Acta Cryst. (1961). **14**, 641

The Refinement of Structures Partially Determined by the Isomorphous Replacement Method

By Michael G. Rossmann and D. M. Blow

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge, England

(*Received* 8 *August* 1960)

When independent results about the phase of a reflexion are obtained, what is the best way of combining them? This problem arises in the multiple isomorphous replacement method, and in a more general form where, in addition, part of the structure is known. The method proposed is to use each result to form a probability function for the phase of the reflexion; the combination of these results is achieved by multiplying the probability functions together. This joint probability function can be fully represented by the magnitude and phase of two vectors **C** and $\mathbf{D}_{is}$ formed by *addition* of components from each result. From these, structure factors can be calculated which provide the 'best' combination of all the data. A discussion of the significance of the vectors **C**, $\mathbf{D}_{is}$ indicates how estimates of the relative importance of structural information and of the isomorphous replacement method may be made.

Introduction

The isomorphous-replacement method has been used by Kendrew *et al.* (1960) to determine the phases of 9,600 reflexions in a study of the protein myoglobin at 2 Å resolution. In the resulting Fourier synthesis, most of the atoms in the helical polypeptide chain can be placed with precision, and a number of the amino acid side chains can be identified. The question now arises, how may an extension or refinement of the structure best be made? The two available techniques—calculation of phases either on the basis of the known part of the structure or by the isomorphous-replacement method—will lead to different sets of phase angles. Is there a satisfactory way of combining these? The large number of reflexions imposes a requirement for a completely automatic method of combination. At the same time, the labour of model building is such that it is worth doing a great deal of preliminary computing in order to reduce the number of cycles of refinement to a minimum.

The problem is approached from the point of view of an earlier paper on the treatment of errors in the isomorphous-replacement method (Blow & Crick, 1959). In a projection subject to error, one may define as 'best Fourier' the Fourier transform which has least mean-square difference from the 'true' Fourier transform when averaged over the whole unit cell. In the usual case, one knows the magnitude F_o of the structure amplitude with accuracy, but its phase is uncertain. If the relative probability of each phase is plotted round a circle of radius F_o on an Argand diagram, the centre of gravity of the probability distribution gives the structure factor, ξ, to be used in calculating the 'best Fourier'. This may be expressed by

$$\xi = F_o \frac{\int_0^{2\pi} \exp(i\alpha) P(\alpha) d\alpha}{\int_0^{2\pi} P(\alpha) d\alpha} . \qquad (1)$$

Here $P(\alpha)d\alpha$ is the relative probability that the phase angle α lies between α and $\alpha+d\alpha$.

In the following paper we present (*a*) a convenient analytical expression, governed by two vectors, that represents the phase probability distribution indicated by a set of isomorphous-replacement data; (*b*) a similar expression, governed by one vector, that represents the corresponding distribution for the 'heavy-atom' or partially known structure method; (*c*) a method for the combination of these to obtain the 'best Fourier' when the two methods are combined simultaneously. This leads to some general strategic considerations in the solution of large structures.

The isomorphous replacement method

In the isomorphous-replacement method Blow & Crick (1959) show that a particular phase angle, α, is related to a particular error $\varepsilon(\alpha)$ in the experimental data of one isomorphous compound. This error can be allotted a probability $\exp\{-\varepsilon^2(\alpha)/2E^2\}$ on the basis of a Gaussian distribution of error with a standard deviation E, which can be estimated, for example, from data for a centrosymmetric zone. In general, J isomorphous derivatives may be used simultaneously. In this case

$$P(\alpha)d\alpha = \prod_{j=1}^{J} \exp\{-\varepsilon_j^2(\alpha)/2E_j^2\} d\alpha$$
$$= \exp\{-\sum_{j=1}^{J} \varepsilon_j^2(\alpha)/2E_j^2\} d\alpha . \qquad (2)$$

ε_j is given by

$$(F_{Hj}+\varepsilon_j)^2=F_o^2+f_j^2+2F_of_j\cos(\alpha-\varphi_j)\,, \qquad (3)$$

where F_o is the structure amplitude of the unsubstituted derivative, F_{Hj} that of the jth heavy atom derivative, and $f_j \exp(i\varphi_j)$ is the calculated structure factor of the substituent atoms in the jth compound (see Fig. 1).

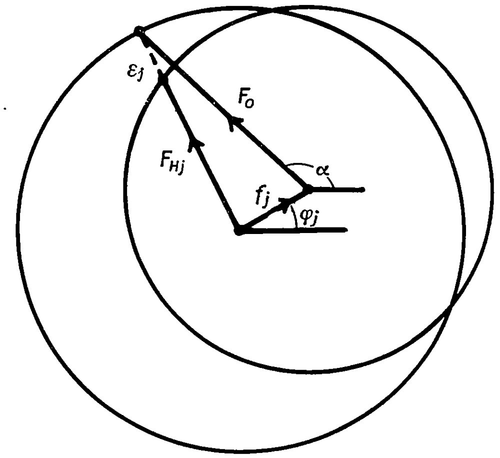

Fig. 1. Vector diagram showing error in the isomorphous replacement technique.

Workers who have used this method (Blow, 1958; Kendrew *et al.*, 1960; Perutz *et al.*, 1960) have so far used numerical methods for the evaluation of (1), by calculating $P(\alpha)$ from (2) and (3) at intervals of 5° or 10°. An analytical expression for the integrals of (1) will now be developed.

Since in all important cases F_o, $F_{Hj} \gg \varepsilon_j$ except where the probabilities are negligible, ε_j^2 can safely be neglected in (3). This leads at once to

$$\varepsilon_j=(F_o^2+f_j^2-F_{Hj}^2)/2F_{Hj}+(F_of_j/F_{Hj})\cos(\alpha-\varphi_j)$$

and

$$\begin{aligned}\varepsilon_j^2/2E_j^2&=e_j-c_j\cos(\alpha-\varphi_j)+2d_j\cos^2(\alpha-\varphi_j)\\&=(e_j+d_j)-c_j\cos(\alpha-\varphi_j)+d_j\cos 2(\alpha-\varphi_j)\,, \qquad (4)\end{aligned}$$

where

$$\begin{aligned}c_j&=(F_{Hj}^2-f_j^2-F_o^2)F_of_j/(2F_{Hj}^2E_j^2)\,,\\d_j&=F_o^2f_j^2/(4F_{Hj}^2E_j^2)\,,\\e_j&=(F_{Hj}^2-f_j^2-F_o^2)^2/(8F_{Hj}^2E_j^2)\,.\end{aligned}$$

This expression shows that $\varepsilon_j^2(\alpha)$ can be expressed with a good accuracy by a Fourier expansion up to the second-order term, and c_j and d_j may be thought of as Fourier coefficients. As will be shown presently, they are, however, like vectorial coefficients with phases φ_j and $2\varphi_j$.

If follows from (4) that

$$\begin{aligned}\sum_{j=1}^{J}\varepsilon_j^2/2E_j^2&=\sum_{j=1}^{J}(e_j+d_j)\\&\quad-\sum_{j=1}^{J}c_j\cos(\alpha-\varphi_j)+\sum_{j=1}^{J}d_j\cos 2(\alpha-\varphi_j)\,. \qquad (5)\end{aligned}$$

The first term of (5) results in a constant multiplier, which appears at top and bottom of (1). The second and third terms provide the required information for calculation of the structure factors, $\boldsymbol{\xi}$, which give rise to the 'best Fourier'.

Let us now write

$$\begin{aligned}C_{is}\cos\Phi_1&=\sum_{j=1}^{J}c_j\cos\varphi_j\,; & C_{is}\sin\Phi_1&=\sum_{j=1}^{J}c_j\sin\varphi_j\,;\\D_{is}\cos 2\Phi_2&=\sum_{j=1}^{J}d_j\cos 2\varphi_j\,; & D_{is}\sin 2\Phi_2&=\sum_{j=1}^{J}d_j\sin 2\varphi_j\,. \qquad (6)\end{aligned}$$

This corresponds to forming vectors $\mathbf{C}_{is}$ and $\mathbf{D}_{is}$ by addition of the c_j and d_j vectors with their corresponding phases, φ_j and $2\varphi_j$. (5) now reduces to

$$\sum_{j=1}^{J}\varepsilon_j^2/2E_j^2=K-C_{is}\cos(\alpha-\Phi_1)+D_{is}\cos 2(\alpha-\Phi_2). \qquad (7)$$

Finally, by combining (1), (2) and (7) we have

$$\boldsymbol{\xi}=F_o\times\frac{\int_0^{2\pi}\exp\{i\alpha\}\exp\{C_{is}\cos(\alpha-\Phi_1)-D_{is}\cos 2(\alpha-\Phi_2)\}d\alpha}{\int_0^{2\pi}\exp\{C_{is}\cos(\alpha-\Phi_1)-D_{is}\cos 2(\alpha-\Phi_2)\}d\alpha}\,.$$

Putting $\psi=\alpha-\Phi_1$,

$$\boldsymbol{\xi}=F_o\exp(i\Phi_1)\times\frac{\int_0^{2\pi}\exp\{i\psi+C_{is}\cos\psi-D_{is}\cos 2(\psi+\Phi_1-\Phi_2)\}d\psi}{\int_0^{2\pi}\exp\{C_{is}\cos\psi-D_{is}\cos 2(\psi+\Phi_1-\Phi_2)\}d\psi}$$

or most briefly

$$\boldsymbol{\xi}=F_o\exp(i\Phi_1)\mathfrak{g}(C_{is},D_{is},\Phi_1-\Phi_2)\,. \qquad (8)$$

This last form shows that the integrals are functions of the three variables C_{is}, D_{is}, $(\Phi_1-\Phi_2)$. Their ratio is a complex number $\mathfrak{g}(C_{is},D_{is},\Phi_1-\Phi_2)$ which can readily be calculated on a computer. Some values are given in Table 1.

The validity of the approximation introduced by neglecting ε^2 was tested by computation of some phase

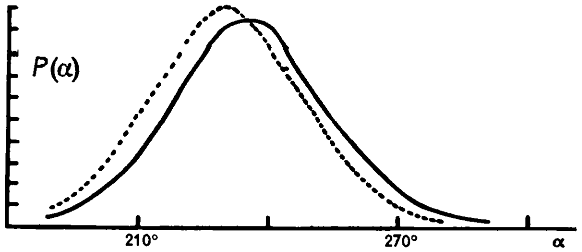

Fig. 2. Dashed line shows probability distribution for reflexion (260) of haemoglobin with approximations suggested in this paper; continuous line shows the same distribution calculated without error.

probability curves from data for five isomorphous compounds of horse haemoglobin. These were calculated both by using equation (3) (accurate) and equation (4) (approximate). The two curves for the reflexion showing the poorest agreement are shown in Fig. 2. The error, about 5°, is quite insignificant.

All the information available about a structure factor from isomorphous replacements can thus be expressed in terms of five parameters: F_o and the real and imaginary parts of $\mathbf{C}_{is}$ and $\mathbf{D}_{is}$, given by

$$\mathbf{C}_{is}=\sum_{j=1}^{J}\frac{(F_{Hj}^2-f_j^2-F_o^2)F_of_j}{2F_{Hj}^2E_j^2}\exp(i\varphi_j),$$

$$\mathbf{D}_{is}=\sum_{j=1}^{J}\frac{F_o^2f_j^2}{4F_{Hj}^2E_j^2}\exp(i2\varphi_j)\ .$$

For practical application in a computer this has several advantages:

(i) It is necessary to store only five parameters per reflexion as a complete record of the results, rather than the $(3J+1)$ parameters required for observed structure amplitudes and calculated heavy-atom contributions.

(ii) If further information becomes available about a reflexion at a later stage, it can be readily added into the $\mathbf{C}_{is}$ and $\mathbf{D}_{is}$ vectors.

(iii) Only one integration is required (or a single reference to a three-dimensional table of $\boldsymbol{\xi}$), rather than J such operations which would be required if the isomorphous replacements were treated separately.

The heavy atom method

A probability distribution $P(\alpha)$ can also be derived in the case where partial knowledge of the structure is available, either from the positions of a small number of heavy atoms or from a large number of light atoms. The two cases are formally identical, and are here both referred to as the 'heavy-atom method'.

Let $\mathbf{f}_H=f_H\exp(i\varphi_H)$ be a structure factor calculated from that part of structure which is known. Sim (1959) has derived the probability function $P(\alpha)$, by making the assumption that the contribution of the unknown atoms conforms to Wilson statistics (1949), and has a mean square value Σ. Following Sim, the probability may be written

$$P(\alpha)d\alpha=K\exp\{c_H\cos(\alpha-\varphi_H)\}d\alpha\ , \qquad (9)$$

where

$$c_H=2F_of_H/\Sigma\ ,$$

and

$$K=2\pi I_0(c_H)\ .$$

I_0 is the modified zero-order Bessel function of imaginary argument ic_H.

Substituting (9) into (1), and changing the variable to $\psi=\alpha-\varphi_H$, gives

$$\boldsymbol{\xi}=F_o\exp(i\varphi_H)\frac{\int_0^{2\pi}\cos\psi\exp\{c_H\cos\psi\}d\psi}{\int_0^{2\pi}\exp\{c_H\cos\psi\}d\psi}\ , \qquad (10)$$

which may be shown to reduce to

$$\boldsymbol{\xi}=F_o\exp(i\varphi_H)I_1(c_H)/I_0(c_H)\ . \qquad (11)$$

(Sim, 1960; Watson, 1922).

One feature of Sim's treatment which may readily be improved is the assumption that the 'known' part of the structure is perfectly accurate. In addition to the contribution to Σ from the L atoms whose position is unknown, there will be a further contribution due to error in the parameters of the H 'known' atoms. If the hth known atom has been assigned a position $\mathbf{r}_h$, while its true position is $\mathbf{r}_h+\boldsymbol{\delta}_h$, and its scattering factor is f_h, then

$$(\mathbf{f}_H)_{\text{true}}=\sum_{h=1}^{H}f_h\exp\{2\pi i(\mathbf{r}_h+\boldsymbol{\delta}_h).\mathbf{s}\}\ .$$

Here $\mathbf{s}$ is the reciprocal-lattice vector, and assuming the angle $2\pi\boldsymbol{\delta}_h.\mathbf{s}$ is small,

$$(\mathbf{f}_H)_{\text{true}}\approx(\mathbf{f}_H)_{\text{calc.}}+\sum_{h=1}^{H}2\pi i(\boldsymbol{\delta}_h.\mathbf{s})f_h\exp\{2\pi i\mathbf{r}_h.\mathbf{s}\}\ .$$

The last term is a vector sum made up of the contributions of each 'known' atom to the total structure factor. Its mean square value is

$$\tfrac{4}{3}\pi\mathbf{s}^2\sum_{h=1}^{H}\sigma_h^2f_h^2\ ,$$

where σ_h is the r.m.s. value of $\boldsymbol{\delta}_h$, or the standard error in position of the hth atom. This quantity needs to be added to the mean square contribution of the 'unknown' atoms, giving

$$\Sigma=\sum_{l=1}^{L}f_l^2+\tfrac{4}{3}\pi\mathbf{s}^2\sum_{h=1}^{H}\sigma_h^2f_h^2\ . \qquad (12)$$

This latter term, although probably negligible in the early stages, becomes increasingly important in later cycles when more of the structure is known. This is particularly true of the terms with large $\mathbf{s}$, or if atomic positions determined at a low resolution (small $\mathbf{s}$) are used for phase determination at a higher resolution.

Alternatively, Σ could be estimated for a range of $\mathbf{s}$ by the study of a centric zone, in the same way as E_j has been determined for the isomorphous-replacement technique.

It has been assumed that each atom of the structure can be assigned into one of the two groups 'known' and 'unknown'. In the case of an uncertain atom a possible procedure would be to put part of its weight into the 'known' set and part into the 'unknown'. However, while this is the best procedure for determining the rest of the structure, it will not decide whether this atom really exists.

Table 1. *The function* $\mathfrak{g}\,(x, y, \omega)$

The function is set out in blocks of constant ω.
The left hand blocks give $100\times$ the magnitude of $\mathfrak{g}$ and the right hand blocks its phase in degrees

	$\lvert g \rvert$ x 100									*arg g* degrees								
X	Y= 0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	Y= 0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
w = 0°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	19	14	10	8	6	5	4	3	360	360	360	360	360	360	360	360	360
1.0	45	35	27	20	15	12	10	8	7	360	360	360	360	360	360	360	360	360
1.5	60	49	38	29	22	18	14	12	10	360	360	360	360	360	360	360	360	360
2.0	70	60	48	38	30	23	19	16	14	360	360	360	350	360	360	360	360	360
2.5	76	67	57	46	36	29	24	20	17	360	360	360	360	360	360	360	360	360
3.0	81	73	64	53	43	35	29	24	21	360	360	360	360	360	360	350	360	360
3.5	84	78	69	59	49	40	33	28	24	360	360	360	360	360	360	360	360	360
4.0	86	81	74	64	55	46	38	32	28	360	360	360	360	360	360	360	360	360
4.5	88	84	77	69	60	51	43	36	31	360	360	360	360	360	360	360	360	360
5.0	89	86	80	73	64	55	47	40	35	360	360	360	360	360	360	360	360	360
w = 15°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	20	16	14	13	13	12	12	12	360	9	20	31	41	48	53	57	60
1.0	45	37	31	27	26	25	25	24	24	360	8	19	31	41	48	53	57	59
1.5	60	51	44	39	37	36	36	36	36	360	8	19	30	40	47	53	56	59
2.0	70	62	54	49	47	46	46	46	46	360	8	18	29	39	46	52	56	58
2.5	76	69	62	58	55	54	54	54	55	360	7	17	27	37	45	51	55	58
3.0	81	75	69	64	62	62	62	62	62	360	7	16	26	36	44	50	54	57
3.5	84	79	74	70	68	67	68	68	68	360	6	15	24	34	42	48	53	56
4.0	86	82	78	74	72	72	73	73	74	360	6	14	23	32	41	47	52	55
4.5	88	85	81	78	76	76	76	77	78	360	5	13	21	31	39	46	50	54
5.0	89	86	83	80	79	79	80	81	81	360	5	12	20	29	37	44	49	53
w = 30°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	22	21	21	22	22	23	23	23	360	13	26	36	43	47	50	51	53
1.0	45	41	40	40	41	42	43	43	44	360	13	26	35	41	46	49	51	52
1.5	60	56	55	55	57	58	59	59	60	360	12	25	34	41	46	48	50	52
2.0	70	66	65	66	68	70	71	72	72	360	12	24	33	40	45	48	50	51
2.5	76	73	73	74	76	78	79	80	81	360	11	22	32	39	43	46	49	50
3.0	81	79	78	79	81	83	85	86	86	360	10	21	30	37	41	45	48	49
3.5	84	82	82	83	85	87	88	89	90	360	10	20	29	36	40	44	46	48
4.0	86	85	84	86	88	89	91	92	92	360	9	18	27	34	39	43	45	47
4.5	88	87	86	88	89	91	92	93	94	360	8	17	26	32	37	41	44	46
5.0	89	88	88	89	90	92	93	94	95	360	8	16	24	31	36	40	43	45
w = 45°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	25	26	28	29	30	31	31	32	360	14	24	31	35	37	39	40	41
1.0	45	46	48	51	53	55	56	57	57	360	13	23	30	34	37	38	40	40
1.5	60	61	64	67	70	72	73	74	75	360	13	22	29	33	36	38	39	40
2.0	70	71	74	77	80	82	83	84	85	360	12	21	28	32	35	37	38	39
2.5	76	78	81	84	86	88	89	90	91	360	11	20	27	31	34	36	37	38
3.0	81	82	85	87	89	91	92	93	94	360	10	19	25	30	33	35	36	37
3.5	84	85	87	89	91	93	94	95	95	360	10	18	24	29	32	34	35	36
4.0	86	87	89	91	93	94	95	96	96	360	9	17	23	27	30	33	34	36
4.5	88	89	90	92	93	95	95	96	97	360	9	16	22	26	29	32	33	35
5.0	89	90	91	93	94	95	96	96	97	360	8	15	21	25	28	31	32	34
w = 60°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	27	31	33	35	36	37	38	38	360	11	17	22	24	25	26	27	27
1.0	45	50	55	59	61	63	65	65	66	360	10	17	21	24	25	26	27	27
1.5	60	65	70	75	77	79	81	81	82	360	10	16	20	23	24	25	26	27
2.0	70	75	80	83	86	88	89	90	90	360	9	16	20	22	24	25	26	26
2.5	76	81	85	88	90	92	93	93	94	360	9	15	19	21	23	24	25	26
3.0	81	85	88	91	92	94	95	95	96	360	8	14	18	20	22	23	24	25
3.5	84	87	90	92	94	95	95	96	96	360	8	13	17	19	21	23	24	24
4.0	86	89	91	93	94	95	96	96	97	360	7	12	16	19	20	22	23	24
4.5	88	90	92	94	95	96	96	97	97	360	7	12	15	18	20	21	22	23
5.0	89	91	93	94	95	96	96	97	97	360	6	11	15	17	19	20	22	22
w = 75°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	29	33	36	39	40	41	42	42	360	6	9	11	12	13	13	14	14
1.0	45	52	59	63	66	68	69	70	71	360	6	9	11	12	13	13	13	14
1.5	60	68	74	78	81	83	84	85	86	360	5	8	10	12	12	13	13	13
2.0	70	77	82	86	88	90	91	92	92	360	5	8	10	11	12	12	13	13
2.5	76	83	87	90	92	93	94	95	95	360	5	8	10	11	11	12	12	13
3.0	81	86	90	92	94	95	95	96	96	360	4	7	9	10	11	12	12	12
3.5	84	88	91	93	94	95	96	96	97	360	4	7	9	10	11	11	12	12
4.0	86	90	92	94	95	96	96	97	97	360	4	6	8	9	10	11	11	12
4.5	88	91	93	94	95	96	97	97	97	360	4	6	8	9	10	11	11	11
5.0	89	92	93	95	96	96	97	97	97	360	3	6	7	9	10	10	11	11
w = 90°																		
0	0	0	0	0	0	0	0	0	0	–	–	–	–	–	–	–	–	–
0.5	24	30	34	38	40	41	42	43	43	360	0	0	0	0	0	0	0	0
1.0	45	53	60	64	67	69	71	72	72	360	0	0	0	0	0	0	0	0
1.5	60	69	75	79	82	84	85	86	87	360	0	0	0	0	0	0	0	0
2.0	70	78	83	87	89	91	92	92	93	360	0	0	0	0	0	0	0	0
2.5	76	83	88	90	92	94	94	95	95	360	0	0	0	0	0	0	0	0
3.0	81	86	90	92	94	95	96	96	97	360	0	0	0	0	0	0	0	0
3.5	84	89	91	93	95	95	96	97	97	360	0	0	0	0	0	0	0	0
4.0	86	90	92	94	95	96	96	97	97	360	0	0	0	0	0	0	0	0
4.5	88	91	93	94	95	96	97	97	97	360	0	0	0	0	0	0	0	0
5.0	89	92	94	95	96	96	97	97	97	360	0	0	0	0	0	0	0	0

Referring again to equation (10), and comparing it with the function $\mathfrak{g}$ defined by equation (8), it is evident that for the heavy atom method

$$\xi = F_o \exp(i\varphi_H)\,\mathfrak{g}(c_H, 0, 0)\,.$$

The results of the heavy-atom method have thus been reduced to the same mathematical form as those of the isomorphous-replacement method, and it will next be shown that they may readily be combined.

The combination of the isomorphous-replacement and the heavy-atom method

The case may now be considered, which has arisen in the myoglobin work of Kendrew, where isomorphous-replacement data are available, and also the structure is partially known. In this case the two contributions to $P(\alpha)$ must be multiplied together, with the result

$$P(\alpha)d\alpha = \exp\{C_{is}\cos(\alpha-\Phi_1) + c_H\cos(\alpha-\varphi_H) - D_{is}\cos 2(\alpha-\Phi_2)\}d\alpha\,.$$

C_{is} and c_H may be combined in just the same way as the c_j's were combined in equation (6):

$$C\cos\Phi = \sum_{j=1}^{J} c_j\cos(\alpha-\varphi_j) + c_H\cos(\alpha-\varphi_H)$$

$$C\sin\Phi = \sum_{j=1}^{J} c_j\sin(\alpha-\varphi_j) + c_H\sin(\alpha-\varphi_H)\,. \quad (13)$$

By exactly the same steps as lead to equation (8) we reach the result

$$\xi = F_o\exp(i\Phi)\,\mathfrak{g}(C, D_{is}, \Phi-\Phi_2)\,. \quad (14)$$

Thus all available information can be summarized in terms of three variables which determine the definite integrals of $\mathfrak{g}$.

The same function $\mathfrak{g}(x, y, \omega)$ gives analytical expression to the weighting scheme of Blow & Crick (1959) when applied to the results from any number of isomorphous replacements and the heavy atom method. The various cases are set out in Table 2.

E_j and Σ may be determined with moderate accuracy by analysis of the results from a centrosymmetric zone, in the way demonstrated by Blow & Crick (1959). In addition, a theoretical estimate of Σ may be made from (12). Although errors in E_j and Σ will result in inaccurate relative weighting of the different sets of data, it is unlikely that these will have an important effect on the phase angles.

Table 2. *The form of the function* $\mathfrak{g}$ *in various methods*

'Heavy atom' method	$\mathfrak{g}(c_H, 0, 0)$	(Sim, 1960)
Single isomorphous pair	$\mathfrak{g}(c_1, d_1, 0)$	(Blow & Rossmann, 1961)
Single isomorphous pair combined with 'heavy atom' method	$\mathfrak{g}(C_1, d_1, \Phi-\varphi_1)$	
Multiple isomorphous replacement method	$\mathfrak{g}(C_{is}, D_{is}, \Phi_1-\Phi_2)$	
Multiple isomorphous replacement combined with 'heavy atom' method	$\mathfrak{g}(C, D_{is}, \Phi-\Phi_2)$	

The function $\mathfrak{g}$ (x, y, ω)

The nature of the function

$$\mathfrak{g}(x, y, \omega) = \frac{\int_0^{2\pi}\exp\{i\psi + x\cos\psi - y\cos 2(\psi+\omega)\}d\psi}{\int_0^{2\pi}\exp\{x\cos\psi - y\cos 2(\psi+\omega)\}d\psi} \quad (15)$$

may now be considered. It can be thought of as representing the centre of gravity of a circular wire of radius unity with a density

$$\exp\{x\cos\psi - y\cos 2(\psi+\omega)\}\,.$$

Since the centre of gravity of such a wire can never be outside the wire, it follows that the magnitude of $\mathfrak{g}$ always lies between 0 and 1, approaching unity as we approach certainty about the true phase of a reflexion.

$\mathfrak{g}$ must always be real if the weight distribution is symmetrical about $\psi=0$. This occurs not only when $y=0$ (as in the case of heavy atom technique) but also when $|\omega| = 0, \pi/2, \pi, \ldots.$ In the case of a single isomorphous pair ($J=1$) (8) reduces to $\xi = F_o\exp(i\varphi_1)\mathfrak{g}(c_1, d_1, 0)$. We know that this case results in a probability distribution for the phase angle α which is symmetrical about φ_1. Thus ω may be regarded as expressing the asymmetry of the phase probability distribution.

When $x \gg y$ then $x\cos\psi$ is the governing factor of the probability distribution. Hence we obtain a unimodal distribution with a peak whose sharpness increases as x becomes large. Conversely when y becomes $\gg x$ the $y\cos 2(\psi+\omega)$ term governs the probability distribution. This results in a bimodal distribution with peaks tending towards $\pi/2-\omega$, $3\pi/2-\omega$. If $\omega=0$, $\mathfrak{g}\rightarrow 0$ as y increases; but for other values of ω (implying an asymmetric distribution), as y increases and sharpens the distribution $|\mathfrak{g}|$ increases, and its phase swivels round towards $(\pi/2-\omega)$. For smaller, but still considerable values of y, the probability distribution has two peaks, which coalesce into one as y becomes small compared to x. It can easily be shown that the distribution must be unimodal when $x \geq 4y$ and $\omega=0$.

The strategy of a structure determination

Apart from its direct application to the combination of structural and isomorphous-replacement data, the analysis given above is useful in evaluating the relative merits of different strategies in a large structure determination.

First consider the effect of increasing the number of isomorphous compounds. This may be done by considering the effect of adding one member to the vector summation for $\mathbf{C}_{is}$ and $\mathbf{D}_{is}$, defined by (6). It may be seen from (3) and (4) that if α_0 is a phase angle for which $\varepsilon_j=0$, so that

$$2F_o f_j \cos(\alpha_0-\varphi_j)=F_{Hj}^2-F_o^2-f_j^2\,,$$

then

$$c_j=4d_j\cos(\alpha_0-\varphi_j)\,,$$

while $d_j\approx f_j^2/4E_j^2$ is a positive number. If the phases are assumed to be random, the problem corresponds to a random-walk problem, giving a mean square value for $\mathbf{D}_{is}$

$$\langle D_{is}^2\rangle=\sum_{j=1}^{J} d_j^2\approx\sum_{j=1}^{J} f_j^4/(16E_j^4)\,. \qquad (16)$$

Thus in the case where the f_j/E_j's are of the same order for the different isomorphous derivatives, the r.m.s. value of $\mathbf{D}_{is}$ increases as $\sqrt{J}$. Note that two isomorphous substituents j, k at closely adjacent sites will lead to similar values of φ_j, φ_k, making $\mathbf{D}_{is}$ larger than it would be if the phases were random.

For a rough approximation it can be assumed that α_0 is the 'true' phase, and has the same value for all j. In the vector summation

$$\mathbf{C}_{is}=\sum_{j=1}^{J}4d_j\cos(\alpha_0-\varphi_j)\exp(i\varphi_j)$$

it is seen that terms with phase φ_j far from α_0 are scaled down by the factor $\cos(\alpha_0-\varphi_j)$, and thus $\mathbf{C}_{is}$ tends to have the phase of α_0. Expanding

$$\mathbf{C}_{is}=\sum_{j=1}^{J}4d_j\,[\cos\alpha_0\cos^2\varphi_j+\sin\alpha_0\sin\varphi_j\cos\varphi_j + i\cos\alpha_0\sin\varphi_j\cos\varphi_j+i\sin\alpha_0\sin^2\varphi_j]\,.$$

Assuming, as before, a random distribution of φ_j, we see that the mean value of $\mathbf{C}_{is}$ is

$$\langle\mathbf{C}_{is}\rangle=\sum_{j=1}^{J}2d_j\exp(i\alpha_0)\approx\sum_{j=1}^{J}[f_j^2/2E_j^2]\exp(i\alpha_0)\,. \qquad (17)$$

If all $\langle f_j^2/E_j^2\rangle$ are of the same order of magnitude, $\langle|\mathbf{C}_{is}|\rangle$ increases as J and $\langle|\mathbf{C}_{is}|\rangle/\langle D_{is}^2\rangle^{\frac{1}{2}}$ as $\sqrt{J}$. This expresses the fact that, as the number of isomorphous compounds is increased, the distribution is sharpened and there is a decrease of their tendency towards being bimodal.

Consider now the heavy-atom method. If a large number of atomic positions have been found in a partially known structure, we may assume that Wilson statistics apply, and the mean square value of f_H is

$$\langle f_H^2\rangle=\sum_{h=1}^{H}f_h^2\,,$$

while $\langle F_o^2\rangle=\langle f_H^2\rangle+\Sigma$. Hence the mean square value of c_H is

$$\langle c_H^2\rangle=4\langle F_o^2\rangle\langle f_H^2\rangle/(\langle F_o^2\rangle-\langle f_H^2\rangle)^2.$$

Let $\langle f_H^2\rangle=r\langle F_o^2\rangle$ so that we may say briefly 'a fraction r of the structure is known'. Then the r.m.s. value of c_H is

$$\langle c_H^2\rangle^{\frac{1}{2}}=2r^{\frac{1}{2}}/(1-r)\,. \qquad (18)$$

Remembering that C characterizes the sharpness of the phase probability distribution and thus the accuracy of the phase determination, we are now in a position to answer questions about the relative power of the isomorphous-replacement method and the heavy atom method under specific conditions. Suppose, for example, a fraction r of the structure has been revealed by using J isomorphous replacements, is it more advantageous to use one more isomorphous replacement, or to use the heavy-atom method? We can calculate the probable value of $|\mathbf{C}|$ in each case. Let $\langle\mathbf{C}_J\rangle$ be the mean value of $\mathbf{C}$ when J isomorphous replacements are used. Then from (17)

$$\langle\mathbf{C}_{J+1}\rangle=\mathbf{C}_J+\frac{f_{J+1}^2}{E_{J+1}^2}\exp(i\alpha_0)\approx\frac{J+1}{J}\langle\mathbf{C}_J\rangle\,;$$

while using (18), adding heavy-atom data gives

$$\langle C_{J+H}^2\rangle^{\frac{1}{2}}=\left[C_J^2+\frac{2r}{(1-r)^2}\right]^{\frac{1}{2}}.$$

It may be mentioned that the isomorphous-replacement method tends to increase $\mathbf{D}_{is}$ as well as $\mathbf{C}$, so that a comparison based only on the value of $|\mathbf{C}|$ tends to over-estimate the value of the isomorphous-replacement method. Nevertheless, to make the example more definite, let us consider typical data for the protein haemoglobin. These would be

$$\langle f_j^2\rangle=100^2,\quad E_j=50\quad\text{so that}\quad\langle C_J\rangle=2J.$$

Thus

$$\langle C_{J+1}\rangle^2\approx 4(J+1)^2$$

and

$$\langle C_{J+H}^2\rangle\approx 4J^2+\frac{2r}{(1-r)^2}\,.$$

The heavy-atom method will therefore be more powerful if the latter quantity is greater, namely if $r>0{\cdot}73$ $(J=2)$, $0{\cdot}77$ $(J=3)$, $0{\cdot}79$ $(J=4)$. (In the case $J=1$ it would certainly be necessary to take account of the effect of $\mathbf{D}_{is}$).

These results are sensitive to the relative magnitudes of the inaccurately known E_j and Σ values. This type of calculation can therefore only give an order of magnitude for r.

More detailed comparisons could readily be made. From the above it can be seen how quantitative expression can be given to the convergence of sets of isomorphous-replacement data or heavy-atom refinements, towards an accurate set of phases. In principle it would be possible to go further and use the actual values of $|\mathbf{g}|$ which are the 'figures of merit' (Dickerson *et al.*, 1961) to give the standard error of electron density (Blow & Crick, 1959).

This study has been stimulated by a fruitful interchange of ideas with other members of this Unit. In particular we are indebted to Dr F. H. C. Crick, who we hope will recognize his influence throughout the work.

The tabulation of the function $g(x, y, \omega)$ was made possible by the use of the University of Cambridge Mathematical Laboratory's electronic computer EDSAC 2.

References

Blow, D. M. (1958). *Proc. Roy. Soc.* A, **247**, 302.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794.

Blow, D. M. & Rossmann, M. G. (1960). *Acta Cryst.* **14** (In press).

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *Conference on Computing Methods and the Phase problem.* (Glasgow, 1960, London: Pergamon Press.)

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature, Lond.* **185**, 422.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). *Nature, Lond.* **185**, 416.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813.

Sim, G. A. (1960). *Acta Cryst.* **13**, 511.

Watson, G. N. (1922). *The theory of the Bessel Functions*, p. 77. Cambridge: University Press.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Acta Cryst. (1961). **14**, 647

Preparation and Structure of $Ba_5Ta_4O_{15}$ and Related Compounds*

By Francis Galasso and Lewis Katz

Department of Chemistry, University of Connecticut, Storrs, Connecticut, U.S.A.

(*Received* 3 *August* 1960)

The preparation and characterization of a new ternary oxide of tantalum, $Ba_5Ta_4O_{15}$, and isomorphous compounds, $Sr_5Ta_4O_{15}$ and $Ba_5Nb_4O_{15}$, are described. The barium tantalum compound was found to belong to the trigonal system; the axes of the primitive hexagonal cell are $a = 5·79$, $c = 11·75$ Å. Single crystals of $Ba_5Ta_4O_{15}$ grown in a lead II oxide flux were used in determining the structure. Anion deficiencies have been produced in these compounds by preparing them with tetravalent niobium and tantalum.

Introduction

In a general survey of the barium–tantalum–oxygen system, the existence of several phases have been noted. At a ratio of barium to tantalum of one half, two different phases have been reported, one a hexagonal phase, $Ba_{0.44}TaO_{2.92}$ (Galasso, Katz & Ward, 1958), and the other a tetragonal phase, $Ba_{0.5}TaO_3$ (Galasso, Katz & Ward, 1959*a*) with the tetragonal bronze structure (Magnéli, 1949). When the Ba/Ta ratio was increased to three, the phase $Ba(Ba_{0·5}Ta_{0·5})O_{2.75}$ was obtained with an ordered cubic perovskite structure (Brixner, 1958; Galasso, Katz & Ward, 1959*b*). At an intermediate ratio, the compound $Ba_5Ta_4O_{15}$ has been identified. The preparation and structure of this compound and isomorphous phases is the subject of this paper.

Experimental

Preparation of $Ba_5Ta_4O_{15}$

The best preparation of $Ba_5Ta_4O_{15}$ resulted from mixing barium carbonate in 2% excess of the amount indicated in the reaction

$$5\,BaCO_3 + 2\,Ta_2O_5 \rightarrow Ba_5Ta_4O_{15} + 5\,CO_2$$

and heating in air at 1150 °C. for 24 hr. The white product obtained gave an X-ray powder pattern which was indexed with the aid of single crystal data on the basis of a hexagonal cell with $a = 5·79$ Å and $c = 11·75$ Å.

The density was found pycnometrically to be 7·9 g.cm.$^{-3}$, and using the above parameters the unit cell content weight was calculated to be 1622, as compared to 1650 for the formula $Ba_5Ta_4O_{15}$. Analysis gave 44·04% Ta, 42·70% Ba, as compared to the theoretical 43·85% Ta, 41·64% Ba for $Ba_5Ta_4O_{15}$.

Preparation of $Ba_5Ta_4^{IV}O_{13}$

The compound $Ba_5Ta_4^{IV}O_{13}$ was prepared in an evacuated sealed silica capsule using tantalum pentoxide and tantalum metal as a source of tantalum IV, and barium oxide as the source of barium. As is frequently the case with oxygen deficient compounds, the powder pattern of the reduced (blue) phase could be indexed using the same parameters as for the oxidized phase.

---

* Part of this research was sponsored by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.