Sparks. We wish to thank Dr W. Mahler and Prof. A. B. Burg for the sample of $(PCF_3)_4$. The figures were prepared by Mrs Maryellen Reinecke.

References

ATOJI, M. & LIPSCOMB, W. N. (1954). *Acta Cryst.* **7**, 173.
BARTELL, L. S. (1960). *J. Chem. Phys.* **32**, 832.
BARTELL, L. S. & BROCKWAY, L. O. (1960). *J. Chem. Phys.* **32**, 512.
BOWEN, H. J. M. (1954). *Trans. Faraday Soc.* **50**, 463.
CRUICKSHANK, D. W. J. (1956). *Acta Cryst.* **9**, 757.
DUNITZ, J. D. & SCHOMAKER, V. (1952). *J. Chem. Phys.* **20**, 1703.
HAMILTON, W. C. (1955). *Acta Cryst.* **8**, 199.
HOUTEN, S. VAN, & WIEBENGA, E. H. (1957). *Acta Cryst.* **10**, 156.
HULTGREN, R., GINGRICH, N. S. & WARREN, B. E. (1935). *J. Chem. Phys.* **3**, 351.
KUCHEN, W. & BUCHWALD, H. (1956). *Angew. Chem.* **68**, 791.
LEUNG, Y. C. & WASER, J. (1956). *J. Phys. Chem.* **60**, 539.
LEUNG, Y. C., WASER, J., VAN HOUTEN, S., VOS, A., WIEGERS, G. A. & WIEBENGA, E. H. (1957). *Acta Cryst.* **10**, 574.
MAHLER, W. & BURG, A. B. (1957). *J. Amer. Chem. Soc.* **79**, 251.
MAHLER, W. & BURG, A. B. (1958). *J. Amer. Chem. Soc.* **80**, 6161.
MAXWELL, L. R., HENRICKS, S. B. & MOSLEY, V. M. (1935). *J. Chem. Phys.* **3**, 699.
MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.
PAULING, L. (1960). *The Nature of the Chemical Bond*, 3rd ed. Ithaca: Cornell University Press.
SPARKS, R. A., PROSEN, R. J., KRUSE, F. H. & TRUEBLOOD, K. N. (1956). *Acta Cryst.* **9**, 350.
SPENCER, C. J. & LIPSCOMB, W. N. (1961). *Acta Cryst.* **14**, 250.
SUTTON, L., editor (1958). *Interatomic Distances*. London: The Chemical Society, Burlington House, W. 1.
VIERVOLL, H. & ØGRIM, O. (1949). *Acta Cryst.* **2**, 277.
WHEATLEY, P. J. (1960). *J. Chem. Soc.*, p. 523.

Acta Cryst. (1962). **15**, 569

The Crystal Structure of Tosyl-L-prolyl-L-hydroxyproline monohydrate

BY J. FRIDRICHSONS AND A. MCL. MATHIESON

Division of Chemical Physics, Chemical Research Laboratories, C.S.I.R.O., Melbourne, Australia

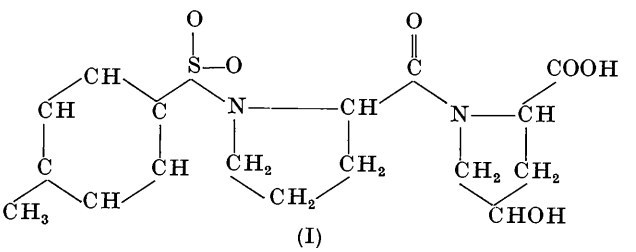
(*Received* 4 *July* 1961)

The structure of the synthetic peptide tosyl-L-prolyl-L-hydroxyproline, $C_{17}H_{22}O_6N_2S.H_2O$, has been determined to establish the stereochemical configuration of the sequence—pro-hypro—and to determine the feasibility of using sulphur as the 'heavy-atom' and the tosyl system as a marker group for the analysis of compounds of comparable complexity. The compound crystallizes in the monoclinic space group $P2_1$ with

$$a = 6{\cdot}291,\ b = 7{\cdot}689,\ c = 19{\cdot}640\ \text{Å};\ \beta = 99°\ 27{\cdot}5',\ Z = 2\ .$$

Intensity data were collected for 0–5 layers about the b-axis. With generalized Patterson functions $P_k(u, w)$, the sulphur atom and a few others on the same y level were located. Subsequent image-seeking operations on modulus generalized Patterson functions and on the zero-layer Patterson distribution indicated the molecular structure which, with the aid of models of the component units —tosyl, proline and hydroxyproline—was established by ϱ_o and $\Delta\varrho$ projections. The y parameters were fixed by three-dimensional ϱ_o (and ϱ_c) syntheses.

It has been established from hydrolysis studies of collagen and gelatin (Schroeder *et al.*, 1954; Kroner *et al.*, 1955) that the proline and hydroxyproline components of these substances occur commonly in the sequence -gly-pro-hypro- and this sequence is now generally accepted in discussing model structures of collagen and gelatin (Rich & Crick, 1955; Cowan & Gavin, 1955; Ramachandran & Kartha, 1955). The structure analysis of a peptide involving the sequence -pro-hypro- seemed therefore to be of interest, insofar as the knowledge of the steric relationships in it might be of some assistance for further developments of these models. We had available for this purpose the synthetic peptide tosyl-L-prolyl-L-hydroxyproline (Beecham, 1957)—$C_{17}H_{22}O_6N_2S$ (I)—and this substance



(I)

because of its p-toluene-sulphonyl (tosyl) group,

proved also to be an excellent test compound for the investigation of the limits of applicability of the 'heavy-atom' method and of image-seeking methods of structure analysis in relation to compounds of comparable complexity. The sulphur atom alone in a molecule of this size (25 'light' atoms) would appear to be rather too lightweight for direct application of the 'heavy' atom method as the ratio, $Z_H^2/\Sigma Z_L^2$, (Lipson & Cochran, 1953; Mathieson, 1955) is only 0·24. However, being linked in the form of the sulphonyl group to the *p*-tolyl ring, the whole tosyl group could be considered as a 'marker' group with sufficient weight to initiate the analysis provided the orientation of the tolyl ring could be determined. Similarly, the tosyl group, once located, would be useful for the investigation and comparison of different image-seeking functions using a progressively increasing number of search points. The results and conclusions derived from these image-seeking tests will be described elsewhere and here they will be mentioned only to the extent that they concern directly the structure determination.

Preliminary results of the analysis were presented earlier (Beecham, Fridrichsons & Mathieson, 1958).

Experimental

The crystals, platelike in shape, belong to the monoclinic space group $P2_1$ with cell dimensions,

$$a=6\cdot291,\ b=7\cdot689,\ c=19\cdot640\ \text{Å};\ \beta=99°\ 27\cdot5',$$

determined against a standard (Pt, $a=3\cdot9231$ Å). The density determined by flotation is 1·415 g.cm.$^{-3}$. The initial micro-analysis, as well as infra-red examination, indicated anhydrous material with formula $C_{17}H_{22}O_6N_2S$ (Beecham, 1957) which leads to a calculated density of 1·355 g.cm.$^{-3}$ for $Z=2$. The structure analysis was initiated on this assumption and only later did it become apparent that the material which, for the analysis, had been recrystallized several times, again from water, and for which the melting point gave little indication of change, was a monohydrate (see analysis section), the calculated density being then 1·419 g.cm.$^{-3}$.

A strong tendency towards twinning was observed. The angular separation of the twin components is however so small that it cannot be detected optically and high-angle equi-inclination Weissenberg photographs had to be used as a diagnostic test for the selection of suitable single crystals.

Intensity data were collected at room temperature on an equi-inclination Weissenberg goniometer for 0–5 layers around the *b*-axis, using multiple-film packs and visual intensity estimation against timed-exposure standards of selected single reflections. The 0*kl* intensities were used to correlate the intensities of the *b*-axis layers. Each layer was given exposures of the order of 200 hr. with a crystal of approximate dimensions $0\cdot35\times0\cdot25\times0\cdot075$ mm. and Cu $K\alpha$ radiation at 30 kV. and 20 mA. Even with such long exposures, the population of reflections in the high-angle region was low, emphasizing the need for experimental conditions which would give more such reflections for precise structure determination of organic compounds of comparable complexity within reasonable time limits (Mathieson, 1961). A total of 1436 reflections of a theoretical 1800 were measured, 247 in the [010] zone. For the *Lp* corrections, the chart by Kaan & Cole (1949) was used, scaling factors and average temperature factors being derived from Wilson (1942) plots. No absorption corrections were applied.

Structure factors and Fourier syntheses for the [010] projections were calculated on the University of Sydney computer, SILLIAC, using programmes devised by Freeman (1957, 1958). The three-dimensional calculations were carried out on the University of New South Wales computer, UTECOM (≡ DEUCE) using coordinated programmes devised by Dr J. S. Rollett. Interatomic distances and angles were computed on a Standard Telephones and Cable Ltd. ZEBRA computer on a programme designed by Dr J. C. Schoone. Scattering curves used for C, N and O were those of Berghuis *et al.* (1955). For S, the values from *Internationale Tabellen zur Bestimmung von Kristallstrukturen* (1935) were used until more accurate values for scattering curves derived by Dawson (1960*a*) became available.

Structure analysis

As a preliminary to analysis, it was considered worthwhile to review possible restrictions on the structure of I arising from structural information on its component units—the tosyl group (Mathieson & Robertson, 1949), proline (Mathieson & Welsh, 1952) and hydroxyproline (Zussman, 1951*a*, *b*; Donohue & Trueblood, 1952). The *p*-methylthiophenyl and the carboxyl groups are planar as is the peptide group and the associated part of the hydroxyproline ring. The prolyl ring is nearly planar. The structure of the molecule may therefore be viewed as composed of four linked planar groups. That the plane of the carboxyl group lies at right angles to the peptide-hydroxyproline plane may be deduced from the analyses of glycyltryptophane.$2H_2O$ (Pasternak, 1956) and glycyltyrosine.HCl (Smits & Wiebenga, 1953). The restrictions on the peptide group mean that the bonds which permit spatial adjustment of the planar groups relative to one another are –S–N– and $-C_{prol.}-C_{pept.}-$. The question arose whether there is a further restriction possible with respect to adjacent planar groups free to rotate. At the time, there was not a great deal of structural evidence but the analyses of sucrose (Beevers & Cochran, 1947) and cytidine (Furberg, 1950) both indicated that in such circumstances the rings tend to orient so that their planes are at right angles. So, for I, we have a probable molecular model with each of the four planar units oriented at right angles with respect to its neighbour(s).

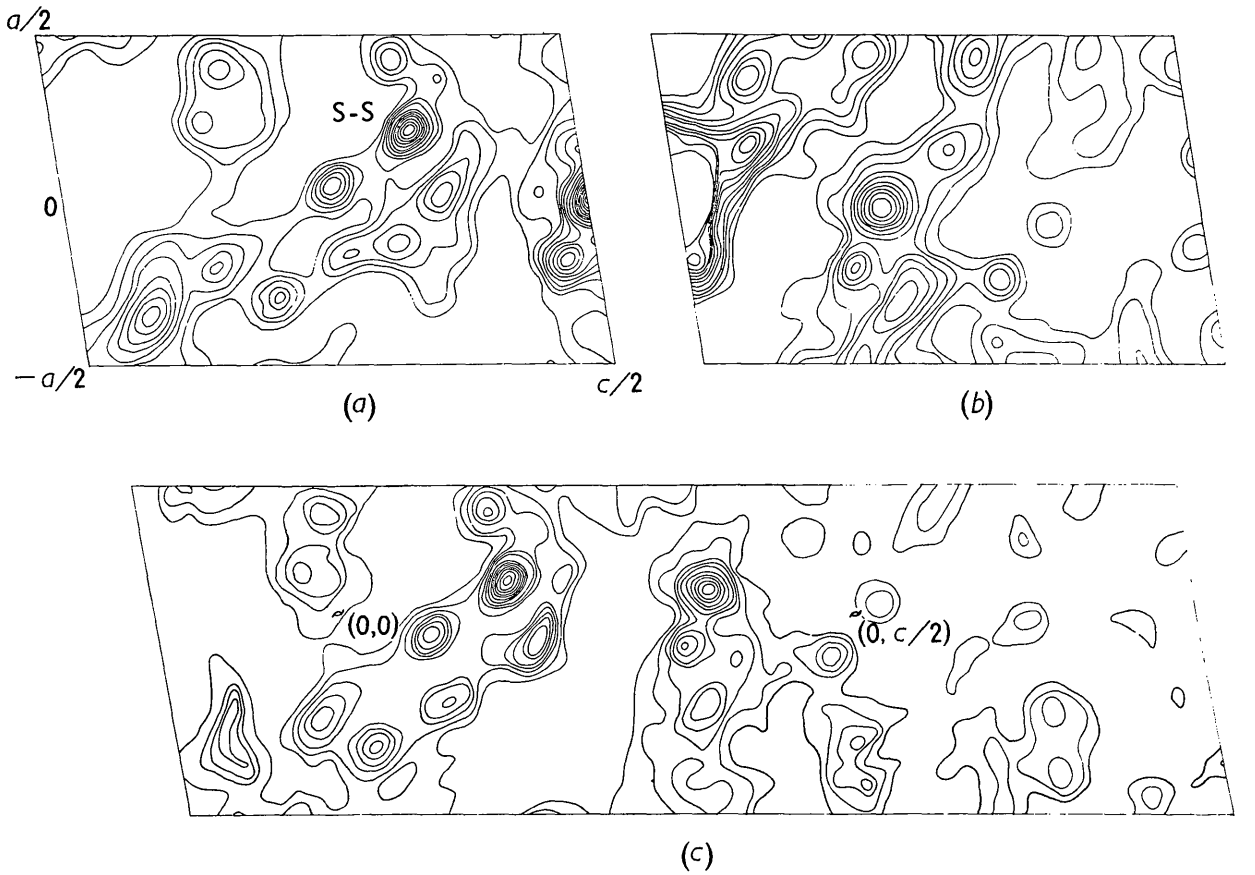


Fig. 1. (*a*) Harker section $H(x, \frac{1}{2}, z)$. (*b*) Harker section $H(x, 0, z)$. (*c*) Distribution derived as a minimum function from the superposition of $H(x, \frac{1}{2}, z)$ and $H(x, \frac{1}{2}, z)$ displaced parallel to the S–S vector.

As to the packing of the molecules, the dimensions of the unit cell suggested that the molecules are extended approximately in the direction of the *c* axis and the chemical structure indicated that the crystal structure could be of the dipolar type e.g. methionine (Mathieson, 1952). It may therefore be considered significant that the *ab* cross-sectional area is 48 Å^2 which may be compared with the value of 43 Å^2 deduced by Brown (1951) as adequate to accommodate two benzene rings. Considered in relation to the space group, this observation suggested that the aromatic ends of adjacent molecules must interpenetrate.

The first step in the analysis of the diffraction data was the location of the sulphur atom and, if possible, some indication of the attached tolyl ring. As the computing facilities then available precluded a three-dimensional Patterson synthesis, generalized Patterson functions $P_k(u, w)$ (Cochran & Dyer, 1952) were calculated for all layers about the *b* axis i.e. $k=0$–5. As could be expected, the normal 0-layer vector map because of severe overlap and the relatively small weight of the sulphur atom did not permit identification of the S–S vector. The generalized distributions were therefore combined to give Harker sections $P(u, \frac{1}{2}, w)$ and $P(u, 0, w)$, Fig. 1. The $P(u, \frac{1}{2}, w)$ distribution revealed several relatively well-resolved peaks and the most prominent was assumed to be the S–S vector. Superposition of the two Harker sections with a relative displacement equal to the S–S vector permitted extraction of a minimum distribution to reveal peaks common to the two functions and corresponding to atoms on the same *y* level as the S atoms. The resultant contour map of the minimum distribution, Fig. 1(*c*), revealed what appeared to be the main C–C · · · C–S axis of the tosyl group as well as some off-axis atoms. From this it could be inferred that the ring lies not far from the *y* level of sulphur and the atomic separation in projection, in conjunction with rough estimates of the *y* parameters from the generalized Patterson projections, permitted assessment of the inclination of the tolyl ring. This disposition of the *p*-methyl-thiophenyl group, with the ring plane not far from perpendicular to the *b* axis and with its main length almost at right angles to 001 is in accord with optical measurements which show that the smallest refractive index (1·55) is parallel to the *b* axis and the largest (1·62) near the direction of *c*.

The probable location of the two oxygen atoms of the sulphonyl group were fixed by inspection of the peaks around the S–S vector in $P(u, \frac{1}{2}, w)$, taking into consideration steric limitations imposed by the bond lengths and angles in the tetrad around S and also rough estimates of the y parameters from the generalized Patterson functions. This procedure also resulted in a tentative positioning of N_{pro} which is linked to the tosyl group.

With the tosyl group defined, the next step was to locate the other atoms of the molecule. It was decided to use image-seeking techniques (Buerger, 1951) for this purpose since a molecule of this size would be expected to provide a relatively stringent test of the capabilities of this method of structure analysis. Preliminary investigations of image-seeking on the zero-layer Patterson function were disappointing since the distributions obtained, although tending to outline the regions of the molecule, did not seem to be reliable enough to use as a basis for model-building. Thus, some regions well defined in the minimum $[P(u, \frac{1}{2}, w),\ P(u, 0, w)]$ function were but poorly indicated. An attempt was therefore made to improve upon the results derived from the zero-layer data.

Earlier experience with modulus functions (Fridrichsons & Mathieson, 1955) suggested a further use for the generalized Patterson functions. These were combined to form

$$\sum_{1}^{5} |P_k(u, w)| \ .$$

In dealing with a vector distribution, as opposed to an electron-density distribution, it appears advisable to omit the zero-layer function since, in it, all vectors are present and cause considerable overlap. Image-seeking by extraction of the minimum function was carried out on the function

$$\sum_{1}^{5} |P_k(u, w)|$$

with the tosyl group as search unit. The resultant distribution was regarded as sufficiently significant to derive a plausible molecular model taking into consideration known structural features for the component units—tosyl, prolyl and hydroxyproline.

With the parameters derived from the first trial model, the first set of $h0l$ structure factors gave a reliability index of 0·52. Adjustment of parameters indicated by the first ϱ_o and $\Delta\varrho$ syntheses reduced it to 0·38 but further cycles did not produce any considerable improvement.

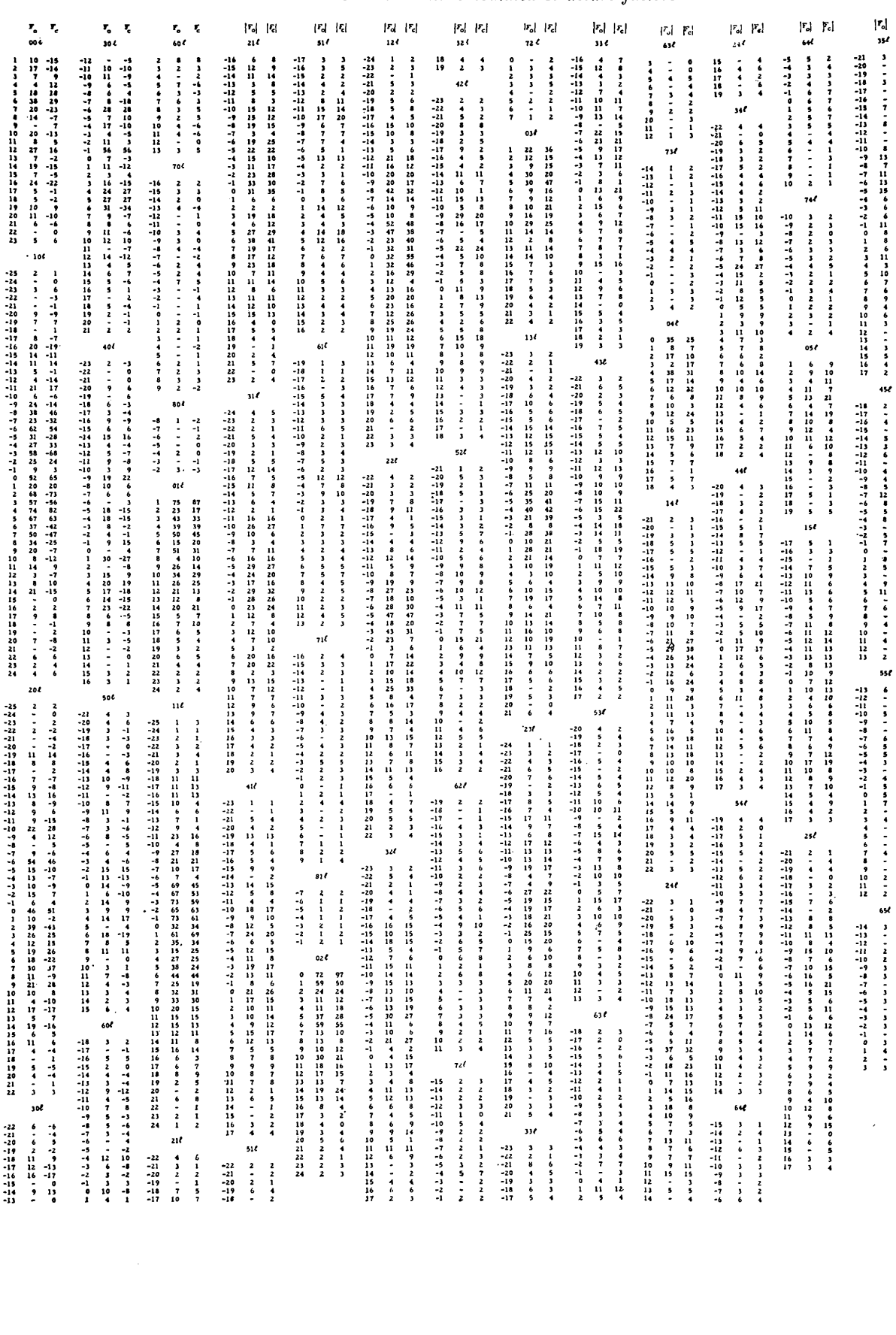
Because it was relatively early in the analysis, it was thought that there might be some misinterpretation of the vector maps, resulting in a partially wrong structure. This possibility was tested by careful reassessment of previous deductions complemented by data from a sharpened ($B \approx 0$) zero-layer Patterson projection with the origin peak removed. Image-seeking on this sharpened Patterson with the S–S and the next strongest vector (assumed to be the through-centre vector of nearly superimposed N and O atoms) yielded a distribution remarkably similar to the previous syntheses. In view of this it seemed certain that the chosen tosyl orientation was correct and that further refinement of the structure would depend on proper adjustments in the prolyl and hydroxyproline groups. Using scale models of these units and adjusting them to fit with the $h0l$ projection some repositioning of these groups and of the carboxyl group led to structure factors with $R = 0{\cdot}33$. The corresponding ϱ_o synthesis did not differ much from the earlier maps, the main improvement being a more uniform value of peak heights for atoms of the same type. For the prolyl group which in projection is viewed almost edge-on, with one oxygen of the sulphonyl group practically coinciding with the nitrogen, the definition

Fig. 2. The projected electron-density distribution $\varrho(x, z)$

Table 1. *Atomic parameters of the molecule*

Atom	x/a	y/b	z/c
C_1	0·933	0·973	0·091
C_2	1·015	0·897	0·037
C_3	0·884	0·884	−0·028
C_4	0·671	0·947	−0·037
C_5	0·596	1·027	0·017
C_6	0·731	1·047	0·082
C_7	0·521	0·937	−0·107
C_8	1·042	0·662	0·212
C_9	0·836	0·900	0·255
C_{10}	0·758	0·714	0·269
C_{11}	0·925	0·566	0·265
C_{12}	0·902	0·967	0·327
C_{13}	0·533	1·080	0·310
C_{14}	0·454	1·225	0·353
C_{15}	0·568	1·197	0·428
C_{16}	0·783	1·115	0·418
C_{17}	0·865	0·975	0·471
N_1	1·033	0·853	0·225
N_2	0·749	1·067	0·346
O_1	1·307	0·982	0·158
O_2	1·047	1·116	0·216
O_3	1·079	0·980	0·366
O_4	0·518	1·391	0·329
O_5	0·797	0·830	0·468
O_6	1·029	1·038	0·523
O_7	1·401	0·699	0·401
S	1·105	0·990	0·172

Table 2. *Observed and calculated structure factors*



is still too poor, allowing only an approximate positioning of the atoms.

Another feature in this ϱ_o map was the presence of an additional peak at $x=0{\cdot}406$, $z=0{\cdot}400$ (noted less obviously in preceding maps) where no scattering matter should have been expected. In the corresponding $\Delta\varrho$ map, this peak represented the highest positive excursion of approximately 4·5 e.Å^{-2} indicating quite definitely an additional atom. The possibility that the material was a hydrate, which earlier had seemed precluded by the micro-analytical data and infra-red measurements, had therefore to be carefully reconsidered. To test the diffraction data further, the contribution of an oxygen atom located at the position of the additional peak was added to the last set of structure amplitudes with a considerable improvement of R from 0·33 to 0·27. Subsequent confirmation of the water molecule was obtained from careful estimation of the cell contents, and also by micro-chemical analysis and determination of weight loss on heating to 110 °C. of the same batch of crystals as had been used in the X-ray analysis. With the water molecule incorporated, the refinement proceeded smoothly to $R=0{\cdot}20$, Fig. 2, at which stage it was decided to turn to the three-dimensional data.

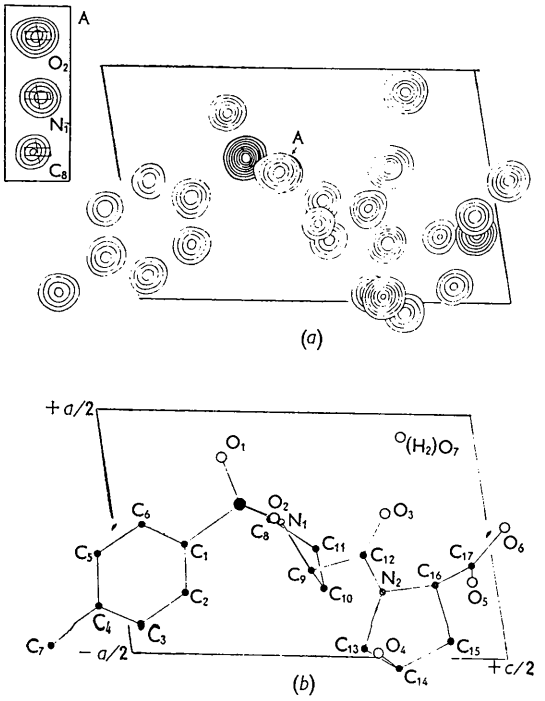


Fig. 3. (*a*) The electron-density distribution of the molecule indicated by sections of $\varrho_o(x, y, z)$, selected near the atom centres. The three atoms C_8, N_1 and O_2 which superimpose adjacent to the S atom are separately indicated at *A* relative to the same cell subdivision for reference. (*b*) The corresponding line diagram of the molecule.

Approximate values of the y parameters for the majority of the 26 atoms had already been deduced in building the molecular model and these were supplemented by values derived from one-dimensional line syntheses of the generalized Patterson functions at the respective x, z positions. Critical comparison of these values combined with the known steric restrictions led to considerable re-arrangement of the prolyl ring involving modifications of the x, z coordinates. With these parameters *hkl* structure amplitudes were calculated and the resultant phases used in a three-dimensional syntheses of electron density. The second set of structure factors did not reveal rapid improvement and for the next step the calculation of $\varrho_o(x, y, z)$, Fig. 3(*a*), was accompanied by $\varrho_c(x, y, z)$. The changes in atomic parameters were derived from the slopes of the $\varrho_o-\varrho_c$ distribution parallel to the axial directions. No further refinement was carried out, for reasons of economy, and the final parameters are given in Table 1. Comparison of the structure factors calculated from these parameters with measured values are recorded in Table 2, the agreement being indicated by the R value for the various layers 0·22(0), 0·21(1), 0·27(2), 0·30(3), 0·34(4), 0·37(5)—mean 0·27. There is considerable scope for refinement of this structure but facilities for an efficient procedure were not available.

Discussion

The interatomic distances and angles in the molecule and the intermolecular approach distances, Table 3, Figs. 3 and 4, were derived from the atomic parameters, Table 1. Because of the more limited range both of the indices k and of the refinement in this direction, the y parameters are less accurate than the x, z parameters. However, estimation of the internal accuracy by comparison of bonds of the same type suggests that the bond lengths are within 0·06 Å of their correct value. In general the bond lengths are normal within the limits suggested. In the case of the carboxyl group, the dimensions suggest that the hydrogen atom is localised on O_6.

With regard to the various groups in the molecule, the *p*-methylthiophenyl group is planar within experimental error. The peptide group is dimensionally in close accord with previous average values (Corey & Donohue, 1950). However, the molecular model reveals that the peptide group is not strictly coplanar. Viewed along the bond $C_{12}N_2$, the line C_9O_3 tilts in one direction while the line $C_{13}C_{16}$ tilts in the opposite direction. The distortion although small appears definite and may be correlated with the intramolecular close approaches between C_{13} and C_9C_{10} and between O_3 and the carboxyl group $C_{17}O_5O_6$ both of which are partly relieved by the slight rotation. Of the hydroxyproline ring, C_{15} lies close to the plane $C_{13}N_2C_{16}$ while C_{14} is a considerable distance (*circa* 0·31 Å) out of this plane. The disposition of the hydroxyl O_4 is *trans* to the carboxyl group as in hydroxyproline itself. The carboxyl group lies approximately at right

angles to the main plane of the peptide–hydroxyproline system consistent with the earlier observations on glycyltryptophane (Pasternak, 1956) and glycyltyrosine (Smits & Wiebenga, 1953). The prolyl group shows, however, some measure of flexibility with regard to the conformation of the ring system. In copper proline, Mathieson & Welsh (1952) found that the atom corresponding to C_{11} (in this structure) lay out of the plane formed of the remaining four atoms (3C+N) and *trans* to the carboxyl group, an arrangement akin to that in hydroxyproline (Zussman, 1951*a*, *b*; Donohue & Trueblood, 1952). However, Leung & Marsh (1958) in their analysis of leucyl-prolyl-glycine showed that the corresponding atom of the pyrrolidine ring gave a somewhat anomalous electron-density distribution which could be ascribed either to violent asymmetric vibration or to equal occupation of sites above and below the plane formed of the other atoms of the pyrrolidine ring. The authors tend towards the latter explanation. In the prolyl-hydroxyproline peptide, the corresponding atom (C_{11}) has swung *cis* with respect to the peptide group. It would appear from the evidence of the analyses of proline, leucyl-prolylglycine and tosyl-prolyl-hydroxyproline that the prolyl ring, in contrast to the hydroxyprolyl ring, is rather flexible and its final orientation in each specific structure may depend largely upon its immediate atomic environment. In the present compound the approach distance C_{11}–C_7^I = 3·95 Å, Fig. 4. If C_{11} were in the *trans* position this distance would be much too small.

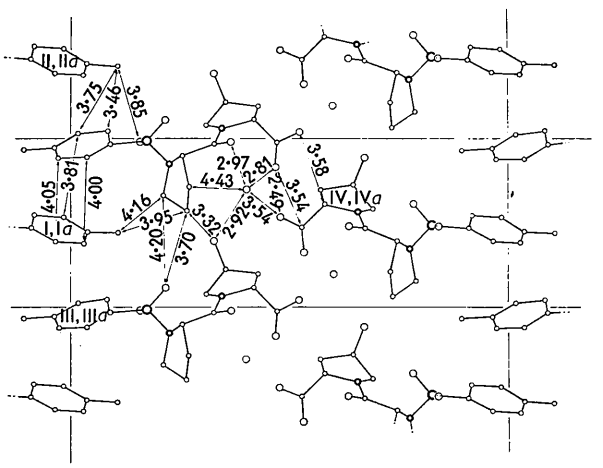


Fig. 4. The arrangement of the molecules as viewed down the *a* axis. Selected approach distances are shown.

The overall shape of the molecule conforms quite closely to the probable shape deduced earlier from the available data i.e. composed of four intersecting planar (or near planar) atomic groups adjusted so that the adjacent planes lie approximately at right angles. The packing of the molecules is illustrated in Figs. 3 and 4 which also show some of the main approach distances, Table 3. The basic arrangement is of polar molecules with their active (carboxylic) ends associated to form dipolar layers extended parallel to the *ab* plane. The molecules are further cross-linked through the agency of the water molecule which is closely allied to the peptide C=O. The relatively inert ends of the molecules (tolyl groups) interpenetrate to permit the maximum number of van der Waals approaches.

There are aspects of the structures of the various collagens and of polymeric proline compounds which

Table 3. *Interatomic distances, bond angles and intermolecular approach distances*

Interatomic distances

Atoms	Bond lengths	Atoms	Bond lengths
C_1–C_2	1·38 Å	C_{16}–C_{17}	1·52 Å
C_2–C_3	1·40	C_1–S	1·78
C_3–C_4	1·41	S–O_1	1·35
C_4–C_7	1·54	S–O_2	1·38
C_4–C_5	1·37	S–N_1	1·60
C_5–C_6	1·43	N_1–C_8	1·50
C_6–C_1	1·38	N_1–C_9	1·49
C_8–C_{11}	1·56	C_{12}–O_3	1·24
C_{11}–C_{10}	1·56	C_{12}–N_2	1·32
C_{10}–C_9	1·56	N_2–C_{13}	1·43
C_9–C_{12}	1·51	C_{14}–O_4	1·45
C_{13}–C_{14}	1·54	N_2–C_{16}	1·45
C_{14}–C_{15}	1·54	C_{17}–O_5	1·19
C_{15}–C_{16}	1·53	C_{17}–O_6	1·41

Bond angles

Atoms	Angle	Atoms	Angle
C_1–C_2–C_3	119·2°	C_{11}–C_{10}–C_9	115·2°
C_2–C_3–C_4	119·7	C_{10}–C_9–N_1	98·8
C_3–C_4–C_7	121·9	N_1–C^9–C_{12}	109·3
C_3–C_4–C_5	120·0	C_{10}–C^9–C_{12}	100·7
C_7–C_4–C_5	118·1	C_9–C_{12}–O_3	132·7
C_4–C_5–C_6	120·8	C_9–C_{12}–N_2	110·9
C_5–C_6–C_1	117·7	O_3–C_{12}–N_2	114·0
C_6–C_1–C_2	122·4	C_{12}–N_2–C_{13}	125·4
C_2–C_1–S	118·1	C_{12}–N_2–C_{16}	114·8
C_6–C_1–S	119·3	N_2–C_{13}–C_{14}	98·3
C_1–S–O_1	105·8	C_{13}–C_{14}–C_{15}	106·6
C_1–S–O_2	115·5	C_{13}–C_{14}–O_4	108·5
C_1–S–N_1	109·9	O_4–C_{14}–C_{15}	109·3
O_1–S–O_2	120·7	C_{14}–C_{15}–C_{16}	103·0
O_1–S–N_1	118·6	C_{15}–C_{16}–N_2	103·7
O_2–S–N_1	85·5	C_{16}–N_2–C_{13}	116·2
S–N_1–C_8	120·3	N_2–C_{16}–C_{17}	117·6
S–N_1–C_9	115·6	C_{15}–C_{16}–C_{17}	114·7
C_8–N_1–C_9	111·6	C_{16}–C_{17}–O_5	123·4
N_1–C_8–C_{11}	107·8	C_{16}–C_{17}–O_6	140·8
C_8–C_{11}–C_{10}	95·1	O_5–C_{17}–O_6	124·7

Intermolecular approach distances

Atoms	Distance	Atoms	Distance
C_2–C_3^I	4·00 Å	C_{11}–O_4^{III}	3·32 Å
C_3–C_2^I	4·05	O_7–O_4^{IIIa}	2·92
C_5–C_5^{Ia}	3·81	O_7–C_{10}	4·43
C_5–C_7^{II}	3·75	O_7–O_3	2·97
C_6–C_7^{II}	3·44	O_7^I–O_5	2·81
O_1–C_7^{IIa}	3·85	O_7–O_6^{IV}	3·54
C_8–C_7^{Ia}	4·16	O_5–C_{17}^{IV}	3·54
C_{11}–C_7^{Ia}	3·95	O_6–C_{16}^{IV}	3·58
C_8–O_2^{III}	4·20	O_5–O_6^{IV}	2·49
C_{11}–O_2^{III}	3·70		

remain uncertain at present and for this reason data on peptides such as leucyl-prolyl-glycine and (tosyl)-prolyl-hydroxyproline may prove useful in elucidating certain points. From the analysis of tosyl-prolyl-hydroxyproline, the following factors may be mentioned. The main one is probably the evidence presented regarding the flexibility of the proline ring, evidence which adds to that from the analysis of leucyl-prolyl-glycine but here the factor producing the *cis* conformation of the ring appears reasonably certain—namely the influence of the *p*-methyl group of an adjacent molecule. This flexibility, which applies only to the prolyl ring, the hydroxyprolyl system being stabilized by the hydroxy group, may permit considerable adjustment of molecular models of proline-containing peptides or proteins and also influence to some extent packing of adjacent parts of the molecule and of neighbouring molecules. Such a molecular adjustment may play a part in the configurational changes observed in poly-L-proline (Steinberg, Harrington, Berger, Sela & Katchalski, 1960).

We have noted that there appears to be a small twist of the peptide group between the prolyl and hydroxyproline units and this deviation from planarity, if substantiated, may also confer a further measure of flexibility in the construction of (hydroxy-) proline-containing peptides or proteins. This relaxation from strict planarity had been considered as a possibility by Cowan & McGavin (1955) in their study of the structure of poly-L-proline, the possibility of a twist having been suggested by the analysis of N,N'-diglycyl-cystine (Yakel & Hughes, 1954) and it has been utilised by Sasisekharan (1959) in his analysis of poly-L-hydroxyproline *A*.

On the basis of the experimental evidence of primary sorption for kangaroo-tail tendon (Rougvie & Bear, 1953), Burge, Cowan & McGavin (1957) have suggested that, since this figure is close to that calculated for one molecule per polar side chain plus one for every two carbonyl groups, the molecular model for collagen can have two possible sites for a water molecule such that it can form hydrogen bonds simultaneously with two carbonyl atoms—(i) with O_5 and O_4^1 of the next chain or (ii) with O_4 and O_5 of the same chain. It is of course not necessary that the system of hydrogen bonds in tosyl-prolyl-hydroxyproline should bear a close relationship to that in collagen but certain similarities in general disposition permits its use as a guide. Thus the water molecule in tosyl-prolyl-hydroxyproline is located so as to approach three oxygens, (*a*) the OH of hydroxyproline, which may be regarded as utilising its H for hydrogen-bonding and also the C=O oxygens, (*b*) of the peptide group and (*c*) of the carboxyl group. The linkage to C=O oxygens of adjacent molecules favours the first possibility of Burge, Cowan & McGavin.

In the present analysis, considerable use has been made of generalized functions. The generalized Patterson functions for the *k*-layers have been used as components of the three-dimensional vector function and in this respect they have one practical advantage over the full three-dimensional distribution in that they are visually readily comprehended. However our main preoccupation with generalized functions has been their particular property that the data in each layer contains essentially the same information regarding two of the three positional parameters for each atom and only partial information regarding the third parameter. In the analysis of a structure use can be made of this property to focus attention first on the derivation of the (x, z) (say) location for each atom and only subsequently deriving the third parameter. This approach differs from structure analysis by simple projections in that every *k*-layer as well as the 0-layer contributes information, thus markedly improving the ratio of the number of data to the number of parameters to be determined (Mathieson, 1961; Fridrichsons & Mathieson, 1961). Use has been made of this type of approach in the analysis of cryptopleurine methiodide (Fridrichsons & Mathieson, 1955) and encouraged by its success in dealing with ϱ functions, a similar application to generalized vector functions was attempted. It can be argued that, as the modulus projection $|\varrho_k(x, z)|$ for each *k*-layer should be, in general features, identical with the distribution for the zero-layer, $\varrho_o(x, z)$, apart from overlap effects, so the modulus of the generalized Patterson function $|P_k(u, w)|$ should be closely allied to the zero-layer Patterson $P_0(u, w)$. Furthermore, for this type of compound, it may be argued that the multiplicity of Z_L–Z_L vectors will tend to neutralize one another and the smaller number of Z_H–Z_L vectors be accentuated. Combination of the modulus functions will tend to accentuate common features (cf. the use of combined moduli functions $\varrho_0+|\varrho_1|+|\varrho_5|$ to extract the best projection coordinates (Fridrichsons & Mathieson, 1955)). Since not all layers will necessarily contain data on all atoms, a simple summation rather than a minimum combination is more appropriate. For this reason, the moduli of the 1–5 layers were combined to form

$$\sum_1^5 |P_k(u, w)| .$$

In dealing with a vector distribution it appears advisable to omit the zero-layer function since, in it, all vectors are present and cause considerable overlap. The combined generalized Patterson function* derived in this manner proved successful in locating the general distribution of the molecule by means of image-seeking with a minimum function. Later work with the normal zero-layer Patterson confirmed the conclusions from the combined generalized functions

* It would not appear likely that this technique can be applied indiscriminately but it may be found useful in particular cases.

and also permitted fairly extensive tests to be made on this method of structure analysis, the process being pushed to the limit to reveal certain limitations. These matters will be dealt with in detail later. However it is of interest to note that the present analysis which required the location of S+26(CNO) represents a somewhat more severe test for the image-seeking procedure than was provided by the analysis of diglycine.HCl i.e. Cl+10(C, N, O) (Buerger & Hahn, 1957) in which image-seeking with the minimum function was carried out on the three-dimensional Patterson function.

One further point may be made with regard to the analysis of diglycine HCl. Buerger & Hahn (1957) have claimed that, in that structure, the Cl atom could not be used as a basis for a 'heavy-atom' solution because $Z_H/Z_L=17/81=0{\cdot}21$. However a more suitable guide in this matter is to assess $\Sigma Z_H^2/\Sigma Z_L^2$ (Mathieson, 1955) which in this case is 0·58, a low value but not outside the range for compounds which can be solved by the direct use of the 'heavy atom'. There are many examples of analyses of this size of which we may quote a recent one, aspartic acid.HCl i.e. Cl+9(C, N, O) (Dawson, 1960*b*). Viewed in this way, it would appear that once the Cl ions in diglycine HCl had been located, analysis would have been possible by phasing by means of the Cl atoms. The structure analysis of tosyl-prolyl-hydroxyproline was initiated partly to determine if it would be possible to solve a structure where the index $Z_H^2/\Sigma Z_L^2$ was as low as 0·24. The completion of the analysis confirmed that this was indeed possible but against this claim we must weigh the considerable amount of information already available on components of the molecule. It therefore remains uncertain if this low index would permit analysis *ab initio* of an unknown structure of this magnitude.

In a similar manner it was thought that this compound would permit assessment of the value in analysis of a 'marker group' i.e. a group of light atoms of known structure compared with the usefulness of the heavy atom. The conclusion is that the 'marker group' is of much less assistance than a 'heavy' atom even though in the aggregate it outweighs the 'heavy' atom. The value of any sign or phase fixing component depends upon the concentration of electron density and hence in the directness of its influence on phase (see Mathieson (1955), Fig. 2). Although the tolyl group is convenient as a marker in that it is highly symmetrical, planar and rigid, its diffuseness renders it much less useful for structure analysis than the relatively light S atom.

Our thanks are due to Mr J. H. Weymouth of the Cement and Refractories Section of these laboratories for measurement of the main refractive indices and we are grateful to Mr R. G. Smart of the University of New South Wales for his assistance in data-handling on UTECOM.

References

Beecham, A. F. (1957). *J. Amer. Chem. Soc.* **79**, 3262.

Beecham, A.F., Fridrichsons, J. & Mathieson, A.McL. (1958). *J. Amer. Chem. Soc.* **80**, 4739.

Beevers, C. A. & Cochran, W. (1947). *Proc. Roy. Soc.* A, **190**, 257.

Berghuis, J., Haanappel, IJ. M., Potters, M., Loopstra, B. O., MacGillavry, C. H. & Veenendaal, A.L. (1955). *Acta Cryst.* **8**, 478.

Brown, C. J. (1951). *Acta Cryst.* **4**, 100.

Buerger, J. M. (1951). *Acta Cryst.* **4**, 531.

Buerger, J. M. & Hahn, T. (1957). *Z. Kristallogr.* **108**, 419.

Burge, R. E., Cowan, P. M. & McGavin, S. (1958). *Recent Advances in Gelatin and Glue Research*, p. 25. London: Pergamon Press.

Cochran, W. & Dyer, H. B. (1952). *Acta Cryst.* **5**, 634.

Corey, R. B. & Donohue, J. (1950). *J. Amer. Chem. Soc.* **72**, 2899.

Cowan, P. M. & McGavin, S. (1955). *Nature, Lond.* **176**, 1062.

Dawson, B. (1960*a*). *Acta Cryst.* **13**, 403.

Dawson, B. (1960*b*). *Acta Cryst.* **13**, 1034.

Donohue, J. & Trueblood, K. (1952). *Acta Cryst.* **5**, 419.

Freeman, H. C. (1957). *Aust. J. Chem.* **10**, 95.

Freeman, H. C. (1958). *Aust. J. Chem.* **11**, 99.

Fridrichsons, J. & Mathieson, A. McL. (1955). *Acta Cryst.* **8**, 761.

Fridrichsons, J. & Mathieson, A. McL. (1962). *Acta Cryst.* **15**, 119.

Furberg, S. (1950). *Acta Cryst.* **3**, 325.

Internationale Tabellen zur Bestimmung von Kristallstrukturen (1935). Vol. 2. Berlin: Borntraeger.

Kaan, G. & Cole, W. F. (1949). *Acta Cryst.* **2**, 38.

Kroner, T. D., Tabroff, W. & McGarr, J. J. (1955). *J. Amer. Chem. Soc.* **77**, 335.

Leung, Y. C. & Marsh, R. E. (1958). *Acta Cryst.* **11**, 17.

Lipson, H. & Cochran, W. (1953). *The Determination of Crystal Structures*. London: Bell.

Mathieson, A.McL. & Robertson, J.M. (1949). *J. Chem. Soc.*, p. 724.

Mathieson, A. McL. & Welsh, H. K. (1952). *Acta Cryst.* **5**, 599.

Mathieson, A. McL. (1952). *Acta Cryst.* **5**, 332.

Mathieson, A. McL. (1955). *Rev. Pure Appl. Chem.* **5**, 113.

Mathieson, A. McL. (1961). *Pure and Applied Chemistry.* **2**, 505.

Pasternak, R. A. (1956). *Acta Cryst.* **9**, 341.

Ramachandran, G. N. & Kartha, G. (1955). *Nature, Lond.* **176**, 59.

Rich, A. & Crick, F. H. (1955). *Nature, Lond.* **176**, 915.

Rougvie, M. A. & Bear, R. S. (1953). *J. Amer. Leath. Chem. Ass.* **48**, 735.

Sasisekharan, V. (1959). *Acta Cryst.* **12**, 903.

Schroeder, W. A., Kay, L. M., LeGette, J., Honnen, L. & Green, F. C. (1954). *J. Amer. Chem. Soc.* **76**, 3556.

Smits, D. W. & Wiebenga, E. H. (1953). *Acta Cryst.* **6**, 531.

Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M. & Katchalski, E. (1960). *J. Amer. Chem. Soc.* **82**, 5263.

Wilson, A. J. C. (1942). *Nature, Lond.* **150**, 152.

Yakel, H. L. & Hughes, E. W. (1954). *Acta Cryst.* **7**, 291.

Zussman, J. (1951*a*). *Acta Cryst.* **4**, 72.

Zussman, J. (1951*b*). *Acta Cryst.* **4**, 493.