Acta Cryst. (1976). B32, 11

The Crystal and Molecular Structure of α-Glycine by Neutron Diffraction – a Comparison

BY L. F. POWER AND K. E. TURNER

James Cook University of North Queensland, Townsville, Australia, 4811

AND F. H. MOORE

Australian Institute of Nuclear Science and Engineering, Lucas Heights, Australia

(*Received* 18 *November* 1974; *accepted* 25 *April* 1975)

The crystal and molecular structure of α-glycine by neutron diffraction was determined simultaneously by the authors and by Jönsson & Kvick [*Acta Cryst.* (1972), B**28**, 1827–1833]. As the two determinations were of comparable precision, a detailed comparative study was undertaken. The comparison indicates that the derived parameters probably do not belong to the same population, whereas the intensity data sets have been shown to belong to the same population. Reasons for the statistical variations between the derived parameter sets have been investigated.

Introduction

At the commencement of this structural determination of α-glycine, only one preliminary report (Burns & Levy, 1958) of the neutron structural determination of any of the naturally occurring amino acids had appeared in the literature. The desirability of obtaining a precise knowledge of hydrogen bonding in biologically important molecules is obvious, and it was decided to begin a systematic study of the important amino acids by neutron diffraction. Towards the completion of our own determination of the structure of α-glycine by neutron diffraction, the results of a similar study were published by Jönsson & Kvick (1972). Since the two data sets were of similar precision and had been refined with the same full-matrix least-squares refinement program (*LINUS*) the opportunity for a detailed comparative study was taken.

Experimental

Large single crystals were grown by the slow cooling of an aqueous solution of glycine. A crystal of 30 mm^3 volume was selected and mounted with the b crystallographic axis approximately parallel to the φ axis of the Australian Institute of Nuclear Science and Engineering's computer-controlled four-circle diffractometer 2TanA. Two sets of equivalent reflexions, comprising 2640 observations, were collected with the ω–2θ scan technique. A standard reflexion, monitored every 25 observations, showed no systematic variation during the course of the data collection.

The neutron wavelength was 0·981 Å, and the calculated linear absorption coefficient for neutrons was $2{\cdot}24 \pm 0{\cdot}05$ cm^{-1}. Crystal data, taken from the accurate X-ray study of Marsh (1958), are as follows. α-Glycine $H_3N^+.CH_2.COO^-$; monoclinic; space group $P2_1/n$; $a = 5{\cdot}1020(8)$,* $b = 11{\cdot}9709(17)$, $c = 5{\cdot}4575(15)$ Å; $\beta = 111{\cdot}705(17)°$; $Z = 4$.

The data were corrected for absorption and Lorentz effects and averaged to yield a unique set of 1208 reflexions, all of which were included in the least-squares refinement. The absorption correction calculation was carried out by Gaussian integration with the crystal shape defined by 14 boundary planes. The structure was refined with the Brookhaven least-squares program *LINUS*. The function minimized was $\sum w(|F_c^2| - |F_o^2|)^2$ where w, the weight of each reflexion, is defined as $w^{-1} = \sigma^2(F_o^2)$ and $\sigma^2(F_o^2) = \sigma_1^2(F_o^2) + \sigma_2^2(F_o^2)$; $\sigma_1^2(F_o^2)$ is that part of the variance arising from Poisson counting statistics and $\sigma_2^2(F_o^2)$ is a polynomial of the form $\alpha + \beta F_o^2 + \gamma F_o^4$, and is that part of the variance arising from noncounting sources. The coefficients of the polynomial were determined from an analysis of the equivalent reflexions and have the values $\alpha = -0{\cdot}355 \times 10^{-1}$, $\beta = 0{\cdot}158 \times 10^{-1}$, $\gamma = -0{\cdot}250 \times 10^{-4}$.

Structure refinement

Initial parameters for all atoms were those of Marsh (1958). Refinement based on F_o^2 was carried out in the usual manner. All atoms were assigned anisotropic thermal parameters and together with an isotropic extinction parameter were allowed to vary in the final least-squares cycles. Final agreement values are listed in Table 1. Jönsson & Kvick found that the inclusion of six anisotropic extinction parameters significantly improved the refinement. Accordingly this approach, using the unaveraged data set, was tried. However, no improvement was obtained for either type I (mosaic-spread dominated) or type II (particle-size dominated)

* Here and throughout this paper, values in parentheses are the estimated standard deviations of the least significant digits.

anisotropic extinction refinement. The isotropic extinction parameter g' [using the notation of Coppens & Hamilton (1970)] was 1·02(2). This value is almost four times the value obtained by Jönsson & Kvick (1972). Hence the extinction properties of the two crystals used are quite different.

Table 1. *Final agreement indices for α-glycine*

	$R(F_o^2)$	$R_w(F_o^2)$	$R(F_o)$	S	$(m-n)$
This investigation	0·050	0·067	0·045	1·43	1116
Jönsson & Kvick (1972)					
Isotropic extinction	0·034	0·0446	–	2·03	665
Type I anisotropic extinction	0·030	0·0422	0·032	1·93	660
Type II anisotropic extinction	0·030	0·0422	0·032	1·93	660

$$R(F_o^2)=\sum||F_o^2|-|F_c^2||/\sum|F_o^2|$$
$$R_w(F_o^2)=[\sum w(|F_o^2|-|F_c^2|)^2/\sum w|F_o^4|]^{1/2}$$
$$R(F_o)=\sum||F_o|-|F_c||/\sum|F_o|$$
S, the standard deviation of unit weight
$$=[\sum w(|F_o^2|-|F_c^2|)^2/(m-n)]^{1/2}$$

where m is the total number of observations and n is the number of parameters varied.

Final positional and thermal parameters, together with those of Jönsson & Kvick, are shown in Table 2. The notation is that of Jönsson & Kvick and is shown in Fig. 1. Neutron scattering lengths used were C = 0·665, N = 0·940, O = 0·577 and H = −0·372. Bond lengths and angles are summarized in Table 3 while final observed and calculated structure factors are shown in Table 4.

Comparison of results

It was felt that before embarking on a comparison of the two structural determinations, the present data should be rigorously examined. Accordingly each pair of equivalent reflexions was examined for agreement in terms of that expected for a normal distribution. The

Table 3. *Bond lengths and angles for α-glycine*

	This investigation	Jönsson & Kvick
C(1)–O(1)	1·251 (1) Å	1·250 (1) Å
C(1)–O(2)	1·252 (1)	1·251 (1)
C(2)–N	1·475 (1)	1·476 (1)
C(1)–C(2)	1·525 (1)	1·526 (1)
N—H(1)	1·053 (2)	1·054 (2)
N—H(2)	1·036 (2)	1·037 (2)
N—H(3)	1·029 (2)	1·025 (2)
C(2)–H(1)	1·094 (2)	1·090 (2)
C(2)–H(2)	1·087 (2)	1·089 (2)
C(2)–C(1)–O(1)	117·38 (7)°	117·46 (6)°
C(2)–C(1)–O(2)	117·08 (7)	117·09 (7)
O(1)–C(1)–O(2)	125·53 (8)	125·45 (8)
C(1)–C(2)–N	111·79 (6)	111·85 (5)
C(2)–N—H(1)	112·08 (11)	112·09 (10)
C(2)–N—H(2)	112·00 (12)	111·73 (12)
C(2)–N—H(3)	110·22 (14)	110·37 (12)
H(1)–N—H(2)	109·05 (16)	108·71 (15)
H(1)–N—H(3)	107·02 (17)	107·13 (16)
H(2)–N—H(3)	106·20 (19)	106·56 (17)
C(1)–C(2)–H(4)	108·55 (13)	108·81 (12)
C(1)–C(2)–H(5)	110·38 (13)	110·50 (11)
N—C(2)–H(4)	108·41 (13)	108·51 (12)
N—C(2)–H(5)	109·47 (13)	109·05 (12)
H(4)–C—H(5)	108·14 (21)	108·03 (18)

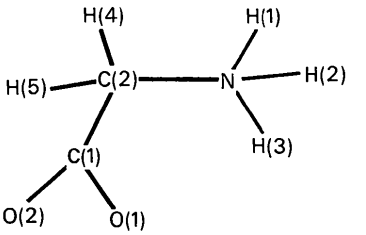


Fig. 1. Numbering scheme for α-glycine.

Table 2. *Final positional and thermal parameters for α-glycine*

The positional parameters are given as fractional coordinates (× 10⁵). The anisotropic thermal parameters (× 10⁵) are defined as $\exp[-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+2\beta_{12}hk+2\beta_{13}hl+2\beta_{23}kl)]$. For each atom our parameters are given on the first line and those of Jönsson & Kvick (1972) on the second.

	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
O(1)	30522 (16)	9430 (7)	23535 (15)	2326 (28)	549 (5)	1314 (22)	216 (9)	788 (18)	133 (8)
	30494 (17)	9439 (9)	23539 (16)	2050 (32)	521 (7)	1181 (28)	207 (12)	786 (24)	120 (11)
O(2)	−14752 (17)	14174 (8)	10725 (17)	2114 (28)	654 (6)	2058 (28)	33 (10)	1246 (21)	−103 (10)
	−14722 (17)	14150 (10)	10708 (18)	1791 (32)	639 (8)	1873 (33)	35 (13)	1141 (27)	−117 (13)
N	30128 (10)	8992 (4)	−25897 (9)	2240 (18)	463 (3)	1293 (14)	189 (5)	835 (11)	59 (5)
	30116 (10)	*8984 (5)*	*−25904 (10)*	*1945 (22)*	*448 (5)*	*1162 (19)*	*177 (7)*	*771 (16)*	*49 (7)*
Cl	7495 (12)	12484 (5)	6613 (12)	1815 (21)	293 (3)	1384 (18)	−27 (6)	812 (14)	−26 (5)
	7504 (13)	12486 (6)	6619 (13)	1552 (25)	274 (4)	1193 (24)	−34 (8)	735 (19)	−23 (8)
C(2)	6459 (13)	14482 (5)	−21332 (12)	2008 (22)	365 (4)	1344 (18)	120 (6)	749 (15)	118 (6)
	6474 (13)	14485 (7)	−21308 (13)	1696 (26)	356 (5)	1227 (24)	120 (9)	682 (20)	111 (9)
H(1)	28950 (34)	10012 (16)	−45449 (30)	3403 (59)	747 (12)	1948 (47)	187 (20)	1380 (41)	118 (18)
	28972 (34)	10036 (16)	−45414 (31)	3267 (63)	718 (15)	1768 (58)	185 (24)	1253 (49)	122 (22)
H(2)	49508 (33)	11880 (17)	−13110 (32)	2783 (54)	871 (14)	2440 (52)	125 (21)	1033 (42)	−39 (20)
	49450 (33)	11929 (18)	−13184 (34)	2450 (60)	863 (17)	2260 (62)	72 (24)	834 (52)	−72 (24)
H(3)	29914 (43)	537 (14)	−22566 (35)	5204 (87)	534 (10)	2743 (57)	504 (24)	1504 (56)	101 (18)
	29935 (42)	561 (17)	−22613 (37)	5059 (90)	510 (13)	2713 (71)	555 (27)	1538 (64)	123 (23)
H(4)	7780 (46)	23474 (14)	−24239 (39)	5752 (95)	466 (9)	3663 (70)	424 (24)	2527 (68)	423 (20)
	7688 (46)	23444 (16)	−24322 (41)	5633 (98)	444 (13)	3698 (79)	439 (28)	2677 (73)	416 (25)
H(5)	−13379 (33)	11484 (20)	35697 (33)	2589 (55)	1084 (17)	2252 (51)	6 (24)	553 (41)	−168 (23)
	−13322 (33)	11439 (20)	35718 (35)	2428 (59)	1055 (20)	2214 (64)	−65 (27)	538 (51)	−226 (26)

Table 4. *Observed and calculated structure factors for α-glycine*

Column headings are k, l, $F_o^2 \times 10$, $F_c^2 \times 10$ and $\sigma(F_o^2) \times 10$.

The data may be brought to the absolute scale by the application of a scale factor (k) of the form kF^2 absolute $= F^2$ table, where $k = 2{\cdot}225$ (8). The data have been corrected for absorption.

results are detailed in Table 5. In addition, a normal probability plot (Abrahams & Keve, 1971) was constructed for the two equivalent sets of data. The plot was linear, passed through the origin and had a slope of about 1·06. On the basis of this plot it may be concluded with some confidence that there is no systematic difference between the equivalent data sets.

Table 5. *Agreement between equivalent reflexions for α-glycine*

Range*	Percentage of total	Expected value
0–0·5	41·0	38·3
0–1·0	67·8	68·3
0–2·0	91·8	95·45
0–3·0	97·7	99·73
Rejected (>3·0)	1·4	–

* Number of estimated standard deviations from the mean of each member of a group of equivalent reflexions.

δR plots were constructed for the present data set and that of Jönsson & Kvick. In each case the data were corrected for extinction. F_c values calculated from the present data were used for the δR plot of the present data while the F_c values published by Jönsson & Kvick (1972) were used in the δR plot of their data. In each case a substantially linear plot, passing through the origin, was obtained. The slopes of the linear portions were 1·35 (this investigation) and 1·68 (Jönsson & Kvick). The amount by which the slope of a δR plot exceeds unity gives a measure of the underestimation of individual estimated standard deviations of F_o. It is suggested on the basis of the δR plots that neither refined data set contains bias but that the estimation of the standard deviation of F_o is poorer in the case of Jönsson & Kvick than in the present case. This is not a surprising result as Jönsson & Kvick have adopted an empirical approach to the estimation of non-counting errors in the data while in this determination an attempt to estimate these errors from an analysis of the agreement between equivalent reflexions has been made. The failure of either δR plot to achieve unit slope was not surprising since the δR plots, and indeed the least-squares refinements, have assumed F_c to be exact. This is clearly an invalid assumption. F_c contains many possible errors, principally the approximations made in assuming the atomic thermal motion to be ellipsoidal and the experimental error associated with the neutron scattering lengths.

A full normal probability plot was constructed with the 807 extinction corrected observed structure factors common to both the present data set and that of Jönsson & Kvick (Abrahams & Keve, 1971). The normal order statistic δm_i, is defined as $\delta m_i = [F(1)_i - KF(2)_i]/[\sigma^2 F(1)_i + K^2\sigma^2 F(2)_i]^{1/2}$, where $F(1)_i$ and $F(2)_i$ are the ith observed structure factors of data sets (1) and (2), and have associated with them estimated standard deviations of $\sigma F(1)_i$ and $\sigma F(2)_i$ respectively. K, the scale factor between data sets (1) and (2), is calculated so that $\sum_{i=1}^{N} \delta m_i^2$ is a minimum (Abrahams & Keve, 1971). The individual δm_i are arranged into descending order of magnitude and plotted against those values expected for a normal distribution. The plot, which is shown in Fig. 2, is linear, passes through the origin and has a slope of about 1·43.

The plot is thus consistent with the following conclusions: (i) the calculated scale factor between the data sets in satisfactory; (ii) the two data sets are from the same population, and (iii) the δm_i are normally distributed.

The deviation of the slope of the plot from unity may be attributed to one or both of the following causes: (i) the pooled standard deviations of δm_i are too small; (ii) the difference between the scaled data sets is larger than expected, *i.e.* one or both data sets are systematically biased. If both data sets are biased then both sets must be biased in the same sense.

In view of the individual δR plots, which are consistent with an underestimation of the estimated standard deviation of F_o for each data set, it is more probable that an underestimation of the pooled standard deviation of δm_i is the major reason for the departure of the normal probability plot from unit slope. Hence it is reasonable to conclude that the two data sets are not systematically different although the pooled standard deviations are underestimated by at least 43 %.

The atomic parameters derived from the above data sets were examined by means of half-normal probability plots (Abrahams & Keve, 1971) and χ^2 tests (Hamilton, 1969).

The half-normal probability plot, which is a convenient method by which to examine the relationship between derived parameters, is constructed as follows. The statistic δp_i, as defined below, is ordered and plotted against those values expected for a half-normal population.

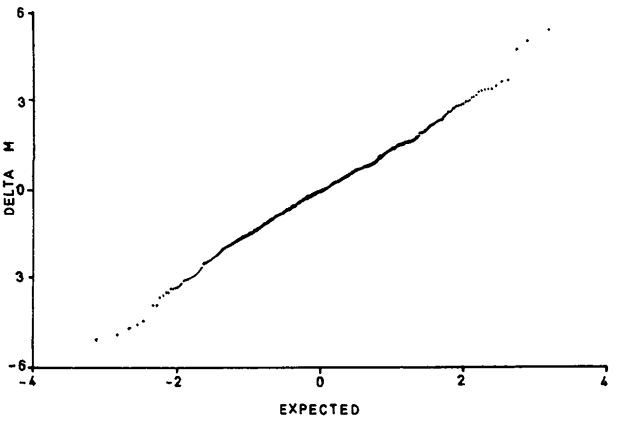


Fig. 2. Normal probability plot for the observed structure factors common to each data set.

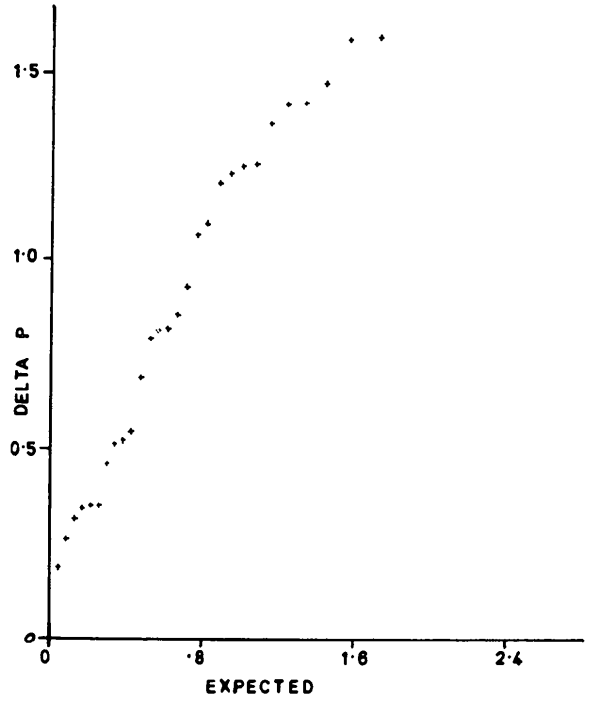
$$\delta p_i = ||p(1)_i| - |p(2)_i||/[\sigma^2 p(1)_i + \sigma^2 p(2)_i]^{1/2}$$

where $p(1)_i$ and $p(2)_i$ are the ith parameters, derived from the sets of $F(1)$ and $F(2)$ structure factors. $p(1)_i$ and $p(2)_i$ have estimated standard deviations of $\sigma p(1)_i$ and $\sigma p(2)_i$ respectively.

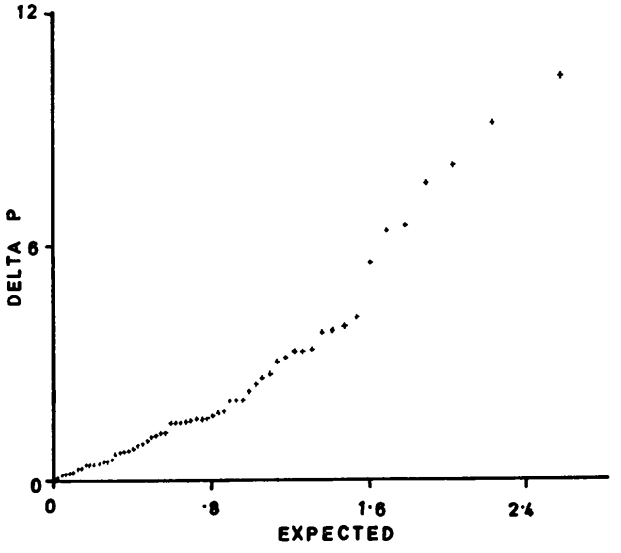
Differences between individual sets of parameters (*e.g.* all x atomic coordinates) may be further examined by means of the χ^2 test. The weighted difference R_i for the ith parameters defined above, may be calculated as

$$R_i = [p(1)_i - p(2)_i]/[\sigma^2 p(1)_i + \sigma^2(2)_i]^{1/2} .$$

The function $R^2 = \sum_{i=1}^{N} R_i^2$, where N is the number of parameters, may be computed for each type of derived parameter (*e.g.* all x atomic parameters) and tested as χ^2. The half-normal probability plots for positional and thermal data are shown in Fig. 3(*a*) and (*b*) respectively, while the χ^2 tests are summarized in Table 6.



(*a*)



(*b*)

Fig. 3. Half-normal probability plots for α-glycine, (*a*) positional, (*b*) thermal parameters.

Table 6. χ^2-*Tests for α-glycine positional and thermal parameters*

Positional parameters			
R_x^2	10·0		
R_y^2	15·2		
R_z^2	9·2		
$\chi^2_{2\cdot5,30}$	46·98		
$\chi^2_{97\cdot5,30}$	16·79		
Thermal parameters			
$R^2_{\beta 11}$	381·9	$R^2_{\beta 12}$	11·6
$R^2_{\beta 22}$	43·7	$R^2_{\beta 13}$	53·4
$R^2_{\beta 33}$	112·2	$R^2_{\beta 23}$	8·1
$\chi^2_{2\cdot5,60}$	83·30	$\chi^2_{97\cdot5,60}$	40·48

The half-normal probability plot for the positional parameters, using the exact values of the half-normal order statistics (Hamilton & Abrahams, 1972), is substantially linear with a slope close to unity. However, the plot does not pass through the origin. The plot thus suggests that the positional parameters derived from the two data sets do not belong to the same population. The calculated values of R^2 (Hamilton, 1969) for the x and z parameters are significantly below the 97·5% probability level of χ^2. This is consistent with the differences between the two sets of parameters being less than that expected for a normal distribution. Hence the results of the χ^2 tests suggest that the x and z parameters derived from the two data sets do not belong to the same population. The R^2 values of the y parameters are sufficiently close to the 97·5% probability level of χ^2 to be considered to belong to the same population.

The thermal parameters may also be considered to come from different populations. The half-normal probability plot for the thermal parameters is markedly non-linear, although the plot passes through the origin, and only the R^2 values of the β_{22} and the β_{13} parameters lie within the 95% probability limits of χ^2.

Discussion

The value of detailed comparisons of independent data sets has been clearly demonstrated. The two data sets, at first examination, are almost identical if only bond

lengths and angles are examined. However, a deeper examination has revealed that while the two independent structure factor sets are self consistent and can be concluded to come from the same population, there is evidence to suggest that the parameters derived from the final refinement of the data should be considered to belong to different populations.

While the derived parameters may be considered to come from different populations, this is only physically shown in the thermal parameters. The positional parameters derived from either data set are virtually identical. Hence the present structural determination of α-glycine is in complete agreement with the molecular geometry and hydrogen bonding reported by Jönsson & Kvick (1972). The thermal parameters, however, show definite systematic differences between the two data sets.

Since the δR plots have indicated that the individual observed structure factor sets may be considered to belong to the same population, the statistical differences indicated between the derived parameter sets must arise through differences in the least-squares refinement. The refinements of the individual observed data sets differ in two aspects.

(*a*) Jönsson & Kvick (1972) excluded from the refinement all data for which $F_o^2 \leq 2\sigma(F_o^2)$ while all data has been included in this investigation.

(*b*) The extinction properties of the two crystals differ markedly.

Hirshfeld & Rabinovich (1973) have pointed out that the exclusion of weak reflexions from the least-squares refinement will bias the data set and inevitably lead to systematic errors in the derived parameters. This will be especially true for the thermal parameters, which depend strongly on the average intensities of many weak reflexions. The magnitude of this effect however is quite small. In a test refinement of the present data excluding all observations for which $F_o^2 \leq 2\sigma(F_o^2)$ the maximum shift in any derived parameter was $0{\cdot}5\sigma$. Most parameters shifts were of the order of $0{\cdot}2\sigma$. These findings are in agreement with those of Stenkamp & Jensen (1975). While the Hirshfeld–Rabinovich effect must contribute to the observed statistical differences between the derived parameter sets the magnitude of the effect is much too small to explain the observed statistical differences.

Therefore, it must be concluded that the observed statistical differences between the derived parameter sets must arise, in the main, from the differences in extinction behaviour between the two crystals. The strong correlation between the thermal parameters and the extinction function is well known. Therefore, it is reasonable to expect that the thermal parameters derived from the two structure determinations could differ significantly since their extinction behaviour differs markedly. This probably represents a weakness in the model rather than any real difference between the thermal parameters derived from either data set.

One of us (KET) acknowledges the tenure of an Australian Institute of Nuclear Science and Engineering Postgraduate Studentship during the course of this work; another (LFP) acknowledges the provision of facilities and funds by the Australian Institute of Nuclear Science and Engineering and financial support from the Australian Research Grants Council. We would like to express our thanks to Dr Jönsson for the provision of the additional data which made this comparison possible.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157–165.

BURNS, J. H. & LEVY, H. A. (1958). *Abstracts, Amer. Cryst. Assoc.* p. 38. Marquette Univ., Milwaukee, Wisconsin.

COPPENS, P. & HAMILTON, W. C. (1970). *Acta Cryst.* A**26**, 71–83.

HAMILTON, W. C. (1969). *Acta Cryst.* A**25**, 194–206.

HAMILTON, W. C. & ABRAHAMS, S. C. (1972). *Acta Cryst.* A**28**, 215–218.

HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A**29**, 510–513.

JÖNSSON, P.-G. & KVICK, A. (1972). *Acta Cryst.* B**28**, 1827–1833.

MARSH, R. E. (1958). *Acta Cryst.* **11**, 654–663.

STENKAMP, R. E. & JENSEN, L. H. (1975). Personal communication.