailed description was given at the XIIth International Congress of Pure and Applied Chemistry at New York, September 1951.

Aided by a grant from the National Foundation for Infantile Paralysis and from the Lederle Laboratories.

[†] Zussman has made a brief report of a determination of the crystal structure of hydroxy-L-proline from two projections alone, with results which are qualitatively the same as ours. In a private communication, Zussman informs us that he has refined his parameters by additional considerations of the ($0kl$) and ($h0l$) data. We intend to make a quantitative comparison between the results of the two independent studies in a later communication.

number of crystals we obtained average values of 0.601:1.1711, with appreciable individual variation. Von Wolff (Fisher, 1902) reported values which correspond in our axial system to 0.6025:1.16833 based on goniometric examination of tablets of hydroxyproline grown from water; Zussman's X-ray data lead to the ratios 0.600:1.169.

The density of several representative crystals was determined to be 1.474 ± 0.003 g.cm.⁻³; there are thus four (calculated 3.99) molecules in the unit cell. Equatorial Weissenberg photographs taken about **a** and **b** showed the absence of all (*h*00), (0*k*0), and (00*l*) reflections with odd values of *h*, *k* and *l* within the respective ranges of observation, 0 to 6, 0 to 10, and 0 to 18. No other systematic extinctions were observed; consequently it may be assumed that the space group is $D_2^4-P2_12_12_1$, and that the four asymmetric molecules in the unit cell lie in general positions.

Collection of intensity data

Two crystals of hydroxyproline which had been dipped in liquid air in an effort to minimize extinction effects were mounted on glass fibers with flake shellac, one with **a** vertical, and one with **b** vertical. Each crystal was reduced to a cylinder, about 0.3 mm. and 0.4 mm. in diameter respectively, with the aid of a narrow strip of filter paper wet first with 90% ethanol and then, in the final stages, with absolute ethanol. Because of this precaution no absorption corrections were needed. Weissenberg photographs were then taken of each crystal for all layer lines with equi-inclination angles less than 30° (*h* = 0 to 3; *k* = 0 to 5). Each exposure was taken with three sheets of Eastman No-Screen film in the camera; the films in each set were developed uniformly and the relative intensities of the spots were measured visually with the aid of the multiple-film technique and intensity strips prepared from a single reflection exposed under standard conditions for known and varying periods of time. For perpendicular incidence on the film, the film factor used (Shoemaker *et al.*, 1950) was 3.7; for non-equatorial layer lines higher film factors were used, since there is greater absorption because of the greater path length of the X-rays through the film.

The intensities were corrected for Lorentz and polarization factors to obtain relative values of $|F_{hkl}|^2$ and the values obtained from each set of films were multiplied by an appropriate scale factor to bring them to the same arbitrary scale. Independent intensity estimates of the same reflections on different films by a given observer showed an average deviation from the mean of about 8%. Furthermore, estimates by different observers of the relative intensities of the spots on a given film generally agreed very well. Thus, the average deviation from the mean of the independent estimates (by each of the authors) of the intensities of the 0*kl* reflections, excluding four of the weakest, was 4.0%.

In all, 751 of the 778 general (*hkl*) reflections accessible to Cu *K*α radiation were in a position to reflect in our experiments; of these, 105 were not observed. The observed values of $|F_{hkl}|^2$ were placed on an absolute scale, and the temperature factor was determined by a method proposed by E. W. Hughes (Shoemaker *et al.*, 1950) and found to be $\exp(-2.23 \sin^2 \theta/\lambda^2)$.

Refinement of parameters

A successful trial structure was obtained by a relatively complete interpretation of a three-dimensional Patterson function. The details of this analysis are described elsewhere (Donohue & Trueblood, 1952). The atomic parameters derived from the Patterson function were subjected to a preliminary refinement by consideration of the structure factors of the (00*l*), (0*k*0) and (*h*00) reflections, and by adjustment of certain intramolecular distances to within about 10% of those expected. The parameters so obtained were then used in the calculation of the structure factors F_{0kl} . Comparison of these with the observed F_{0kl} enabled 59 of the 120 signs to be assigned with confidence. A Fourier projection of electron density on (100) prepared with these 59 terms showed five of the nine atoms clearly resolved. The *y* and *z* parameters were adjusted, the values of F_{0kl} were recalculated, and a second projection was prepared, with 29 additional terms. An attempt was then made to speed the refinement by the use of a method suggested by Booth (1947), in which the quantity $\sum_{F=0} F_{0kl}^2$ is minimized with respect to the parameters, the summation being taken over all planes of observed zero intensity.* These parameters were averaged with those obtained from the second projection on (100) and used in the calculation of a third set of F_{0kl} . These were then used in a least-squares refinement (Hughes, 1941) of the *y* and *z* parameters. A fourth set of F_{0kl} was then calculated with the parameters obtained from the least-squares procedure. A third projection on (100) was then made, with 108 of the observed F_{0kl} . The averages of the parameters from the Fourier projection with those from the least-squares were considered as good as could be obtained from the data of this one zone. The final projection on (100), made with signs determined from the final parameters (Table 2), is shown in Fig. 1.

The *z* parameters from the above procedures were combined with the *x* parameters as refined from their original Patterson values, and used in the calculation of a set of structure factors F_{h0l} . A Fourier projection on (010) was then made with all 75 of the observed (*h*0*l*) reflections. The same structure factors were also used in a least-squares refinement. The least squares

* This method proved in this case to be of no value, and, in fact, was detrimental, for of the eight parameter shifts greater than 0.10 Å which it indicated, four were of incorrect sign, as judged by the final parameters.

and Fourier projection parameters were averaged for use in the subsequent steps. The final projection on (010), made with signs determined from the final parameters, is shown in Fig. 2.

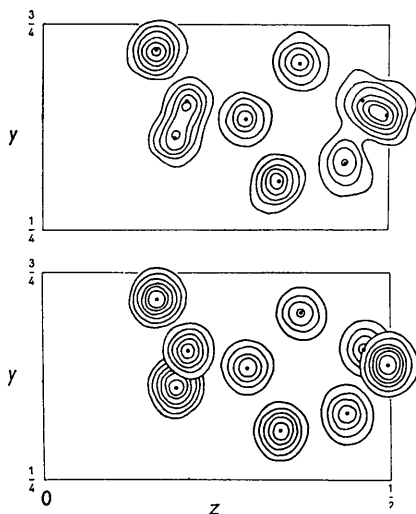


Fig. 1. Above: Fourier projection on (100); below: Fourier sections from $\rho_o(x, y, z)$. Dots indicate positions of atoms according to final parameters. Contours in the projection are drawn at equal intervals of an arbitrary scale: those in the sections are at 2, 4, 6 ... e.Å⁻³.

Average values of the parameters obtained from the refinement of the two zones were then used to calculate a complete set of structure factors $|F_{hkl}|$ (1). A three-dimensional least-squares treatment was next applied to give refined parameters, and the new set of calculated structure factors, $|F_{hkl}|$ (2), showed somewhat improved agreement with the observed $|F_{hkl}|$. In the three-dimensional least-squares procedures the off-diagonal elements of the normal equations were

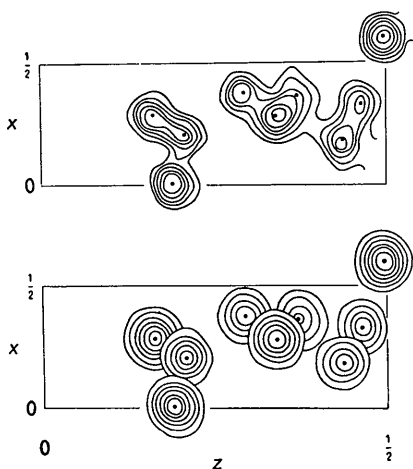


Fig. 2. Above: Fourier projection on (010); below: Fourier sections from $\rho_o(x, y, z)$. Dots indicate positions of atoms according to final parameters. Contours in projection are drawn at equal intervals of an arbitrary scale: those in the sections are at 2, 4, 6 ... e.Å⁻³.

neglected. A second least-squares treatment, on $|F_{hkl}|$ (2), showed that this method had essentially converged, as the average parameter shift was but 0.007 Å, as compared with the average shift of 0.017 Å indicated by the first least-squares treatment. These parameter shifts are small enough to indicate that not much greater improvement in the agreement would result from a set of $|F_{hkl}|$ calculated with the parameters from the second least-squares calculation. The agreement was, however, not as satisfactory as desired. Two procedures designed to improve the agreement were therefore applied: (1) The absolute scale factor was adjusted by minimizing the quantity $\sum_{hkl} w(k|F_o|^2 - |F_c|^2)^2$ with respect to k . (2) Positions for

the hydrogen atoms were deduced from ball-and-stick models, and were refined by consideration of accepted bond distances and angles. The contributions of these atoms to the structure factors of planes for which $\sin \theta < 0.6$ were then calculated. These were then combined with the set $|F_{hkl}|$ (2) to give $|F_{hkl}|$ (3). Although a few of the strongest reflections were noted to have much smaller structure factors than those calculated, a situation indicative of extinction, no correction for extinction could be made because all of the strongest lines are unresolved from at least one other strong line on powder photographs. It is interesting that liquid-air treatment did not eliminate extinction in these crystals.

A third and final least-squares treatment of the $|F_{hkl}|$ data was then carried out. As anticipated, neither correction (1) to the observed $|F_{hkl}|$ data nor the addition of the contributions of the hydrogen atoms had much effect in the least-squares procedure. Furthermore, the planes subject to extinction are the strongest ones, and therefore have very small weighting factors, w , in the least-squares calculation. The average difference between parameters from the second and third least-squares procedures was less than 0.005 Å.

Two three-dimensional electron-density distributions were then computed, one with the phases and amplitudes of $|F_{hkl}|$ (3), termed $\rho_c(x, y, z)$, and the other with the phases of $|F_{hkl}|$ (3) and the amplitudes of the observed $|F_{hkl}|$, termed $\rho_o(x, y, z)$. The intervals in x, y and z were the same as those in the corresponding directions of the Patterson functions (Donohue & Trueblood, 1952).

Two views of the electron density $\rho_o(x, y, z)$ are shown in Figs. 1 and 2. These drawings were made by taking planes parallel to (100) or (010) and passing near the maxima of the peaks. Contours were drawn at intervals of 2, 4, 6, ... e.Å⁻³, and the results were then projected down a_0 , or b_0 . The corresponding projections, made with (0kl) or (h0l) data alone, are shown for comparison.

Small humps in both $\rho_o(x, y, z)$ and $\rho_c(x, y, z)$ were noted at positions near those expected for the hydrogen atoms. The maxima in $\rho_o(x, y, z)$ were not suited for

Table I. *Observed and calculated structure factors*

In each vertical section, the left hand column is the l index, the middle column the observed structure factor, and the right hand column the calculated structure factor, $|F_{hkl}|$ (5) (see text). All structure factors have been multiplied by ten.

0	00 l	2800	0	04 l	72	92	0	08 l	< 36	36	15	82	73
2	140	158	1	97	124		1	< 39	27		16	67	58
4	99	94	2	59	46		2	< 36	18		17	12	18
6	360	621	3	146	116		3	< 38	56			12 l	
8	176	152	4	265	264		4	< 37	34		0	333	347
10	62	50	5	216	203		5	< 35	30		1	273	270
12	34	15	6	48	76		6	< 34	8		2	176	148
14	108	89	7	< 15	24		7	< 33	30		3	271	270
16	32	17	8	41	26		8	< 34	32		4	32	23
18	35	31	9	< 18	18		9	< 31	24		5	206	191
			10	205	228		10	< 26	19		6	114	75
	01 l		11	156	148		11	< 24	3		7	77	78
1	99	92	12	29	13						8	112	98
2	186	201	13	33	28			09 l			9	186	172
3	236	334	14	17	10		1	66	100		10	64	53
4	275	392	15	56	49		2	34	32		11	59	62
5	137	118	16	48	29		3	47	52		12	18	34
6	< 13	29					4	60	66		13	136	129
7	18	3					5	63	80		14	21	18
8	147	129		05 l			6	28	28		15	43	45
9	208	221	1	177	194		7	< 26	25		16	52	49
10	105	110	2	178	219		8	< 18	27		17	33	33
11	43	33	3	26	21		9	< 18	27				
12	88	91	4	32	32			25	36				
13	105	114	5	209	232							13 l	
14	26	11	6	< 13	18			0,10, l			0	358	386
15	64	60	7	108	83		0	43	31		1	236	206
16	14	0	8	139	156		1	52	64		2	88	65
17	< 14	0	9	99	76		2	< 20	6		3	134	108
			10	118	119		3	< 19	4		4	225	193
	02 l		11	84	94		4	< 18	0		5	330	356
0	308	657	12	31	33		5	< 46	48		6	254	263
1	145	111	13	30	22						7	92	73
2	81	75	14	26	22			10 l			8	107	98
3	55	55	15	71	54		1	< 19	0		9	71	41
4	347	522	16	61	49		2	232	260		10	116	117
5	83	64					3	86	100		11	207	205
6	75	66					4	237	248		12	146	140
7	56	41		06 l			5	78	74		13	55	41
8	135	120	0	168	139		6	32	19		14	22	15
9	88	94	1	140	113		7	73	88		15	52	43
10	236	255	2	102	111		8	101	69		16	16	13
11	143	136	3	< 34	9		9	200	197				
12	105	94	4	< 65	70		10	48	49			14 l	
13	47	25	5	< 35	4		11	< 21	14		0	208	202
14	71	60	6	154	166		12	57	51		1	105	91
15	29	16	7	115	100		13	< 22	9		2	199	171
16	59	52	8	58	62		14	65	64		3	125	110
17	< 13	7	9	35	32		15	65	46		4	135	136
			10	42	55		16	19	3		5	120	105
	03 l		11	< 34	35		17	38	29		6	100	86
1	116	126	12	40	46						7	103	90
2	175	168	13	< 28	0			11 l			8	109	114
3	135	115	14	28	10		0	406	728		9	142	133
4	102	90	15	67	67		1	361	451		10	64	66
5	164	174					2	251	250		11	< 19	8
6	82	52					3	149	146		12	41	49
7	136	134		07 l			4	164	166		13	17	17
8	26	34	1	101	137		5	365	440		14	23	13
9	56	63	2	100	81		6	71	58		15	124	116
10	< 22	0	3	34	43		7	13	32		16	45	27
11	67	46	4	146	148		8	208	194				
12	< 23	8	5	82	97		9	136	129			15 l	
13	62	67	6	< 40	67		10	122	119		0	61	54
14	74	76	7	< 36	22		11	107	101		1	144	129
15	29	25	8	< 35	24		12	119	119		2	144	148
16	16	13	9	< 34	37		13	52	54		3	97	66
17	12	7	10	37	33		14	35	39		4	71	92

Table 1 (cont.)

5	153	135		1,10,l	4	122	127	4	121	136	
6	74	66	0	38	55	5	173	5	77	97	
7	105	88	1	< 14	16	6	106	6	25	5	
8	16	14	2	39	35	7	67	7	< 20	30	
9	62	49	3	45	60	8	95	8	25	32	
10	29	22	4	23	36	9	73	9	24	28	
11	67	70				10	96	10	19	25	
12	< 17	19		20l		11	62	11	44	48	
13	14	14	0	285	287	12	75	12	35	39	
14	42	43	1	82	74	13	26	13			
15	79	70	2	23	22	14	37	14	28l		
			3	268	292	15	56	15	58	74	
			4	162	151	16	< 13	16	31	51	
	16l		5	< 32	27			1	33	29	
0	141	166	6	83	87		24l	2	52	66	
1	37	29	7	128	114	0	278	274	4	55	61
2	< 20	37	8	45	30	1	90	99	5	< 24	19
3	< 20	24	9	43	26	2	114	98	6	< 19	29
4	124	124	10	169	204	3	115	91	7	22	36
5	25	26	11	116	136	4	36	21	8	38	41
6	84	112	12	< 22	24	5	76	75	9	< 17	40
7	39	52	13	139	121	6	170	186	10	44	47
8	106	105	14	< 20	30	7	147	180		29l	
9	77	72	15	28	8	8	46	68	0	< 19	24
10	86	92	16	107	94	9	58	69	1	< 18	9
11	33	27	17	102	91	10	< 19	13	2	68	91
12	33	28				11	108	119	3	48	50
13	36	43		21l		12	33	35	4	48	60
14	41	33		0	162	119	13	47	42	42	42
			1	213	202	14	< 18	9	5	< 18	42
	17l		2	183	168	15	12	4	6	< 17	31
0	< 23	41	3	248	256				7	< 14	19
1	146	153	4	71	76		25l			2,10,l	
2	40	68	5	143	130	0	32	52	0	18	3
3	54	61	6	80	88	1	79	72	1	43	51
4	100	100	7	75	67	2	91	93			
5	56	45	8	87	70	3	39	15		30l	
6	30	32	9	144	133	4	20	39	1	203	230
7	77	75	10	91	86	5	59	71	2	47	69
8	59	51	11	49	56	6	72	62	3	21	5
9	30	16	12	54	50	7	101	97	4	118	124
10	57	54	13	69	67	8	134	154	5	126	136
11	26	39	14	35	36	9	117	106	6	69	16
12	22	24	15	29	19	10	57	59	7	144	143
13	< 53	24	16	39	26	11	33	33	8	67	57
			17	32	23	12	63	67	9	29	46
						13	76	62	10	39	41
				22l		14	88	69	11	53	56
0	111	122	0	253	241	15	75	78	12	< 20	0
1	26	21	1	129	121				13	60	42
2	126	102	2	149	135		26l		14	38	34
3	39	42	3	253	238	0	57	72	15	21	18
4	142	142	4	123	101	1	166	177	16	36	23
5	< 22	39	5	53	44	2	< 20	27			
6	61	74	6	206	203	3	59	77		31l	
7	50	64	7	166	164	4	81	86	0	238	216
8	< 19	15	8	63	52	5	85	91	1	62	57
9	< 17	15	9	71	60	6	80	86	2	124	117
10	58	76	10	83	91	7	130	125	3	78	74
11	25	41	11	113	102	8	51	57	4	116	107
			12	77	78	9	< 19	26	5	64	41
			13	92	76	10	60	47	6	143	142
	19l		14	79	85	11	37	36	7	50	70
0	50	50	15	39	37	12	55	53	8	84	63
1	35	31	16	44	38	13	< 15	14	9	86	98
2	< 20	11							10	47	58
3	< 19	20		23l			27l		11	58	75
4	< 19	13		0	95	37	0	31	8	80	90
5	48	51	0	105	101	1	89	81	12	27	27
6	23	39	1	105	111	2	124	114	13	78	78
7	56	49	2	140	129	3	68	76	14	78	78
8	14	9	3						15	25	18

Table 1 (cont.)

6	< 11	4	4	< 26	0	3	39	40	3	< 11	12
7	44	57	5	26	42	4	63	69	4	45	47
			6	43	38	5	48	32			
	60 <i>l</i>									63 <i>l</i>	
0	< 30	34		61 <i>l</i>			62 <i>l</i>		0	25	32
1	< 30	13	0	< 13	10	0	37	37	1	56	56
2	< 29	0	1	47	53	1	< 12	25	2	46	64
3	< 28	28	2	16	19	2	40	52	3	27	43

use in locating the hydrogen atoms with any degree of precision. A plane of $\rho_o(x, y, z)$, perpendicular to (100) and including the peaks of O_1 (A) and O_3 ($B_{0\bar{1}1}$) is shown in Fig. 3. Contours are drawn at 1, 2, 4, 6, . . .

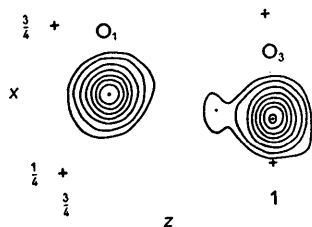


Fig. 3. Fourier section from $\rho_o(x, y, z)$. The section is perpendicular to (100), and passes through the center of the maxima of O_1 and O_3 . Dots indicate the positions of O_1 , O_3 , and $H(O_3)$ according to final parameters. Contours are drawn at 1, 2, 4, 6 . . . e. \AA^{-3} .

e. \AA^{-3} . The hydrogen atom of the hydrogen bond between these two atoms shows quite clearly. The atom centers in these functions were located by the analytical method (Shoemaker *et al.*, 1950), with the difference that 19 instead of 27 points were used, since the eight corner points were in general of such low density, i.e. less than one-third of the maximum, as not to be fitted by the assumed Gaussian function. It is probably more appropriate to use 19 points in locating peak centers when the distance between adjacent points is as large as about 0.2 \AA . The back-shift corrections were found by comparing the positions of the peak centers in the density distribution $\rho_c(x, y, z)$ with the parameters used in its computation. These corrections, which arise largely from the abrupt cut-off of the Fourier series at the limit of Cu $K\alpha$ radiation, and from the omission of the 27 inaccessible reflections, averaged 0.006 \AA ; they were then applied to the positions of the maxima in $\rho_o(x, y, z)$. In this non-centrosymmetric structure, the n -shift rule (Shoemaker *et al.*, 1950) must be applied in finding the true parameters, since the electron-density calculation makes use of calculated phase angles in addition to observed amplitudes. The value $n = 1.6$ was used; this is the value found empirically for threonine, and the proportion of real to complex structure factors is about the same for the two crystals. The parameters obtained are quite insensitive to n ; use of $n = 1.7$, for example, in place of 1.6, gives an average difference of 0.0009 \AA and a maximum difference of 0.0022 \AA . The Fourier parameters were obtained in accordance with the above procedures, and were then averaged

with the set from the third least-squares treatment to give what we term the final parameters. The Fourier and least-squares values differ by an average of 0.008 \AA , with a maximum deviation of 0.023 \AA and a root-mean-square deviation of 0.010 \AA .

The final parameters were used to calculate two sets of structure factors, $|F_{hkl}|$ (4) and $|F_{hkl}|$ (5), which respectively omit and include the contributions of the hydrogen atoms. Set $|F_{hkl}|$ (5) is compared with the observed values of $|F_{hkl}|$ in Table 1.

The stages of the refinement procedure are illustrated in Tables 2 and 3. Table 2 lists the parameters of the first trial structure, as determined from the Patterson function and preliminary one-dimensional refinement; the parameters obtained from the several two-dimensional refinements; and the final

Table 2. Refinement of parameters

		First trial structure*	After two-dimensional refinement	Final parameters
C_1	x	0.191	0.201	0.2062
	y	0.581	0.558	0.5617
	z	0.206	0.207	0.2086
C_2	x	0.384	0.377	0.3763
	y	0.517	0.520	0.5212
	z	0.300	0.293	0.2938
C_3	x	0.387	0.377	0.3647
	y	0.646	0.654	0.6495
	z	0.369	0.370	0.3711
C_4	x	0.335	0.319	0.3290
	y	0.563	0.565	0.5646
	z	0.450	0.463	0.4637
C_5	x	0.195	0.190	0.1866
	y	0.380	0.407	0.4088
	z	0.440	0.441	0.4375
N	x	0.255	0.275	0.2774
	y	0.364	0.371	0.3704
	z	0.342	0.340	0.3407
O_1	x	0.010	0.012	0.0077
	y	0.481	0.471	0.4746
	z	0.191	0.190	0.1907
O_2	x	0.269	0.278	0.2794
	y	0.685	0.684	0.6850
	z	0.165	0.163	0.1626
O_3	x	0.582	0.596	0.5979
	y	0.518	0.528	0.5274
	z	0.508	0.496	0.4974

* Before first two-dimensional refinement.

Table 3. *Percentage discrepancies* during refinement*

	All planes	<i>hkl</i>	<i>h0l</i>	<i>0kl</i>	<i>hkl</i> only	$\sin \theta < 0.6$	Comments
$ F_{hkl} $ (1)	20.0	37.7	24.9	28.4	19.9	—	After two-dimensional refinements
$ F_{hkl} $ (3)	16.5	—	—	—	15.8	—	After two three-dimensional refinements; including H atoms
$ F_{hkl} $ (4)	16.8	23.1	21.6	21.8	14.5	14.4	Final parameters, excluding H atoms, and five planes showing extinction†
$ F_{hkl} $ (5)	16.7	28.7	18.6	21.8	14.4	15.9	Final parameters, including H atoms; all planes included
$ F_{hkl} $ (5)	14.8	16.0	15.6	15.5	14.4	11.1	Final parameters, including H atoms; five planes showing extinction† omitted

* Percentage discrepancy (R) = $100 \frac{\sum_{hkl} |F_{hkl}|_o - |F_{hkl}|_c}{\sum_{hkl} |F_{hkl}|_o}$.

† The five planes showing extinction are: (110), (006), (014), (020) and (024). Although secondary extinction was apparently important for these planes, the nature of the powder pattern for the crystal precluded any correction for extinction.

parameters. Table 3 shows the percentage discrepancy of the calculated and observed structure factors at various stages during the refinement. The final overall percentage discrepancy (14.8%) is somewhat greater than in some comparable structures (e.g. L-threonine, 11.2%; acetylglycine, 13.4%; DL-alanine, 14.6%) but is entirely satisfactory. The larger discrepancy in the present analysis may be related to the fact that there is a larger fraction of absent reflections (not including space-group extinctions) in hydroxyproline (14.0%) than in the other structures (e.g. threonine, 10.2%).

The average standard error in a parameter, calculated in the usual way (Shoemaker *et al.*, 1950) with the least-squares data, is 0.009 Å. The corresponding probable error in an interatomic distance is $0.009 \times \sqrt{2} \times 0.6745 = 0.009$ Å. A conservative estimate of the limit of error for an interatomic distance is thus about 0.03 Å. The corresponding limit of error for bond angles near 110° and bond lengths near 1.5 Å is 1.0°.

The parameters assigned to the hydrogen atoms are listed in Table 4. The parameters of H(C₃) and H'(C₄)

Table 4. *Parameters of hydrogen atoms*

		<i>x</i>	<i>y</i>	<i>z</i>
1	H(C ₂)	0.588	0.506	0.271
2	H(C ₃)	0.544	0.718	0.362
3	H'(C ₃)	0.189	0.726	0.356
4	H'(C ₄)	0.199	0.640	0.505
5	H(C ₅)	0.249	0.313	0.486
6	H'(C ₅)	-0.030	0.418	0.435
7	H(N)	0.434	0.290	0.347
8	H'(N)	0.140	0.319	0.298
9	H(O ₃)	0.558	0.528	0.564

differ slightly from those used in calculating the contributions of the hydrogen atoms to the structure factors, as adjustments were made in the final stages on the basis of van der Waals contacts between molecules. The nine hydrogen atoms are designated as follows: the heavy atom to which the hydrogen atom is covalently bound follows in parentheses, a prime following the H indicates that the hydrogen atom is

on the same side of the pyrrolidine ring as the carboxyl group, an unprimed *H* indicates the hydrogen atom is on the same side of the ring as the hydroxyl group.

Discussion of the structure

The hydroxyproline molecule

Two schematic drawings of the hydroxyproline molecule as it exists in the crystal, showing the interatomic distances and bond angles, are presented in Fig. 4; the expected *trans* configuration of the carboxyl and hydroxyl groups relative to the ring is confirmed. The bond distances and bond angles in this molecule are listed in Tables 5 and 6 together with the corresponding values for other amino acids.

The interatomic distances in hydroxyproline are in general close to the expected values. The average C-C distance is significantly smaller than the classical value of 1.54 Å; this situation also obtains in the other amino acids. More striking is the apparent alternation

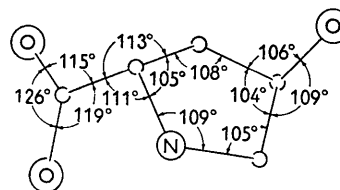
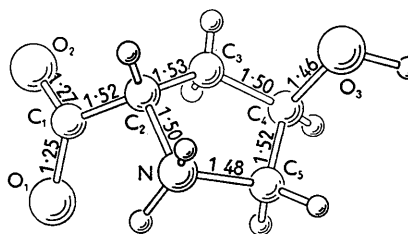


Fig. 4. The hydroxy-L-proline molecule, showing interatomic distances and bond angles.

Table 5. *Intramolecular interatomic distances*

Bond	Hydroxy-L-proline	L-Threonine*	DL-Alanine†	DL-Serine‡
C ₁ -C ₂	1.516 Å	1.517 Å	1.536 Å	1.528 Å
C ₂ -C ₃	1.532	1.542	1.513	1.513
C ₃ -C ₄	1.503	1.505	—	—
C ₄ -C ₅	1.524	—	—	—
C ₅ -N	1.482	—	—	—
N-C ₂	1.503	1.490	1.496	1.491
C ₁ -O ₁	1.254	1.236	1.211	1.268
C ₁ -O ₂	1.269	1.253	1.273	1.261
C ₄ -O ₃	1.460	—	—	—
C ₃ -O ₃	—	1.424	—	1.425
N...O ₁	2.666	2.672	2.688	2.638

* Shoemaker *et al.* (1950).

† Donohue (1950).

‡ Shoemaker, Barieau, Donohue & Lu (private communication).

Table 6. *Intramolecular bond angles*

Angles	Hydroxy-L-proline	L-Threonine	DL-Alanine	DL-Serine
(a) External angles				
O ₁ -C ₁ -O ₂	126.1°	126.9°	125.4°	125.3°
O ₁ -C ₁ -C ₂	118.5°	117.0°	121.3°	117.4°
O ₂ -C ₁ -C ₂	115.4°	116.1°	113.2°	117.2°
C ₁ -C ₂ -C ₃	113.3°	113.4°	111.2°	110.3°
C ₁ -C ₂ -N	110.8°	110.4°	108.3°	110.0°
C ₃ -C ₄ -O ₃	106.1°	—	—	—
C ₃ -C ₄ -O ₃	109.2°	—	—	—
C ₂ -C ₃ -O ₃	—	104.1°	—	112.0°
C ₄ -C ₃ -O ₃	—	110.5°	—	—
(b) Angles within the ring in hydroxy-L-proline and their analogs				
N-C ₂ -C ₃	104.5°	108.0°	110.4°	111.1°
C ₂ -C ₃ -C ₄	107.6°	112.5°	—	—
C ₃ -C ₄ -C ₅	103.9°	—	—	—
C ₄ -C ₅ -N	105.5°	—	—	—
C ₅ -N-C ₂	109.4°	—	—	—

in these bond lengths, which is remarkably similar in hydroxyproline and threonine (which contains a four-carbon chain); whether this similarity is more than coincidence cannot be said at this time. In the same way, the close parallelism in the C-C distance in the two similar three-carbon compounds, alanine and serine, may be no more than coincidence; moreover, the variation is in the opposite sense (C₁-C₂ being larger than C₂-C₃) to what it is in the compounds with longer carbon chains. Similar alternations in bond length which have been tentatively reported in hexamethylenediamine and its dihydrochloride, and in adipic and sebacic acids by Robertson and his coworkers (Binnie & Robertson, 1949, 1950; Morrison & Robertson, 1949*a, b*) suggest that these alternations are indeed real.

The C-N distance is, within experimental error, the same as that found in other amino acids. On the other hand, the C-O distance involving the hydroxyl group (1.46 Å) appears significantly larger than that in threonine and serine, where the distance is essentially equal to the sum of the covalent radii (1.43 Å). This bond lengthening may be simply a steric effect. The

energy required to stretch a carbon-oxygen single bond by 0.03 Å may be estimated from the force constants for such a bond to be about 0.5 kcal. In view of the crowding, discussed below, of the comparatively bulky hydroxyl group by the adjacent methylene groups it is not surprising that some relief from the sidewise compressional strain is sought; the bond stretching provides a possible mode for such relief.

The five-membered pyrrolidine ring in the hydroxyproline molecule as it exists in the crystal is appreciably puckered; C₄, the carbon which bears the hydroxyl group, is about 0.4 Å from a plane defined (within 0.03 Å) by the other four atoms of the ring. Furthermore, C₄ is on the opposite side of this plane from the carboxyl group. The angle between the two planes, i.e. C₅-N-C₂-C₃ and C₃-C₄-C₅, is 17°. This distortion, or puckering, of the ring is clearly evident from an inspection of the values of the bond angles within the ring in Table 6. If the ring were planar the average value of these angles would be 108°. However, as one atom moves out of the plane the average angle necessarily becomes smaller. The puckering of this ring is not unexpected in view of the convincing evidence (Pitzer, 1945; Kilpatrick, Pitzer & Spitzer, 1947) that a puckered configuration of cyclopentane is more stable than the planar one. The explanation of this effect lies in the repulsion of non-ring atoms bonded to adjacent ring atoms in the planar structure since in the planar structure these non-bonded atoms are forced to occupy an eclipsed position. When the ring becomes puckered, a staggered position is possible, with a consequent gain in stability; however, this gain in stability is, in part, offset by the distortion of the bond angles within the ring to values appreciably below that of the tetrahedral angle, 109° 28'.

It seems reasonable that the ring atom which moves out of the plane will be one of the two to which a bulky group is attached, i.e. either C₂ (to which the carboxyl group is attached) or C₄ (which carries the hydroxyl group), since moving a large group from the eclipsed position should provide more relief from strain than moving a hydrogen atom. The fact that the largest angle within the ring is that at the nitrogen atom is reasonable if one considers the unsubstituted ring and assumes that the substitution of a nitrogen atom for a carbon atom does not change the force constants. With this assumption, the fact that C-N bonds are slightly shorter than C-C bonds necessarily means that the nitrogen atom must be nearer the center of the ring, and thus that the angle at the nitrogen atom must be greater than those at the carbon atoms.

One more final feature of the structure of the molecule is worth noting: C₂ and the three atoms of the carboxyl group are precisely planar, the sum of the three angles around C₁ being 360.0°. This situation is of course to be expected on theoretical grounds, and inspection of Table 6 shows that it holds (within 0.1°) for the other amino acids there listed, as it should for

all carboxylic acids. None of the atoms C_1 , C_2 , O_1 or O_2 deviates by as much as 0.004 Å from a plane derived from their coordinates by the method of least squares. It is of interest that the nitrogen atom lies only 0.05 Å from this plane. The near equality of the two carbon-oxygen bonds in the carboxyl group supports strongly the assignment of the *Zwitterion* structure to these molecules in their crystals.

Intramolecular distances and bond angles involving the hydrogen atoms are summarized in Table 7.

Table 7. *Intramolecular distances and angles involving hydrogen atoms*

$C_2-H(C_2)$	1.11 Å	$H(C_2)-C_2-C_1$	109°
$C_3-H(C_3)$	1.09	$-C_2-C_3$	109
$C_3-H'(C_3)$	1.11	$-C_2-N$	110
$C_3-H'(C_4)$	1.11	$H(C_3)-C_3-C_2$	105
$C_5-H(C_5)$	1.10	$-C_3-C_4$	118
$C_5-H'(C_5)$	1.09	$-C_3-H'(C_3)$	109
$N-H(N)$	1.03	$H'(C_3)-C_3-C_2$	107
$N-H'(N)$	1.01	$-C_3-C_4$	110
$O_3-H(O_3)$	0.98	$H'(C_4)-C_4-C_3$	105
$O_1 \cdots H'(N)$	2.10	$-C_4-C_5$	111
$O_2 \cdots H(C_2)$	2.64	$-C_4-O_3$	119
$O_2 \cdots H'(C_3)$	2.80	$H(C_5)-C_5-C_4$	109
$C_1 \cdots H'(C_3)$	2.50	$-C_5-N$	110
$O_3 \cdots H(C_3)$	2.45	$-C_5-H'(C_5)$	111
$O_3 \cdots H(C_5)$	2.50	$H'(C_5)-C_5-C_4$	114
$H(O_3) \cdots H'(C_4)$	2.15	$-C_5-N$	107
		$H(N)-N-C_5$	107
		$-N-C_2$	109
		$-N-H'(N)$	107
		$H'(N)-N-C_5$	116
		$-N-C_2$	108
		$H(O_3)-O_3-C_4$	97

The molecular environment

Two views of the structure are shown in Figs. 5 and 6. In these figures and in tables of intermolecular

distances the molecule with coordinates (x, y, z) is labeled M . Molecules labeled A , B , and C are related to M by the operation of the twofold screw axes parallel to a_0 , b_0 and c_0 respectively; the coordinates of these molecules are $(\frac{1}{2}+x, \frac{1}{2}-y, 1-z)$, $(1-x, \frac{1}{2}+y, \frac{1}{2}-z)$ and $(\frac{1}{2}-x, 1-y, \frac{1}{2}+z)$. Molecules equivalent to these in adjacent unit cells have the additional designation of a lattice translation vector. Intermolecular contacts of molecule M are summarized in Table 8.

As may be seen in Figs. 5 and 6, the hydroxyproline crystal is held together by a system of hydrogen bonds, the most important of which is that between the nitrogen atom and the oxygen atom O_2 of the carboxyl group, e.g. between molecules M and B_{010} . The angles $C_2-N \cdots O_2$ and $C_5-N \cdots O_2$ are 101° and 113°. The second hydrogen bond formed by the nitrogen atom is quite weak—the $N \cdots O$ distance is 3.17 Å, and the $H'(N) \cdots O_2$ distance of 2.44 Å is only about 0.2 Å shorter than would be expected for a simple van der Waals contact, whereas in strong $N-H \cdots O$ bonds the $H \cdots O$ distance is 0.7 Å or more shorter than the sum of the van der Waals radii. The values for the angles are $C_2-N \cdots O_2 = 133^\circ$ and $C_5-N \cdots O_2 = 81^\circ$ for this long hydrogen bond. The reason for the weakness of this bond is quite clear; the tendency in α -amino acids for the nitrogen atom to lie rather close to one of the oxygen atoms of the carboxyl group is quite marked (see Table 5). In the case of DL-serine and hydroxy-L-proline, the nitrogen atom lies very nearly in the plane of the carboxyl group. In DL-serine it is possible, nevertheless, for all three hydrogen atoms of the $-NH_3^+$ group to take part in strong hydrogen bonds (to three different oxygen atoms, the $N \cdots O$ distances being 2.79, 2.81 and 2.87 Å) because this group is free to rotate so that

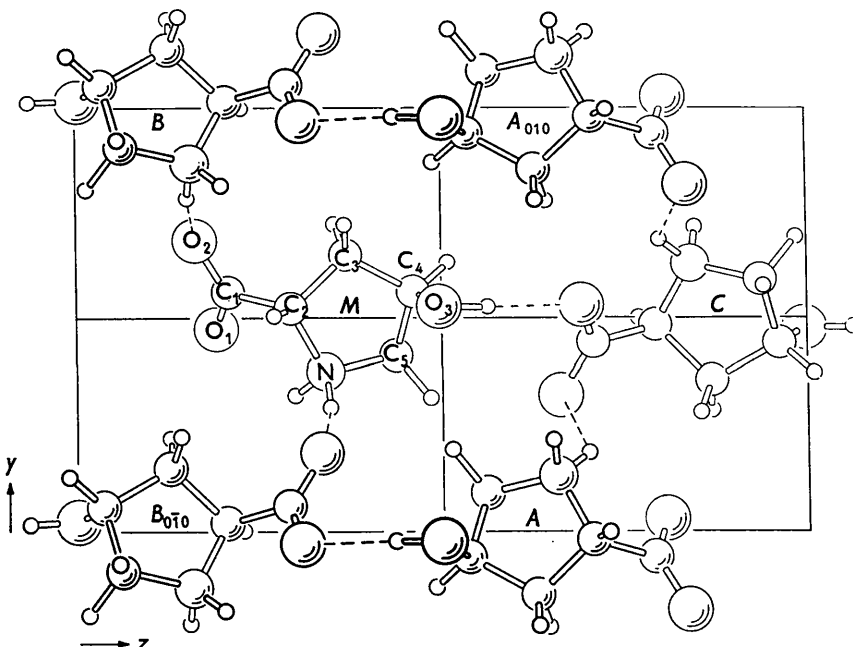
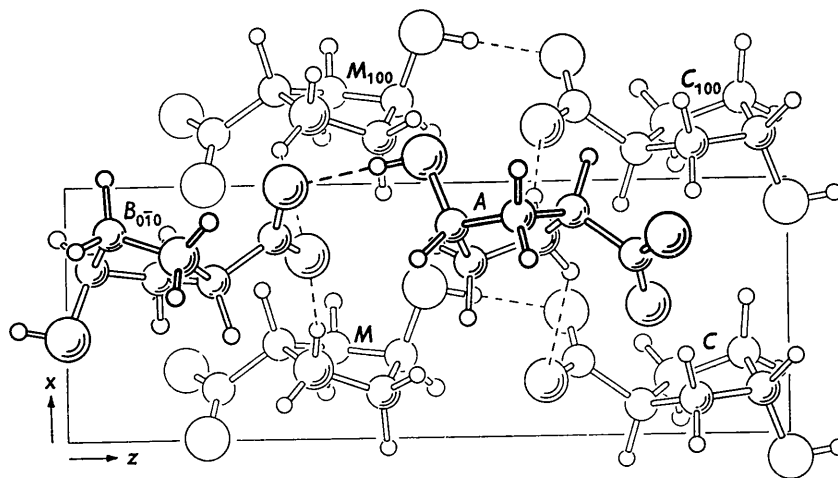


Fig. 5. View of the structure along the x axis.


 Fig. 6. View of the structure along the y axis.

the disposition of the hydrogen atoms is staggered with respect to the carboxyl oxygen. In hydroxy-L-proline, on the other hand, the closure of the five-membered ring makes this rotation impossible, and one hydrogen atom, $H'(N)$, is eclipsed with respect to O_1 of the carboxyl group. The favorable position

of a hypothetical oxygen atom which would be involved in a strong $N-H \cdots O$ bond here, i.e. 2.8 Å from N, on an extension of the $N-H'$ bond, would be only 2.1 Å from O_1 , an improbably short distance. Since these steric factors make the formation of two strong hydrogen bonds by the $>NH_2^+$ group impossible the one which is formed is unusually short, the $N \cdots O$ distance being 2.69 Å, as compared with values from 2.79 to 3.10 Å for the nine $N-H \cdots O$ bonds in threonine, serine and alanine.

Table 8. Intermolecular contacts

From atom X on molecule M	to atom Y	on molecule	Distance $X \cdots Y$
(a) Hydrogen-bond contacts			
O_3	O_1	C	2.80 Å
$H(O_3)$	O_1	C	1.83
N	O_2	$B_{0\bar{1}0}$	2.69
$H(N)$	O_2	$B_{0\bar{1}0}$	1.69
N	O_2	$B_{\bar{1}10}$	3.17
$H'(N)$	O_2	$B_{\bar{1}10}$	2.44
(b) Other distances			
C_5	O_3	$M_{\bar{1}00}$	3.22
$H'(C_5)$	O_3	$M_{\bar{1}00}$	2.25
C_4	C_3	$A_{\bar{1}10}$	4.07
$H'(C_4)$	$H(C_3)$	$A_{\bar{1}10}$	2.36
O_3	C_3	$A_{0\bar{1}0}$	3.53
$H(O_3)$	$H'(C_3)$	$A_{0\bar{1}0}$	2.40
C_5	C_5	A	4.05
$H'(C_5)$	$H(C_5)$	A	2.48
C_3	O_1	$B_{\bar{1}00}$	3.40
$H'(C_3)$	O_1	$B_{\bar{1}00}$	2.39
C_5	O_2	$B_{\bar{1}10}$	3.30
$H(C_5)$	O_2	$B_{\bar{1}10}$	2.52
C_2	O_2	$B_{0\bar{1}0}$	3.34
$H(C_2)$	O_2	$B_{0\bar{1}0}$	2.90
C_4	O_1	C	3.34
$H'(C_4)$	O_1	C	3.29
C_5	O_2	C	3.29
$H(C_5)$	O_2	C	2.52

Moderately strong $O-H \cdots O$ bonds from the hydroxyl group to O_1 of the carboxyl group of a neighboring molecule hold the crystal together in the c direction. The angle $C_4-O_3 \cdots O_1$ is 99° . This $O \cdots O$ distance is 2.80 Å, and is thus appreciably longer than the hydroxyl \cdots carboxyl oxygen bonds in threonine and serine, 2.66 and 2.67 Å respectively. The hydroxyl oxygen, O_3 , is also in quite close contact with a hydrogen atom, $H'(C_5)$, in the molecule directly above in the a direction. The $H \cdots O$ distance here is about what is expected for a weak hydrogen bond, so it is possible that the interaction is significant in this case. There are three other $H \cdots O$ contacts (see Table 8) which are somewhat shorter than the 2.6 Å expected for a simple van der Waals interaction, and these too may correspond to weak hydrogen bonds. Other examples of weak $C-H \cdots O$ interactions have been mentioned by Hunter (1946).

In this structure, as in L-threonine and DL-serine, the hydrogen atom of the hydroxyl group enters into hydrogen-bond formation with the carboxyl oxygen O_1 , the one which lies closer to the nitrogen atom in the same molecule. The reason for this apparent preference is not clear—it may be related to the apparent preferential binding of hydroxyl groups to carboxyl groups rather than to quaternary nitrogen groups in native proteins (Klotz & Urquhart, 1949).

We wish to thank Prof. Robert B. Corey for the original suggestion of this problem, and for his

continued interest and discussions during the course of the work. We are also indebted to Prof. V. Schomaker and Prof. J. H. Sturdivant for helpful discussions. Many of the calculations were done by Miss Lillian Casler.

Note added 27 December 1951.—Zussman (1951b) has now published a brief account of his additional work on the (*Ok*l) and (*h*0l) data. It is interesting to compare his results with those of the present study. Comparing our final parameters, Table 2, with those in Table 4 of Zussman's paper, we find an average difference of 0.020 Å, and a maximum difference of 0.082 Å; for the nine intramolecular bond distances and three hydrogen bond distances, the average difference is 0.018 Å, the maximum difference 0.074 Å; for the twelve interbond angles, the average difference is 2.1°, the maximum difference 5.1°. The agreement between the results of Zussman's work, in which about 220 reflections in two prism zones were used, and the present work, in which about 650 reflections were used, is quite satisfactory, the differences being approximately what is expected by consideration of the probable errors of the two determinations.

The most important discrepancy between the two investigations is that Zussman found a difference of 0.11 Å between the two carboxyl C—O distances, whereas in our study these distances differed by only 0.015 Å. This discrepancy arises from a difference of 0.074 Å in the length of the C₁—O₁ bond (C₆—O₇ in Zussman's notation) and a difference, in the opposite sense, of 0.021 Å for C₁—O₂(C₆—O₈). Zussman interprets his difference of 0.11 Å as indicating that in this crystal the hydroxyproline molecule is not in the *Zwitterion* form. We do not agree with this interpretation. First of all, there is evidence of a statistical nature which tends to show that the difference in the lengths of the two C—O bonds is of questionable significance: if the test suggested by Cruickshank (1949) is applied to Zussman's data to determine the probability that a difference of 0.11 Å between two bonds each with a standard error* of 0.045 Å is due to random errors, one finds $P = 4.2\%$, or 'possibly significant'. If the same test is applied to our results, where the difference is 0.015 Å and the standard error 0.015 Å, we find $P = 24\%$, or 'not significant'.

The conclusion that hydroxyproline, like all of the other amino acids whose structures have been studied in detail, is a *Zwitterion* in the crystal is entirely in accord with the evidence provided by the disposition and lengths of the hydrogen bonds, and, as in other amino acids and peptides, by the effect of the hydrogen bonding on the relative lengths of the C—O distances in the carboxyl group. In alanine and glycylglycine an inequality of the two C—O distances is brought

about by the situation that one oxygen atom is involved in one hydrogen bond of the type N—H···O, the other in two of the same type. In both crystals the two C—O distances differ by 0.06 Å, the shorter one being that of the oxygen atom accepting only one hydrogen bond. The explanation for this difference in terms of the influence of hydrogen bonding on resonance in the carboxyl group was first suggested by Levy & Corey (1941). In threonine and serine, on the other hand, the two carboxyl C—O distances are substantially equal. In each of these crystals, one carboxyl oxygen accepts two hydrogen bonds, N—H···O, of moderate strength (lengths about 2.85 Å), and the other accepts one bond of the type OH···O; moreover, these OH···O bonds are rather strong, the oxygen—oxygen distances each being close to 2.67 Å. If the reasonable assumption is made that one resonance structure of the carboxyl group will be favored over the other if the distribution of charges directed at the two oxygen atoms is markedly unequal, the situation in threonine and serine leads to the conclusion that one rather strong O—H···O bond is about equivalent to two N—H···O bonds of moderate strength. In hydroxyproline, then, where we observe the two carboxyl C—O distances to be very nearly the same, the evidence indicates that the effect of one quite strong N—H···O bond (length 2.69 Å) and one very weak one (length 3.17 Å) is about the same as that of one O—H···O bond of intermediate strength and considerably longer (2.80 Å) than the hydroxyl—carboxyl bonds in threonine and serine.

The question of how best to describe the way in which the pyrrolidine ring is puckered is to some extent arbitrary, although we believe that the discussion we have given above is a very reasonable one in terms of our present knowledge of molecular structure. Our description that atoms C₂—C₃—C₅—N are closely coplanar (within 0.03 Å according to our parameters, within 0.08 Å according to Zussman's) and that atom C₄, the one bonding the hydroxyl group, is about 0.4 Å from this plane seems more satisfactory than the alternate description, that atoms C₂—C₃—C₄—N are coplanar (within 0.08 Å according to our parameters and within 0.04 Å according to Zussman's), with atom C₅ about 0.4 Å from the plane.

Finally, it seems appropriate to comment that we feel the additional labor attendant with the refinement of parameters with complete three-dimensional data to be justified by the resulting increase in accuracy of the results, since any discussion of the various points of molecular structure which depends largely on accurate interatomic distances and bond angles gains appreciably in significance with the increased accuracy. With regard to amino acids and related compounds, molecular dimensions as precise as possible are of particular value in the formulation of proposed structures for the polypeptide chains of proteins. Until recently such proposals were made largely on a qualitative basis, and not until the work

* Zussman estimates a probable error of 0.03 Å in bond length. It is not clear whether this includes the $\pm 1\%$ uncertainty in his axial lengths. Assuming that it does, then the standard error is $0.03 \text{ Å} \times 1.48 = 0.045 \text{ Å}$.

of Pauling & Corey (1950, 1951) and Bragg, Kendrew & Perutz (1950) was it shown that many previously proposed configurations, formerly believed reasonable, were incompatible with the results of careful X-ray studies on a number of compounds related to proteins. It is likely that more such studies will suggest the formulation of additional protein structures and also might well rule out some of the structures now under consideration.

References

- ALBRECHT, G. A. & COREY, R. B. (1939). *J. Amer. Chem. Soc.* **61**, 1087.
- BINNIE, W. P. & ROBERTSON, J. M. (1949). *Acta Cryst.* **2**, 180.
- BINNIE, W. P. & ROBERTSON, J. M. (1950). *Acta Cryst.* **3**, 424.
- BOOTH, A. D. (1947). *J. Chem. Phys.* **15**, 415.
- BRAGG, W. L., KENDREW, J. C. & PERUTZ, M. (1951). *Proc. Roy. Soc. A*, **203**, 321.
- CARPENTER, G. B. & DONOHUE, J. (1950). *J. Amer. Chem. Soc.* **72**, 2315.
- COREY, R. B. (1938). *J. Amer. Chem. Soc.* **60**, 1598.
- COREY, R. B. & DONOHUE, J. (1950). *J. Amer. Chem. Soc.* **72**, 2899.
- CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65.
- DONOHUE, J. (1950). *J. Amer. Chem. Soc.* **72**, 949.
- DONOHUE, J. & TRUEBLOOD, K. N. (1952). *Acta Cryst.* **5**, 414.
- FISCHER, E. (1902). *Ber. deutsch. chem. Ges.* **35**, 2662.
- HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
- HUGHES, E. W. & MOORE, W. J. (1949). *J. Amer. Chem. Soc.* **71**, 2618.
- HUNTER, L. (1946). *Rep. Progr. Chem.* **43**, 153.
- KILPATRICK, J. E., PITZER, K. S. & SPITZER, R. S. (1947). *J. Amer. Chem. Soc.* **69**, 2483.
- KLOTZ, I. M. & URQUHART, J. (1949). *J. Amer. Chem. Soc.* **71**, 1597.
- LEVY, H. A. & COREY, R. B. (1941). *J. Amer. Chem. Soc.* **63**, 2095.
- MORRISON, J. D. & ROBERTSON, J. M. (1949a). *J. Chem. Soc.* p. 987.
- MORRISON, J. D. & ROBERTSON, J. M. (1949b). *J. Chem. Soc.* p. 993.
- NEUBERGER, A. (1948). *Advances in Protein Chemistry*, **4**, 321, 325.
- PAULING, L. (1940). *J. Amer. Chem. Soc.* **62**, 2643.
- PAULING, L. & COREY, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 5349.
- PAULING, L. & COREY, R. B. (1951). *Proc. Nat. Acad. Sci. Wash.* **37**, 235, 241, 251, 256, 261, 272, 282.
- PAULING, L., COREY, R. B. & BRANSON, H. R. (1951). *Proc. Nat. Acad. Sci. Wash.* **37**, 205.
- PITZER, K. S. (1945). *Science*, **101**, 672.
- SHOEMAKER, D. P., DONOHUE, J., SCHOMAKER, V. & COREY, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 2328.
- ZUSSMAN, J. (1951a). *Acta Cryst.* **4**, 72.
- ZUSSMAN, J. (1951b). *Acta Cryst.* **4**, 493.

Acta Cryst. (1952). **5**, 431

The Crystal Structure of Solid Chlorine

BY ROBERT L. COLLIN*

Mellon Institute of Industrial Research, Pittsburgh 13, Pennsylvania, U.S.A.

(Received 14 February 1952 and in revised form 25 February 1952)

The crystal structure of chlorine is based on space group D_{2h}^{18} -*Cmca* with four molecules in a unit cell of $a = 6.29$, $b = 4.50$ and $c = 8.21$ Å. The bond distance within the molecule is 2.02 Å and the shortest non-bonded distance is 3.34 Å.

Introduction

The crystal structure of chlorine at -185° C. has been investigated by Keesom & Taconis (1936). They used both single-crystal and powder techniques and arrived at a structure based on the space group D_{2h}^{16} which gave a Cl-Cl distance of 1.82 Å. This is considerably shorter than the value of 2.01 Å found for the bond distance in gaseous chlorine by electron diffraction (Brockway, 1936). The close intermolecular approach of 2.52 Å found by Keesom & Taconis is also quite surprising in the light of the many X-ray investigations of organic molecules containing chlorine. In no case is the Cl-Cl approach under 3.0 Å, even in the presence

of strong steric effects that tend to pull the non-bonded atoms together.

Because of these anomalous distances the structure proposed by Keesom & Taconis is open to some question and the present investigation has been carried out to check the early work.

Experimental

The chlorine used in this investigation was obtained from a lecture bottle, dried with P_2O_5 and distilled once under vacuum. While the chlorine was kept liquid in a dry-ice bath, thin-walled glass tubes of about 0.5 mm. diameter were filled with the liquid and then sealed at both ends with a flame. Single crystals were grown in the tubes and precession photographs were taken, at around -160° C., according to

* Present address: Department of Physics, The Pennsylvania State College, State College, Pennsylvania, U.S.A.