

Test for Semi-isomorphism using Higher Moments of Intensities*†

BY S. PARTHASARATHY AND M. N. SABESAN

Centre of Advanced Study in Physics, University of Madras, Madras-25, India

(Received 4 December 1970 and in revised form 25 August 1971)

The multiple isomorphous replacement method of determination of non-centrosymmetric crystal structures is the one often used in protein crystallography. For this we require a minimum of three crystals, namely (i) the native protein (ii) the native protein with a group of heavy atoms and (iii) the native protein with a group of heavy atoms at locations different from those in crystal (ii). The isomorphism existing between crystals (ii) and (iii) is called semi-isomorphism. In this paper we derive higher moments of simple functions of normalized X-ray intensities from two semi-isomorphous crystals and show how these can be used to test the isomorphism of the common group of atoms in the two crystals. The theoretical results are also verified using a few hypothetical structures.

1. Introduction

The method of using the higher moments of the sum, difference and product of normalized intensities (denoted respectively by z_+ , z_- and z_x) from a pair of crystals to test for the isomorphism of the pair has been discussed recently by Parthasarathy & Ramachandran (1966) (this paper will be hereafter referred to as PR, 1966) and Parthasarathy (1968). The isomorphism dealt with in this paper (PR, 1966) are the simple replacement and the addition types. However, another kind of isomorphism, which is of particular interest in protein crystallography, occurs in the multiple isomorphous replacement method of phase determination for the non-centrosymmetric crystals (Harker, 1956). It is well known that the theory of multiple isomorphous replacement technique requires at least three isomorphous crystals such that the first crystal contains the native protein (called the Q group), the second crystal contains, besides the Q group, a group of one or more of heavy atoms (called the $P(1)$ group) and the third crystal contains, besides the Q atoms, a group of heavy atoms (called the $P(2)$ group) at locations different from those in the second crystal. The isomorphism which exists between the first and second crystals or the first and third crystals is the usual addition type and the method of distinguishing such a pair from a non-isomorphous pair has been discussed in the earlier paper (PR, 1966). The second and the third crystals are not strictly isomorphous, because of the presence of different $P(1)$ and $P(2)$ groups in the two crystals, and they are not strictly non-isomorphous because they have common group of Q atoms in the two crystals. We shall therefore describe such a type of isomorphism existing between the second and the

third crystals as 'semi-isomorphism'. A statistical test for semi-isomorphism could be useful as a preliminary step in the multiple isomorphous replacement method, since it offers an additional confirmation of the isomorphism of the Q group as obtained from the tests* conducted using data from pairs 1 and 2 and pairs 1 and 3. In this paper, we shall obtain the expressions for the higher moments of the normalized intensity variables z_+ , z_- and z_x and show how they could be used to test the isomorphism of the Q group of atoms in the semi-isomorphous pair of crystals. Since the necessity for using a pair of semi-isomorphous crystals does not arise in the case of centrosymmetric crystals, our main consideration in this paper will be for non-centrosymmetric crystals only.

The notation used in this paper is somewhat different from that used in the earlier paper (PR, 1966) and this is described in § 2. The theoretical expressions of the higher moments of the intensity variables z_+ , z_- and z_x are obtained in § 3 for the semi-isomorphous pair. A brief discussion of the theoretical results is given in § 4.

2. Definition of SI † and NI pairs and the description of the notation

Following the notation in the earlier paper‡ (PR, 1966) let us denote the contents of the unit cell of a pair of crystals (both the crystals of the pair must have the same unit-cell dimensions and space group symmetry) as follows:

* From a computational point of view this test is not difficult to perform, since the normalized intensities of the two derivatives and the native protein will be already available while performing the statistical tests described in the previous paper (PR, 1966).

† In this paper we shall use the abbreviation I for an isomorphous pair (PR, 1966), SI for a semi-isomorphous pair and NI for a non-isomorphous pair of crystals.

‡ Any subscript within parentheses refers to the number of the isomorphous derivative; for example $r_{P(j)l}$ denotes the position of the l th atom in the group $P(j)$.

* Contribution No. 331 from the Centre of Advanced Study in Physics, University of Madras, Madras, India.

† The importance of studying this problem was brought to the notice of the authors by a referee of *Acta Cryst.* (see footnote on p.164 of Parthasarathy & Ramachandran [*Acta Cryst.* (1966), 21, 163]).

Crystal (1)

$$P(1)[f_{P(1)}, \mathbf{r}_{P(1)i}] + Q[f_{Qk}, \mathbf{r}_{Q(1)k}]$$

Crystal (2)

$$P(2)[f_{P(2)}, \mathbf{r}_{P(2)j}] + Q[f_{Qk}, \mathbf{r}_{Q(2)k}]$$

$$i=1 \text{ to } P(1), \quad j=1 \text{ to } P(2), \quad k=1 \text{ to } Q$$

$$P(1) \neq P(2), \quad f_{P(1)} \neq f_{P(2)}$$

$$\mathbf{r}_{P(1)i} \neq \mathbf{r}_{P(2)j}, \quad \text{for any } i \text{ or } j$$

Here the symbol \neq is used to denote that the quantities concerned may or may not be equal and the symbol \neq is used to denote that the quantities concerned should not be equal.

For an *SI* pair,

$$\mathbf{r}_{Q(1)k} = \mathbf{r}_{Q(2)k} = \mathbf{r}_{Qk}, \quad \text{say,}$$

while for an *NI* pair,

$$\mathbf{r}_{Q(1)k} \neq \mathbf{r}_{Q(2)k} \quad \text{for any } R.$$

The structure factor equation for a reflexion \mathbf{H} ($=hkl$) for the *SI* pair will be

$$F_{N(i)} = F_{P(i)} + F_Q, \quad i=1,2 \quad (1)$$

where* $N(i) = P(i) + Q$, $i=1$ to 2 are the total number of atoms in the unit cells of the two crystals (Fig. 1). Following the notation used in the earlier paper (PR, 1966) we shall write

$$\sigma_{N(i)}^2 = \sigma_{P(i)}^2 + \sigma_Q^2, \quad i=1,2 \quad (2)$$

$$\sigma_{1(i)}^2 = \sigma_{P(i)}^2 / \sigma_{N(i)}^2, \quad \sigma_{2(i)}^2 = \sigma_Q^2 / \sigma_{N(i)}^2, \quad i=1,2 \quad (3)$$

$$\sigma_{1(i)}^2 + \sigma_{2(i)}^2 = 1, \quad i=1,2. \quad (4)$$

* Note that $P(i)$ and Q denote the number of atoms as well as the name of the group.

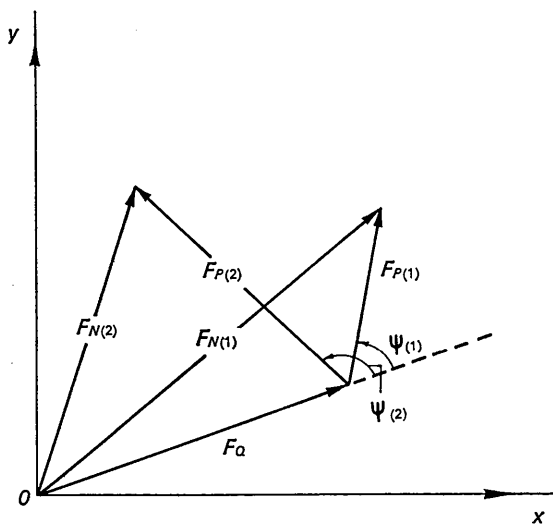


Fig. 1. Argand diagram showing the relation between the structure factors of a semi-isomorphous pair of non-centrosymmetric crystals.

Since both in the *SI* and *NI* types, the two crystals of the pair contains the same number and type of atoms in the Q group, it is clear that $\sigma_{Q(1)}^2 = \sigma_{Q(2)}^2 = \sigma_Q^2$, say, and this property is used in equation (2). The normalized intensity in the two crystals of the *SI* pair will be given by

$$z_{(i)} = |F_{N(i)}|^2 / \sigma_{N(i)}^2$$

$$= [|F_{P(i)}|^2 + |F_Q|^2 + 2|F_{P(i)}||F_Q|\epsilon_i] / \sigma_{N(i)}^2$$

$$= \sigma_{2(i)}^2 y_Q^2 + \sigma_{1(i)}^2 y_{P(i)}^2$$

$$+ 2\sigma_{1(i)}\sigma_{2(i)}y_{P(i)}y_Q\epsilon_i, \quad i=1,2. \quad (5)$$

where y_Q and $y_{P(i)}$ are the normalized structure amplitudes and ϵ_i is the cosine of the angle between $F_{P(i)}$ and F_Q .

Following the earlier paper (PR, 1966) we define the intensity variables z_+ , z_- and z_x as

$$z_{\pm} = z_{(1)} \pm z_{(2)} \quad \text{and} \quad z_x = z_{(1)} \cdot z_{(2)} \quad (6)$$

In the case of an *NI* pair of crystals, the structure factor equation would be given by

$$F_{N(i)} = F_{P(i)} + F_{Q(i)}, \quad i=1,2.$$

so that the normalized intensity would be given by

$$z_{(i)} = \sigma_{2(i)}^2 y_{Q(i)}^2 + \sigma_{1(i)}^2 y_{P(i)}^2$$

$$+ 2\sigma_{1(i)}\sigma_{2(i)}y_{P(i)}y_{Q(i)}\epsilon_i, \quad i=1,2. \quad (7)$$

The quantities y_Q , $y_{P(1)}$, $y_{P(2)}$, ϵ_1 and ϵ_2 in equation (5) are evidently independent random variables and so are $y_{Q(1)}$, $y_{Q(2)}$, $y_{P(1)}$, $y_{P(2)}$, ϵ_1 and ϵ_2 of equation (7). From equations (5) and (7) it is clear that the moments of z_+ or z_x will depend on whether the pair are *SI* or *NI* and hence these higher moments can be used as criteria for testing whether a pair of crystals is of the *SI* or *NI* type. Since the first few moments of z_{\pm} and z_x would suffice in practice (PR, 1966), we shall obtain the expressions of $\langle z_m^n \rangle$ for $n=2, 3$ and 4 when $m=\pm$ and for $n=1, 2$ and 3 when $m=x$ in the next section.

3. Derivations of the higher moments of z_m ($m=\pm, x$) for an *SI* pair

(a) *The sum and difference variables*

From equations (5) and (6) we obtain the expressions for the sum and difference variables as

$$z_{\pm} = a_1 y_Q^2 + a_2 y_Q y_{P(1)} \epsilon_1$$

$$+ a_3 y_Q y_{P(2)} \epsilon_2 + a_4 y_{P(1)}^2 + a_5 y_{P(2)}^2 \quad (8)$$

where we have used the following simplifying notation:

$$a_1 = \sigma_{2(1)}^2 \pm \sigma_{2(2)}^2,$$

$$a_2 = 2\sigma_{2(1)}\sigma_{1(1)},$$

$$a_3 = \pm 2\sigma_{2(2)}\sigma_{1(2)},$$

$$a_4 = \sigma_{1(1)}^2,$$

$$a_5 = \pm 2\sigma_{1(2)}^2. \quad (9)$$

In equations (9) the signs \pm in a_1 , a_3 and a_5 correspond

to z_{\pm} respectively. To obtain the expression of $\langle z_{\pm}^n \rangle$, we have to expand the expression for z_{\pm}^n by using the multinomial theorem and then use the statistical theorems given in equations (17) and (18) of PR (1966). Since the procedure in the present paper is the same as that used in the previous paper (PR, 1966) we shall omit all the intermediate steps and list only the final results. The final expressions* for the second, third and fourth moments of z_{\pm} are:

$$\langle z_{\pm}^2 \rangle = a_4^2 B_{4(1)} + a_5^2 B_{4(2)} + 2a_1^2 + \frac{1}{2}(a_2^2 + a_3^2) + 2(a_1 a_4 + a_1 a_5 + a_4 a_5) \quad (10)$$

$$\langle z_{\pm}^3 \rangle = a_4^3 B_{6(1)} + a_5^3 B_{6(2)} + 3[a_4^2(a_1 + a_5) + \frac{1}{2}a_2^2 a_4] B_{4(1)} + 3[a_5^2(a_1 + a_4) + \frac{1}{2}a_3^2 a_5] B_{4(2)} + 6[a_1^3 + a_1^2(a_4 + a_5)] + a_1 a_4 a_5 + 3[a_2^2(a_1 + \frac{1}{2}a_5) + a_3^2(a_1 + \frac{1}{2}a_4)] \quad (11)$$

$$\langle z_{\pm}^4 \rangle = a_4^4 B_{8(1)} + a_5^4 B_{8(2)} + [4a_4^3(a_1 + a_5) + 3a_2^2 a_4^2] B_{6(1)} + [4a_5^3(a_1 + a_4) + 3a_3^2 a_5^2] B_{6(2)} + 6a_4^2 a_5^2 B_{4(1)} B_{4(2)} + [3a_2^2(\frac{1}{4}a_2^2 + 2a_4 a_5 + 4a_1 a_4) + 3a_4^2(4a_1^2 + a_3^2 + 4a_1 a_5)] B_{4(1)} + [3a_3^2(\frac{1}{4}a_3^2 + 2a_4 a_5 + 4a_1 a_5) + 3a_5^2(4a_1^2 + a_2^2 + 4a_1 a_4)] B_{4(2)} + 6[4a_1^3(a_1 + a_4 + a_5) + 4a_1^2 a_4 a_5 + 3a_1^2(a_2^2 + a_3^2) + \frac{1}{2}a_2^2 a_3^2] + 2a_2^2 a_1 a_5 + 2a_3^2 a_1 a_4. \quad (12)$$

A study of equations (10) to (12) shows that the moments of z_{\pm} depend on the values of $B_{n(i)}$, $i=1,2$ which in turn depend on the number of atoms in the $P(i)$ -groups and the symmetry of the crystal. For the important cases, namely when $P(i)=1,2$ and many (*i.e.* *MA* and *MC* cases of PR, 1966) the values of $B_{n(i)}$ for crystals of the triclinic system are given in Table 1. For other space groups the values of $B_{n(i)}$ could be obtained from Foster & Hargreaves (1963). Substituting the values of $B_{n(i)}$, thus obtained, we can arrive at the final expressions for the higher moments of z_{\pm} in any particular case. In anticipation of the derivation of the higher moments of z_x in the next section it may be noted here that the above arguments for obtaining the final expression of the higher moments of z_{\pm} also holds for the product variable z_x .

Table 1. Values of $B_{n(i)}$ for $n=4,6$ and 8 in a crystal of symmetry $P1$

$P(i)$	n	n	n
	4	6	8
1	1	1	1
2	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{3 \pm}{8}$
<i>MC</i>	3	15	105
<i>MA</i>	2	6	24

* For convenience we shall write $B_{n(i)}$ to stand for $\langle y_{P(i)}^n \rangle$ in these expressions.

(b) The product variable

From equations (5) and (6) we obtain the expression for the product variable:

$$z_x = b_1 y_Q^4 + b_2 y_Q^2 y_{P(1)}^2 + b_3 y_Q^2 y_{P(2)}^2 + b_4 y_{P(1)}^2 y_{P(2)}^2 + b_5 y_Q^3 y_{P(1)} \varepsilon_1 + b_6 y_Q^2 y_{P(2)} \varepsilon_2 + b_7 y_Q y_{P(1)} y_{P(2)} \varepsilon_2 + b_8 y_Q y_{P(1)} y_{P(2)}^2 \varepsilon_1 + b_9 y_Q^2 y_{P(1)} y_{P(2)} \varepsilon_1 \varepsilon_2. \quad (13)$$

The quantities b_i , $i=1$ to 9 in equation (13) are given by

$$b_1 = \sigma_{2(1)}^2 \sigma_{2(2)}^2, \quad b_2 = \sigma_{1(1)}^2 \sigma_{2(2)}^2, \quad b_3 = \sigma_{2(1)}^2 \sigma_{1(2)}^2, \\ b_4 = \sigma_{1(1)}^2 \sigma_{1(2)}^2, \quad b_5 = 2\sigma_{2(1)} \sigma_{1(1)} \sigma_{2(2)}^2, \\ b_6 = 2\sigma_{2(1)}^2 \sigma_{2(2)} \sigma_{1(2)}, \quad b_7 = 2\sigma_{1(1)}^2 \sigma_{2(2)} \sigma_{1(2)}, \\ b_8 = 2\sigma_{2(1)} \sigma_{1(1)} \sigma_{1(2)}^2, \quad b_9 = 4\sigma_{1(1)} \sigma_{1(2)} \sigma_{2(1)} \sigma_{2(2)}. \quad (14)$$

By following the method outlined in a previous paper (Parthasarathy, 1968) we can easily obtain the higher moments of z_x . The final expressions are given below.

$$\langle z_x \rangle = 2b_1 + b_2 + b_3 + b_4 \quad (15)$$

$$\langle z_x^2 \rangle = 2(b_2^2 + \frac{1}{4}b_7^2 + b_2 b_4) B_{4(1)} + 2(b_3^2 + \frac{1}{4}b_8^2 + b_3 b_4) B_{4(2)} + b_4^2 B_{4(1)} B_{4(2)} + 24b_1^2 + 3(b_5^2 + b_6^2) + \frac{1}{2}b_9^2 + 12b_1(b_2 + b_3) + 4b_2 b_3 + 2(b_5 b_8 + b_6 b_7) \quad (16)$$

$$\langle z_x^3 \rangle = 3b_2^2(2b_2^2 + 2b_2 b_4 + b_7^2) B_{6(1)} + 3b_3^2(2b_3^2 + 2b_3 b_4 + b_8^2) B_{6(2)} + 3(b_4^2 b_2 + \frac{1}{2}b_7^2 b_4) B_{6(1)} B_{4(2)} + 3(b_4^2 b_3 + \frac{1}{2}b_8^2 b_4) B_{4(1)} B_{6(2)} + b_4^3 B_{6(1)} B_{6(2)} + 9[2b_2^2(4b_1 + b_3) + b_2^2(4b_2 + b_4) + b_7^2 b_1 + \frac{1}{2}b_7^2 b_2 + 2b_2(2b_1 b_4 + b_5 b_8 + b_6 b_7) + b_5 b_7 b_9] B_{4(1)} + 9[2b_3^2(4b_1 + b_2) + b_3^2(4b_3 + b_4) + b_8^2 b_1 + \frac{1}{2}b_8^2 b_3 + 2b_3(2b_1 b_4 + b_5 b_8 + b_6 b_7) + b_6 b_8 b_9] B_{4(2)} + 3[2b_4^2 b_1 + b_7^2 b_3 + b_8^2 b_2 + \frac{1}{2}b_9^2 b_4 + 4b_2 b_3 b_4 + 2b_4(b_5 b_8 + b_6 b_7) + b_7 b_8 b_9] B_{4(1)} B_{4(2)} + 36[20b_1^3 + 2b_1^2(5b_2 + 5b_3 + b_4) + b_2^2(5b_1 + b_3) + b_6^2(5b_1 + b_2) + \frac{1}{2}b_9^2 b_1 + 2b_1(b_2 b_3 + b_5 b_8 + b_6 b_7) + b_5 b_6 b_9]. \quad (17)$$

4. Discussion of the theoretical results

The higher moments of z_{\pm} and z_x for an *SI* pair are functions of the quantities $\sigma_{1(1)}^2$ and $\sigma_{1(2)}^2$ which can be evaluated by knowing the number and types of atoms in the two crystals. Since the higher moments of the normalized intensities $z_{(i)}$ are also functions of $\sigma_{1(i)}^2$ (Parthasarathy, 1966) the values of the higher moments of z_{\pm} and z_x of the *NI* pair of non-centrosymmetric crystals can be calculated from equation (37) of PR (1966), equation (25) of Parthasarathy (1968), and the equations (36) to (39) of Parthasarathy (1966). A comparison of the experimental values of the higher moments calculated using the observed intensities from

Table 2. Verification of the theoretical results

 E =Experimental value, SI =theoretical value for the SI pair, NI =theoretical value for the NI pair

Sample number	Number of carbon atoms in Q group	$\sigma_{I(1)}^2$	$\sigma_{I(2)}^2$									
				E	$\langle z_{\pm}^2 \rangle$	$\langle z_{\pm}^3 \rangle$	$\langle z_{\pm}^4 \rangle$	$\langle z_{\pm}^2 \rangle$	$\langle z_{\pm}^3 \rangle$	$\langle z_{\pm}^4 \rangle$	$\langle z_{\times} \rangle$	$\langle z_{\times}^2 \rangle$
1	42	0.1	0.1	E	6.87	40.10	312.4	0.310	0.410	1.642	18.87	424.2
				SI	7.61	43.31	328.6	0.369	0.700	1.811	19.53	528.3
				NI	5.99	23.90	118.9	1.990	23.55	1.000	3.98	35.5
2	42	0.1	0.2	E	7.41	40.98	301.7	0.534	1.400	1.721	17.50	446.4
				SI	6.84	40.47	322.7	0.455	0.899	1.600	19.22	455.8
				NI	5.98	23.71	117.2	1.975	22.91	1.000	3.95	34.7
3	42	0.2	0.2	E	6.76	39.76	316.8	0.574	1.384	1.545	18.61	441.7
				SI	7.24	38.89	278.0	0.679	2.300	1.639	15.69	379.3
				NI	5.96	23.54	115.5	1.960	22.2	1.000	3.92	34.0
4	24	0.1	0.1	E	7.37	46.85	420.7	0.371	0.469	1.749	25.40	953.6
				SI	7.61	43.30	328.6	0.369	0.700	1.810	19.53	528.3
				NI	5.99	23.90	118.9	1.990	23.55	1.000	3.98	35.5
5	24	0.1	0.2	E	6.97	40.04	315.4	0.574	1.329	1.600	17.95	440.4
				SI	6.84	40.47	322.7	0.455	0.899	1.600	19.22	455.8
				NI	5.98	23.71	177.2	1.975	22.91	1.000	3.95	34.7
6	24	0.2	0.2	E	6.70	35.84	258.5	0.691	1.561	1.502	14.78	314.4
				SI	7.24	38.89	278.0	0.679	2.300	1.639	15.69	379.3
				NI	5.69	23.54	115.5	1.960	22.28	1.000	3.92	34.0

the two crystals, with their theoretical values, thus obtained, for the SI and NI cases would enable one to decide whether the given pair of crystals is of the SI or NI type.

It may be noted here that, from one of the crystals of the SI pair, if the P group of atoms are removed (this can be done mathematically by making either $\sigma_{I(1)}^2 \rightarrow 0$ or $\sigma_{I(2)}^2 \rightarrow 0$) the isomorphism between the two crystals then turns out to be that of the addition type (PR, 1966). Thus the addition type of isomorphism can also be tested by using the expressions in the present paper by putting $\sigma_{I(1)}^2 = 0$ or $\sigma_{I(2)}^2 = 0$.

It may be worth while to note here that the theory in this and earlier papers enable us to test whether a given pair of crystals is of the I (or SI) type or NI type. These results cannot however be used to obtain the degree of isomorphism between the given pair of crystals. For example, the present theory does not consider the effects of a systematic movement of the Q group on the higher moment test and this is particularly important from a practical point of view. The theory for this problem is under investigation.

5. Verification of the theoretical results

The theoretical results were tested in a few two-dimensional models of the SI type. Since in protein crystals the value of $\sigma_{I(i)}^2$ is small (say $\sigma_{I(i)}^2 < 0.2$), because of the very large number of light atoms (Q group) in the unit cell, the theoretical results were tested on hypothetical models in which $\sigma_{I(i)}^2$ is in this range (Table 2). All the models were based on the plane group pg with two

atoms in the $P(i)$ group (*i.e.* $P(1) = P(2) = 2$). The number of atoms in the Q group (*i.e.* the common part) in the two crystals are also given in Table 2. The scattering powers of the atoms in the $P(i)$ group ($i=1,2$) were so chosen as to give the required values of $\sigma_{I(i)}^2$ ($i=1,2$). The values of the higher moments of z_{\pm} and z_{\times} thus calculated are shown in Table 2, and these correspond to the 'experimental values' (denoted by E in this Table). The theoretical values of the higher moments of z_{\pm} and z_{\times} for the SI and NI cases were calculated as stated in § 4. A comparison of the experimental values of these higher moments with the corresponding theoretical values for the SI and NI pairs confirm that all the structures chosen are of the SI type. A study of this Table also shows that though all the tests prove that the crystals taken are of the SI type, the second and fourth moments of z_{\pm} are the best parameters for testing for the isomorphism of the common group of atoms in the SI pair of crystals.

The authors are grateful to the referees for their useful comments and for having suggested the problem of calculating the higher moments when there are finite errors in the coordinates of the Q atoms.

References

- FOSTER, F. & HARGREAVES, A. (1963). *Acta Cryst.* **16**, 1133.
 HARKER, D. (1956). *Acta Cryst.* **9**, 1.
 PARTHASARATHY, S. (1966). *Z. Kristallogr.* **123**, 27.
 PARTHASARATHY, S. (1968). *Z. Kristallogr.* **126**, 379.
 PARTHASARATHY, S. & RAMACHANDRAN, G. N. (1966). *Acta Cryst.* **21**, 163.