Acta Cryst. (1952). **5**, 414

The Crystal Structure of Hydroxy-L-Proline. I. Interpretation of the Three-Dimensional Patterson Function*

By Jerry Donohue

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

and Kenneth N. Trueblood

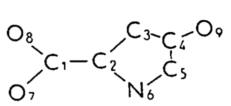
Department of Chemistry, University of California at Los Angeles, Los Angeles, California, U.S.A.

(*Received* 28 *August* 1951)

The methods used in preparing and interpreting the three-dimensional Patterson function of hydroxy-L-proline are described in some detail. No chemical assumptions were made during the interpretation. The structure obtained from this Patterson analysis was sufficiently accurate to confirm the chemical structure, and to be used in subsequent Fourier and least-squares refinements.

1. Introduction

The investigation of the crystal structures of various amino acids and their derivatives is part of a long-range program on the structure of proteins which is in progress at the California Institute of Technology. The determination of the crystal structure of hydroxy-L-proline



is a part of that program. The derivation of the correct trial structure involved the interpretation of a three-dimensional Patterson function by methods which should be of value in the determination of other structures of comparable complexity; since they apparently have not been discussed before it seems worth while to describe them in some detail in the present paper. The experimental details and final results of this structure determination, together with the discussion of the chemical implications, will be presented separately (Donohue & Trueblood, 1952). In the present paper it need only be mentioned that essentially all of the reflections accessible to Cu $K\alpha$ radiation were recorded photographically with a Weissenberg camera, and the intensities of the 646 non-equivalent reflections were then estimated by means of intensity strips and the multiple-film technique. The space group was found to be $P2_12_12_1$; there are four molecules in a unit cell with

$$a_0 = 5\cdot00,\ b_0 = 8\cdot31,\ \text{and}\ c_0 = 14\cdot20\ \text{Å}\,.$$

2. Preparation of the Patterson functions

The values of $|F|^2_{hkl}$ were placed on an absolute scale and the temperature factor B was determined by the method suggested by Hughes (see Shoemaker, Donohue, Schomaker & Corey, 1950, p. 2330). The values of $|F|^2_{hkl}$ were then sharpened (Patterson, 1935) by dividing each by the quantity $\sum_i f_i^2 \exp[-2B\sin^2\theta/\lambda^2]$, where f_i is the form factor of the ith atom (*Internationale Tabellen*, 1935, vol. 2, p. 571). The peak at the origin was then removed (Patterson, 1935), and the resulting, altered values of $|F|^2_{hkl}$ were multiplied by a modification function (Waser, 1944; Schomaker, 1944) $M_n(H) = H^n \exp[-\alpha^2H^2]$, where $H = 2\sin\theta/\lambda$, and, in this case, $n = 4$ and $\alpha = 2\cdot16$. This function (Schomaker, 1947; Shoemaker, 1947) is used to insure the convergence of the Fourier series; moreover, its use has a twofold advantage over that of an artificial temperature factor, i.e. $M_0(H)$: (1) the linear resolving power (as given by the sharpness of a peak) is greater, and (2) the data weighted most heavily are those at intermediate values of $\sin\theta$, and these intensities are generally more accurately estimated than those at low or high values of $\sin\theta$. One disadvantage of $M_4(H)$ is that in Patterson space each peak, having a shape given by the Fourier transform of $M_4(H)$, is surrounded by a shallow negative trough. In practice, however, this negative region is not very troublesome, and it is felt that the enhanced resolution, which increases considerably the number of peaks to work with, outweighs this disadvantage.

The modified $|F|^2_{hkl}$ data were used to compute the asymmetric unit of the Patterson function, $P(u, v, w)$. Simultaneously with this calculation, the untreated data were used to compute a second function, $P'(u, v, w)$. These computations were carried out with the assistance of Prof. V. Schomaker, using a punched-card method devised by him. About 28 man-hours were spent in making these computations, exclusive of plotting.

---

* Contribution No. 1626 from the Gates and Crellin Laboratories. This paper was presented in part at the Second International Congress of Crystallography, Stockholm, June 1951, and in part at the XIIth International Congress of Pure and Applied Chemistry, New York, September 1951.

Aided by a grant from The National Foundation for Infantile Paralysis.

The intervals in u, v, and w were $\frac{1}{20}$, $\frac{1}{30}$, and $\frac{1}{60}$, corresponding to 0·25, 0·28, and 0·24 Å respectively. The function $P'(u, v, w)$ was evaluated mainly in order to obviate any difficulties which might have arisen as a result of the small negative regions about each peak in $P(u, v, w)$ or as a result of shifts in peak positions due to possible error in the scale and temperature factor determination. Comparison of the plots of $P(u, v, w)$ with those of $P'(u, v, w)$ showed, however, that neither of these anticipated difficulties had materialized. In the work which followed, therefore, use was made almost exclusively of $P(u, v, w)$, the modified Patterson function, because of its enhanced resolution.

Inspection of the calculated octant of $P(u, v, w)$ showed 91 separate peaks, including several maxima which could be crudely resolved from obviously composite peaks. Since 171 peaks were to be expected if all interactions between the heavier (non-hydrogen) atoms were resolved, even the superior resolution of the modified Patterson function was sufficient to distinguish only about one-half of the expected peaks.

3. Interpretation

In a preliminary approach to the interpretation of the Patterson function it was assumed that the five-membered ring was nearly parallel to (100), as indicated by the short a axis. Peaks in the section $P(\frac{1}{2}, v, w)$ would then be expected to outline a projection of the ring. Considerable time was spent in trying (unsuccessfully, it turned out) to fit peaks in $P(\frac{1}{2}, v, w)$ with a five-membered ring in various orientations. The failure of this effort will be remarked upon later.

Two independent attacks were then made. The first of these constituted a direct attempt at analysis of the Harker sections. In the space group $P2_12_12_1$, we expect Harker peaks (those due to interactions between crystallographically equivalent atoms) at the following positions in $P(u, v, w)$: $(\frac{1}{2}, \frac{1}{2}-v, w)$, $(u, \frac{1}{2}, \frac{1}{2}-w)$, and $(\frac{1}{2}-u, v, \frac{1}{2})$, where u, v and w are each twice the respective coordinates x, y and z of an atom. It should therefore be possible to find peaks in all three sections which are consistent. Critical examination of the three Harker sections revealed seven possible atom positions, the restriction being made that all interactions lie in a region of height at least 10% that of the highest peak. Of these seven possibilities, only three were consistent with the remainder of $P(u, v, w)$ when interactions between non-equivalent atoms were considered. These three atoms were later identified as O_7, N_6 and O_9. It was then apparent that *all* positive regions, and perhaps even slightly negative regions, must be accepted in determining possible atom positions from Harker sections, and in checking the consistency of the positions with the remainder of $P(u, v, w)$. This is apparent from Fig. 2, 4 and 6, in which the Harker interactions are identified in each of the three Harker sections. Since the inclusion of all positive and slightly negative areas increases tremendously the number of possible positions, this line of approach was discontinued.

In the second approach to the interpretation of $P(u, v, w)$, use was made of a relation first pointed out by Albrecht & Corey (1939) and used by them in the determination of the crystal structure of glycine. It is based on the following considerations. The equivalent points in $P2_12_12_1$ are $M(x, y, z)$, $A(\frac{1}{2}+x, \frac{1}{2}-y, \bar{z})$, $B(\bar{x}, \frac{1}{2}+y, \frac{1}{2}-z)$, $C(\frac{1}{2}-x, \bar{y}, \frac{1}{2}+z)$. Considering for the moment only two atoms, these will give ten interactions in the asymmetric unit of $P(u, v, w)$. These we may designate as the Harker interactions M_1A_1, M_1B_1, M_1C_1, M_2A_2, M_2B_2, M_2C_2, and the non-Harker interactions M_1M_2, M_1A_2, M_1B_2, M_1C_2. Now if it is observed that a peak occurs in a zero-section of $P(u, v, w)$, say $P(u, v, 0)$, we may arbitrarily designate this as M_1M_2. We then find that in the corresponding Harker section, $P(u, v, \frac{1}{2})$, the three peaks M_1C_1, M_1C_2, and M_2C_2 lie on a straight line, the central (non-Harker) peak being separated from the other two by the same vector which separates M_1M_2 from the origin. Since the peak M_1C_2 lies accidentally on the Harker section, it has double weight. This relation between any non-Harker peak in a zero section and a triple of peaks in a Harker section is of course due to the operation of the screw axis.*

The application of this method is illustrated in Figs. 1–6, which show the three zero and Harker sections. These drawings were prepared in the following way: in the asymmetric unit of each section all interactions, calculated with the final parameters, are plotted; the size of the circle is roughly proportional to the weight of the interaction, this weight depending on two factors: the kind of atoms involved in the interaction, and the distance of the interaction from the section. In the case of the Harker sections additional Harker peaks outside the asymmetric unit are plotted to give a complete projection of one molecule; in this space group there are four such projections in each Harker section. In order to facilitate com-

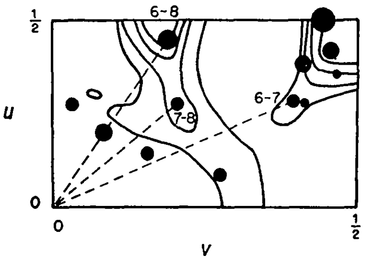


Fig. 1. Patterson section $P(u, v, 0)$.

---

* Stern & Beevers (1950) made use of this relation in their determination of the structure of tartaric acid. In this crystal five of the atoms lie in a plane perpendicular to a 2_1 axis (the space group is $P2_1$), a fact which made possible a relatively quick correlation between the zero and Harker sections, once it was realized that the Harker section alone was of little value.

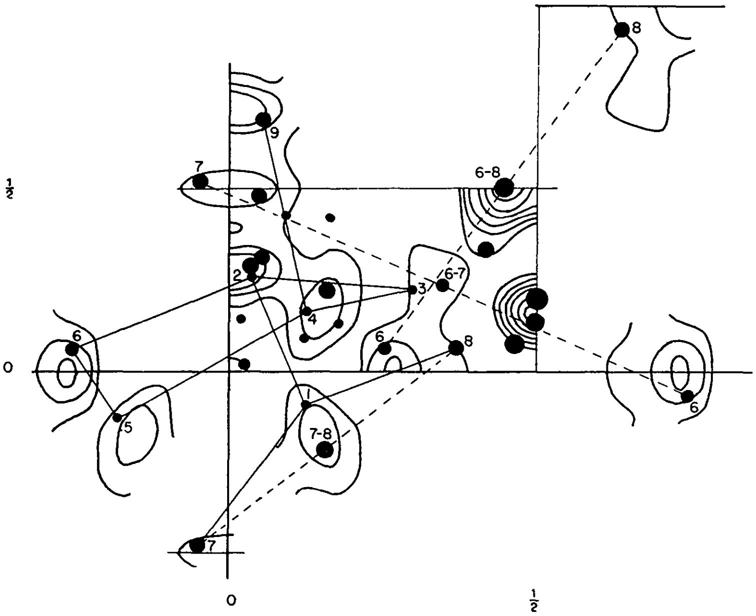


Fig. 2. Patterson section $P(u, v, \frac{1}{2})$.

parison of the zero sections with the Harker sections, the peaks comprising the triple in the Harker sections were chosen to be linear, where convenient; in some cases this leads to the use of peaks outside the asymmetric unit. There are four symmetry-equivalent linear triples in the Harker sections corresponding to four equivalent peaks in a complete zero section and the four molecular projections in the Harker section.

In Fig. 1 are shown the interactions in the section $P(u, v, 0)$. The corresponding interactions in $P(u, v, \frac{1}{2})$ are shown in Fig. 2. By suitable manipulation, it was possible to find three linear triples in $P(u, v, \frac{1}{2})$ in accord with peaks in $P(u, v, 0)$. This procedure is easily accomplished by tracing the Harker section to extend its range on all sides of the asymmetric unit.

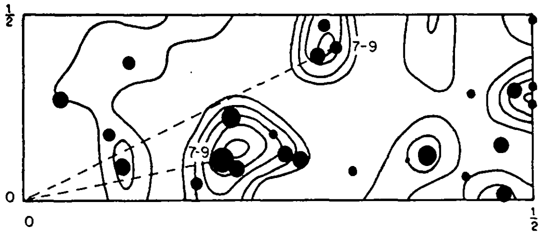


Fig. 3. Patterson section $P(u, 0, w)$.

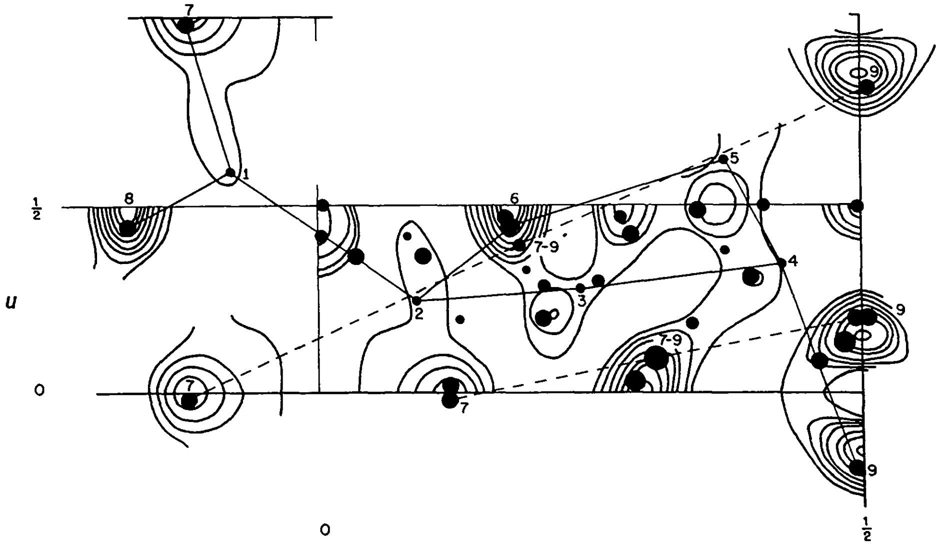


Fig. 4. Patterson section $P(u, \frac{1}{2}, w)$.

The positions of the three peaks of the triple, as given by the zero section, are plotted on a piece of tracing paper which is then moved about over the Harker section until a promising fit is obtained. This consistency alone is not enough to establish beyond doubt that the relations between the positions of these peaks are not mere coincidences; it is necessary also to check other interactions corresponding to the postulated atomic positions. In this way several linear triples consistent with $P(u, v, \frac{1}{2})$ were shown to be spurious.

Only two authentic linear triples were found in $P(u, \frac{1}{2}, w)$, Fig. 4, which were consistent with peaks in $P(u, 0, w)$, Fig. 3. For these, an interesting feature arises: these two triples use the same Harker peaks; therefore only two atoms are involved, and thus their y coordinates, in addition to being nearly the same, must both be near 0, $\frac{1}{4}$, or $\frac{1}{2}$. This fact is useful in identifying the peaks in the other Harker sections with these same atoms.

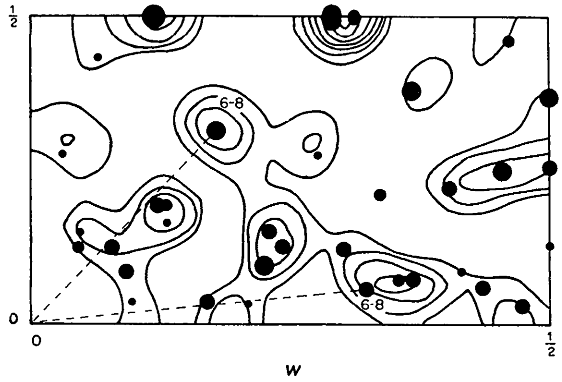


Fig. 5. Patterson section $P(0, v, w)$.

In the case of $P(0, v, w)$, Fig. 5, and $P(\frac{1}{2}, v, w)$, Fig. 6, again only two triples were found, and these again were seen to be due to the same pair of atoms. Special attention should be paid to Fig. 6 in connection with the failure of the earlier attempts to fit a five-membered ring to its peaks. It is obvious from inspection of this figure that any procedure, e.g. the so-called 'implication' method (Buerger, 1946), which depends on the identification of Harker peaks as such, or at least on the appearance of these interactions in regions of relatively high density, is of little aid in structure determinations of this type. It is perhaps worth pointing out also that the decision as to whether or not there is a 2 or 2_1 axis parallel to a_0 (Buerger, 1950) could scarcely be made with confidence by comparing the density of peaks in $P(0, v, w)$ with $P(\frac{1}{2}, v, w)$. The lack of a strikingly higher density in the Harker section compared with the zero section is to be expected when the sections are perpendicular to a short axis. It is interesting to note that this criterion does hold for the sections perpendicular to the long axis (cf. $P(u, v, 0)$ with $P(u, v, \frac{1}{2})$, Figs. 1 and 2).

Comparison of the various interactions which were designated as Harker peaks in the seven different triples discussed above showed that only four atoms needed to be placed to account for them. It was then possible to assign x, y and z parameters to all four of them, and then to calculate the additional interactions which should arise. Twelve of these additional interactions fall near the section $P(\frac{1}{4}, v, w)$, shown in Fig. 7. The agreement between all thirty-six calculated interactions and $P(u, v, w)$ was considered satisfactory, so that some confidence could now be placed in the parameters assigned to these four atoms. Slight

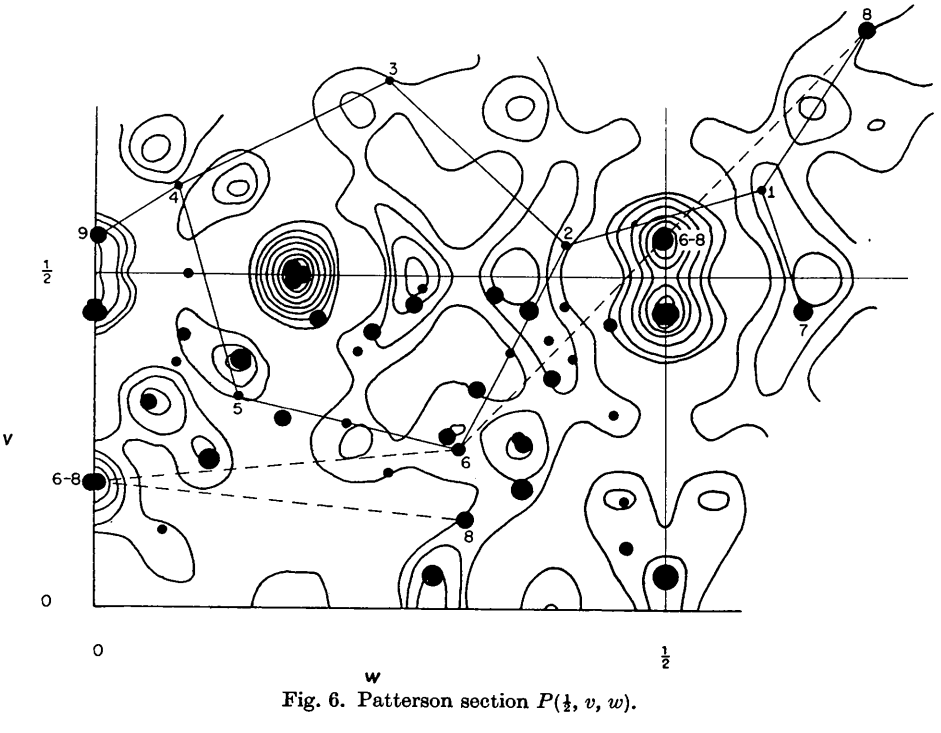


Fig. 6. Patterson section $P(\frac{1}{2}, v, w)$.

Table 1. *Bond distances and angles in hydroxy-*L*-proline*

Distance	$P(u, v, w)$ parameters	Final parameters	Angle	$P(u, v, w)$ parameters	Final parameters
C_2–C_3	1·43 Å	1·53 Å	O_9–C_1–O_8	121°	126°
C_3–C_4	1·66	1·50	O_7–C_1–C_2	114	119
C_4–C_5	1·30	1·52	O_8–C_1–C_2	124	115
C_5–N_6	1·42	1·48	C_1–C_2–C_3	100	113
N_6–C_2	1·57	1·50	C_1–C_2–N_6	114	111
C_1–C_2	1·47	1·52	N_6–C_2–C_3	97	105
C_1–O_7	1·33	1·25	C_2–C_3–C_4	100	108
C_1–O_8	1·18	1·27	C_3–C_4–C_5	100	104
C_4–O_9	1·38	1·46	C_4–C_5–N_6	107	105
$O_9 \cdots O_7$	2·84	2·80	C_5–N_6–C_2	110	109
$N_6 \cdots O_8$	2·80	2·69	C_3–C_4–O_9	119	106
$N_6 \cdots O_8'$	3·02	3·17	C_5–C_4–O_9	117	109

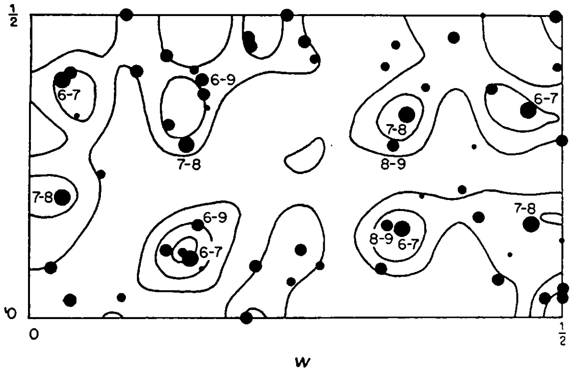


Fig. 7. Patterson section $P(\frac{1}{4}, v, w)$.

adjustments were made in some of the parameters in order to improve the agreement somewhat.

Probable identification of the four atoms as N_6, O_7, O_8 and O_9 could now be made. In the case of O_7 and O_8 this identification was certain because one of the interactions—the one at the extreme left of Fig. 7—fell 2·25 Å from the origin, the expected oxygen–oxygen distance in a carboxyl group; for placing the other atoms, other structural considerations were possible. Nevertheless, it was thought that it would be more desirable to proceed without any chemical assumptions, partly in order to test the power of the three-dimensional Patterson method, and partly because one of the objects of the investigation was to fix the stereochemical configuration of hydroxyproline. The vector convergence method (Beevers & Robertson, 1950) was then applied. Atoms 6 and 7 were chosen as the 'heavy atoms' at whose four equivalent positions the origin of $P(u, v, w)$ was placed in the four successive transcriptions which made up the vector convergence diagram. The interpretation of the vector convergence diagrams was not as straightforward as one might hope—more than a dozen fourfold overlaps or near overlaps were found. Of these, five gave atomic positions which were not grossly inconsistent with $P(u, v, w)$ and, furthermore, these five agreed remarkably closely with completely independent estimates of the probable positions of the five missing atoms which had been predicted on the basis of the identification of the first four atoms as N_6, O_7, O_8 and O_9.

4. Results

Before the refinement procedures which are described in Part II (Donohue & Trueblood, 1952) were undertaken, some of the atomic parameters were adjusted to give more reasonable interatomic distances. It is of interest, however, to compare the parameters obtained directly from the Patterson function with the final parameters which resulted from three-dimensional least-squares and Fourier refinements. This comparison shows that the average difference in a parameter is 0·07 Å, the maximum difference 0·42 Å. It is perhaps more interesting to make a corresponding comparison of interatomic distances and interbond angles; this is done in Table 1. The average discrepancy in the twelve distances of Table 1 is 0·10 Å; for the twelve angles the average discrepancy is 7°. In general, it is probably true that a relatively complete interpretation of a three-dimensional Patterson function is sufficient to determine only the gross features of a structure with some degree of assurance, but that in order to make detailed comments concerning interatomic distances, bond angles and other fine points of structural interest, the conventional Fourier and/or least-squares refinement are necessary.

References

Albrecht, G. A. & Corey, R. B. (1939). *J. Amer. Chem. Soc.* **61**, 1087.

Beevers, C. A. & Robertson, J. H. (1950). *Acta Cryst.* **3**, 164.

Buerger, M. J. (1946). *J. Appl. Phys.* **17**, 579.

Buerger, M. J. (1950). *Acta Cryst.* **3**, 87.

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

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

Patterson, A. L. (1935). *Z. Krystallogr.* **90**, 517.

Schomaker, V. (1944). Private communication.

Schomaker, V. (1947). Private communication.

Shoemaker, D. P. (1947). Ph.D. Thesis, California Institute of Technology.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 2328.

Stern, F. & Beevers, C. A. (1950). *Acta Cryst.* **3**, 341.

Waser, J. (1944). Ph.D. Thesis, California Institute of Technology.