

Theoretical evaluation of the probability of success of the quasi-anomalous method[†]

K CHINNAKALI, S PARTHASARATHY and N ELANGO

Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600025, India

MS received 27 January 1986; revised 21 January 1987

Abstract. Theoretical expressions for the probability of success of the quasi-anomalous method have been derived for triclinic, monoclinic and orthorhombic crystals containing a few (1, 2 or 3) heavy atoms per asymmetric unit besides a large number of light atoms. The results derived take into account data-truncation due to unobserved reflections. Using the theoretical expressions, tables of probability values for the success of the quasi-anomalous method are obtained as a function of the relevant parameters k and σ_1^2 . Corresponding results for triclinic crystals containing many heavy atoms (i.e. $P = MN$ and MC cases) have also been obtained. It is seen that, using suitable heavy atoms to prepare the heavy-atom derivative, probability of success as high as 0.7 could be obtained in the case of proteins containing 1000 to 1500 atoms.

Keywords. Probability; quasi-anomalous method; crystals.

PACS No. 61.10

1. Introduction

It is well known that the phase of a reflection in a non-centrosymmetric crystal containing heavy atoms can be determined from the intensities of the inverse reflections but for a two-fold ambiguity (Ramachandran and Raman 1956, Peerdemann and Bijvoet 1956). A simple method of resolving the two-fold ambiguity consists in choosing the phase closer to that of the heavy atoms and this procedure of phase determination is called the quasi-anomalous method (QAM*). The power of this method is obvious from the few structures that were solved by the QAM immediately after it was proposed (Raman 1959, Dale *et al* 1963, Geurtz 1963, Hall and Maslen 1965, Yow-lam oh and Maslen 1966). Since the QAM method requires twice the amount of data to be collected, the interest in this method declined subsequently. However during recent years interest in the anomalous scattering method has increased particularly in the field of protein crystallography due to the advent of synchrotron radiation as a source for diffraction work as well as the development of faster methods of data collection. It is hence useful to derive theoretical expression from which the probability of success of the QAM (denoted by P_{QAM}) for a given crystal can be computed.

Parthasarathy (1965) obtained the values of P_{QAM} as a function of the parameter σ_1^2 (see (8) for a definition of this quantity) under the following assumptions: (i) The crystal

[†]Contribution No. 702.

*For the principle and other details on QAM, one may refer to the excellent monograph of Ramachandran and Srinivasan (1970).

belongs to the space group $P1$. (ii) All the theoretically possible reflections within a given region of $\sin\theta/\lambda$ could be measured*. These assumptions are not strictly valid in actual crystals. For example, most organic crystals and proteins crystallize in space groups of higher symmetry. Further, the intensities of a certain percentage of reflections measured in a diffractometer may be negative and many crystallographers truncate the data set by omitting such reflections as well as those whose net counts are less than one to two times the standard deviation of this value (Stout and Jenson 1968). In photographic data also a certain percentage of reflections are too weak to be above the background and such 'unobserved' reflections are generally omitted in crystal structure analysis. We shall therefore derive the theoretical expression for P_{QAM} by taking into account the influence of space group symmetry as well as the effect of data truncation due to unobserved reflections. In this paper we shall consider the theory for crystals belonging to triclinic, monoclinic and orthorhombic systems.

2. Notation and preliminary results

Consider a non-centrosymmetric crystal containing N atoms in the unit cell. Let P of these be anomalous scatterers of the same type and $Q (= N - P)$ be normal scatterers of similar scattering power. Let s be the number of equivalent general positions in the unit cell. Let $n = (N/s)$, $p = (P/s)$ and $q = (Q/s)$ be the number of N -, P - and Q -atoms per asymmetric unit. The structure factor equation for reflection $H (= hkl)$ can be written as (see figure 1)

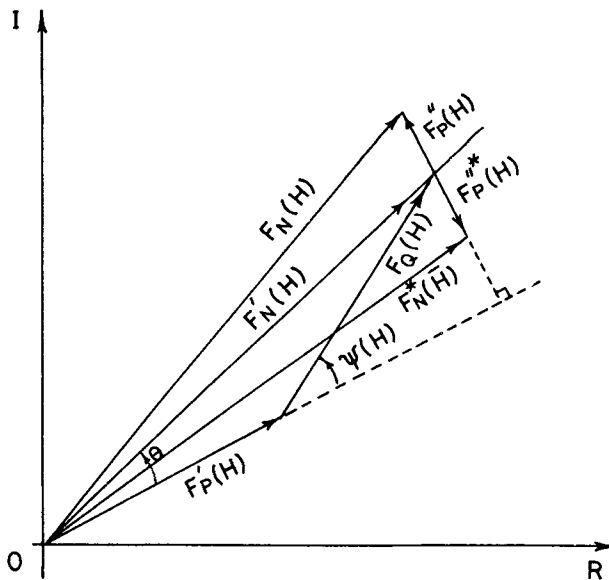


Figure 1. Argand diagram showing the relationship between the various components of the structure factors of the inverse reflections H and \bar{H} in a non-centrosymmetric crystal containing P -anomalous and Q -normal scatterers in the unit cell.

* Owing to this assumption P_{QAM} was found to be a function of σ_1^2 only. See Section 5 for the secondary influence of k on P_{QAM} .

$$F_N(H) = F'_N(H) + F''_P(H), \quad (1)$$

where

$$F'_N(H) = F'_P(H) + F_Q(H).$$

Let θ be defined by

$$\theta = \alpha'_N - \alpha'_P. \quad (2)$$

The mean intensity of the Bijvoet pair is given by (Ramachandran and Raman 1956)

$$\bar{I} = |F'_N|^2 + |F''_P|^2. \quad (3)$$

The normalized intensities and normalized structure factor magnitudes of the inverse reflections H and \bar{H} are defined to be

$$\begin{aligned} z(H) &= [y(H)]^2 = |F_N(H)|^2 / \langle \bar{I} \rangle, \\ z(\bar{H}) &= [y(\bar{H})]^2 = |F_N(\bar{H})|^2 / \langle \bar{I} \rangle, \end{aligned} \quad (4)$$

where $\langle \bar{I} \rangle$ is the local average value of \bar{I} , namely, (see (3))

$$\langle \bar{I} \rangle = \langle |F'_N|^2 \rangle + \langle |F''_P|^2 \rangle. \quad (5)$$

Let z_{\min} be the minimum of the values $z(H)$ and $z(\bar{H})$ and let y_{\min} be that of $y(H)$ and $y(\bar{H})$. Let y_t and z_t denote the threshold values of the normalized structure factor magnitude and the normalized intensity respectively. It is evident that

$$z_{\min} = y_{\min}^2, \quad z_t = y_t^2. \quad (6)$$

The probability of success of QAM is expected to depend on the parameters k and σ_1^2 which are defined by

$$k = f''_P / (f''_P + f'_P), \quad (7)$$

and

$$\sigma_1^2 = \langle |F'_P|^2 \rangle / \langle |F'_N|^2 \rangle. \quad (8)$$

Let us define σ_2^2 by

$$\sigma_2^2 = \langle |F_Q|^2 \rangle / \langle |F'_N|^2 \rangle. \quad (9)$$

It is known that

$$\sigma_1^2 + \sigma_2^2 = 1. \quad (10)$$

The normalized structure factor magnitudes y_Q , y'_P and y'_N are defined by

$$\begin{aligned} y_Q &= [|F_Q|^2 / \langle |F_Q|^2 \rangle]^{1/2}, \quad y'_P = [|F'_P|^2 / \langle |F'_P|^2 \rangle]^{1/2}, \\ y'_N &= [|F'_N|^2 / \langle |F'_N|^2 \rangle]^{1/2}. \end{aligned} \quad (11)$$

Let $\psi(H)$ be the angle defined by

$$\psi(H) = \alpha_Q(H) - \alpha'_P(H). \quad (12)$$

Since $\alpha'_p(\bar{H}) = -\alpha'_p(H)$ and $\alpha_Q(\bar{H}) = -\alpha_Q(H)$, it follows that

$$\psi(\bar{H}) = -\psi(H). \tag{13}$$

In crystals with a single species of anomalous scatterer we have (Srinivasan and Parthasarathy 1976)

$$|F'_p| = k|F'_p|. \tag{14}$$

Using (8) and (14) in (5), the local average value of \bar{I} can be rewritten in terms of k and σ_1^2 as

$$\langle \bar{I} \rangle = (1 + k^2 \sigma_1^2) \langle |F'_N|^2 \rangle. \tag{15}$$

3. Basic equation for probability of success of the QAM

The Bijvoet ratio of that Bijvoet pair for which the intensities $I(H)$ and $I(\bar{H})$ are both greater than the background level can alone be measured. Further, only the reflections for which $|\theta| \leq \pi/2$ have their phases correctly determined by the QAM. Hence the probability of success of QAM can be quantitatively defined as the fractional number of reflections which satisfy the following two conditions simultaneously: (i) $|\theta| \leq \pi/2$ and (ii) $z_{\min} > z_t$ (or equivalently $y_{\min} > y_t$). Using the set theoretic notation of intersection (denoted by the symbol \cap) the simultaneous satisfaction of these two conditions can be written as $(|\theta| \leq \pi/2) \cap (y_{\min} > y_t)$. Using the probability concept we can therefore define P_{QAM} as the probability of occurrence of the event D , where D is defined by

$$D = [(|\theta| \leq \pi/2) \cap (y_{\min} > y_t)]. \tag{16}$$

That is,

$$P_{\text{QAM}} = \text{Pr}(D). \tag{17}$$

It is thus seen that P_{QAM} , for a given crystallographic situation, can be obtained by evaluating the probability function in the right hand side of (17). We shall presently consider this aspect.

4. Derivation of the theoretical expression for P_{QAM}

4.1. Theoretical expression for θ

From figure 1 we can obtain the following results:

$$|F'_N| \sin \theta = |F_Q| \sin \psi \tag{18}$$

$$|F_Q|^2 = |F'_N|^2 + |F'_p|^2 - 2|F'_N| |F'_p| \cos \theta \tag{19}$$

$$|F'_N|^2 = |F'_p|^2 + |F_Q|^2 + 2|F'_p| |F_Q| \cos \psi. \tag{20}$$

It is convenient to rewrite (18) and (19) as

$$\sin \theta = (|F_Q| \sin \psi) / |F'_N| \tag{21}$$

and

$$\cos\theta = (|F'_N|^2 + |F'_P|^2 - |F'_Q|^2) / (2|F'_N||F'_P|) \quad (22)$$

respectively. In terms of the normalized variables y_Q , y'_N and y'_P , (21) and (22) can be respectively written as

$$\sin\theta = (\sigma_2 y_Q \sin(2\pi\psi'_0)) / y'_N \quad (23)$$

and

$$\cos\theta = (y'^2_N + \sigma_1^2 y'^2_P - \sigma_2^2 y'^2_Q) / (2\sigma_1 y'_N y'_P) \quad (24)$$

where we have used the substitution $\psi'_0 = \psi / (2\pi)$. From (20) it is seen that

$$y'^2_N = \sigma_1^2 y'^2_P + \sigma_2^2 y'^2_Q + 2\sigma_1 \sigma_2 y'_P y_Q \cos(2\pi\psi'_0). \quad (25)$$

For further derivation we need a convenient expression for y'_P and this can be obtained by making use of the results of Foster and Hargreaves (1963).

Foster and Hargreaves have shown that the triclinic, monoclinic and orthorhombic space groups can be conveniently classified into 7 categories based on the trigonometric factors of the geometrical structure factor. Among these, the categories 1, 3, 5 and 6 correspond to the non-centrosymmetric case and hence these alone are of interest in our study. y'_P for these four space group categories can be written as (Parthasarathy and Ponnuswamy 1976)

$$y'_P = E_P / \sqrt{\varepsilon p}, \quad (26)$$

where ε is 1, $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{4}$ for the categories 1, 3, 5 and 6 respectively and E_P is given by

$$E_P = \left[\left(\sum_{j=1}^p \xi_{pj} \right)^2 + \left(\sum_{j=1}^p \eta_{pj} \right)^2 \right]^{1/2} \quad (27)$$

See table 1 for the expressions for E_P . Here ξ_{pj} and η_{pj} are the trigonometric parts of the structure factor listed in volume I of the International Tables for X-ray Crystallography (1969). Substituting (26) and (27) in (24) we obtain

$$\cos\theta = (y'^2_N + c_1 E_P^2 + c_2 y'^2_Q) / (c_3 y'_N E_P) \quad (28)$$

where c_1 , c_2 and c_3 are defined to be

$$c_1 = \sigma_1^2 / (\varepsilon p), \quad c_2 = -\sigma_2^2, \quad c_3 = 2\sigma_1 / \sqrt{\varepsilon p}. \quad (29)$$

The theoretical expressions for the normalized intensities $z(H)$ and $z(\bar{H})$ are known to be (Velmurugan and Parthasarathy 1984)

$$z(H) = c_4 E_P^2 + c_5 y'^2_Q + c_6 E_P y_Q [\cos(2\pi\psi'_0) + k \sin(2\pi\psi'_0)], \quad (30)$$

and

$$z(\bar{H}) = c_4 E_P^2 + c_5 y'^2_Q + c_6 E_P y_Q [\cos(2\pi\psi'_0) - k \sin(2\pi\psi'_0)] \quad (31)$$

where

$$c_4 = \frac{(1+k^2)\sigma_1^2}{\varepsilon p(1+k^2\sigma_1^2)}, \quad c_5 = \frac{\sigma_2^2}{(1+k^2\sigma_1^2)},$$

and

$$c_6 = \frac{2\sigma_1\sigma_2}{\sqrt{\epsilon p(1+k^2\sigma_1^2)}} \quad (32)$$

4.2 Expression of P_{QAM} for space group category 5

The joint probability density function of the random variables $y_Q, \psi'_0, \theta'_i, \phi'_i, \psi'_i$ ($i=1$ to p) can be written as (Velmurugan and Parthasarathy 1984)

$$\begin{aligned} P(y_Q, \dots, \psi'_p) &= P(y_Q)P(\psi'_0) \prod_{i=1}^p P(\theta'_i)P(\phi'_i)P(\psi'_i) \\ &= 2y_Q \exp(-y_Q^2), \quad 0 \leq y_Q < \infty, \quad 0 < \psi'_0, \dots, \psi'_p < 1, \end{aligned} \quad (33)$$

where $\theta'_i = \theta_i/2\pi, \phi'_i = \phi_i/2\pi, \psi'_i = \psi_i/2\pi$.

From (30), (31) and (32) it is clear that

$$\theta = \theta(y_Q, \dots, \psi'_p) \quad (34)$$

and

$$y_{\min} = y_{\min}(y_Q, \dots, \psi'_p). \quad (35)$$

In the $(3p+2)$ -dimensional (y_Q, \dots, ψ'_p) -space we have denoted the domain satisfying the conditions $|\theta| \leq \pi/2$ and $y_{\min} > y_i$ simultaneously by D . In view of (34) and (35) we obtain from (33)

$$P_{QAM} = \int \dots \int_D 2y_Q \exp(-y_Q^2) dy_Q \dots d\psi'_p \quad (36)$$

where the $(3p+2)$ -fold integration is over the domain D . The multiple integral on the right hand side of (36) is too difficult to evaluate in the closed form. It may however be evaluated by the Monte Carlo method of integration in a digital computer. In order to facilitate this we make use of the transformation

$$y_Q = u/(1-u), \quad (37)$$

and rewrite (36) as

$$P_{QAM} = 2 \int \dots \int_{D'} [y_Q \exp(-y_Q^2)]_u \frac{du d\psi'_0 \dots d\psi'_p}{(1-u)^2} \quad (38)$$

where D' is the image of D under the mapping (37). The symbolic notation $[\dots]_u$ implies that, during numerical computation of the quantity in square parentheses, y_Q is to be replaced by $u/(1-u)$.

It may be noted that E_p for category 6 is also a function of the $3p$ variables $(\theta'_1, \phi'_1, \psi'_1, \dots, \psi'_p)$. For category 3, E_p is a function of the $2p$ variables $(\theta'_1, \phi'_1, \dots, \phi'_p)$. For category 1, E_p is a function of the p variables $(\theta'_1, \theta'_2, \dots, \theta'_p)$. The explicit expression for E_p for the categories 1, 3, 5 and 6 are given in table 1 (Parthasarathy and Ponnuswamy 1976). Expression (38) is valid for all categories. It is however necessary to note that the multiple integrals involved in these equations are over $(p+2)$ -, $(2p+2)$ -, $(3p+2)$ - and $(3p+2)$ -dimensional spaces for the space group categories 1, 3, 5 and 6 respectively.

4.3 Theoretical expression for P_{QAM} for the many-atom cases*

In this connection we shall take the space group of the crystal to be $P1$. Here it is convenient to make use of the independent variables y_Q, y'_P and ψ (Parthasarathy and Srinivasan 1964). The joint probability density function of y_Q, y'_P and ψ are given by (Velmurugan and Parthasarathy 1984)

$$P(y_Q, y'_P, \psi) = (2/\pi)y'_P y_Q \exp(-y_Q^2 - y'_P^2) \quad \text{for } P = MN \quad (39)$$

$$= \frac{\sqrt{2}}{\pi^{3/2}} y_Q \exp(-y_Q^2 - y'_P^2/2) \quad \text{for } P = MC \quad (40)$$

$$0 \leq y_Q, y'_P < \infty; 0 < \psi < 2\pi.$$

Equation (24) is valid for the present cases. The expressions for $z(H)$ and $z(\bar{H})$ are now given by (Velmurugan and Parthasarathy 1984)

$$z(H) = c_7 y'_P{}^2 + c_5 y_Q^2 + c_8 y'_P y_Q (\cos \psi + k \sin \psi), \quad (41)$$

and

$$z(\bar{H}) = c_7 y'_P{}^2 + c_5 y_Q^2 + c_8 y'_P y_Q (\cos \psi - k \sin \psi), \quad (42)$$

where

$$c_7 = \frac{(1+k^2)\sigma_1^2}{(1+k^2\sigma_1^2)}, \quad c_8 = \frac{2\sigma_1\sigma_2}{(1+k^2\sigma_1^2)}.$$

By following the arguments used for deriving (38) we can show that for the present cases

$$P_{QAM} = \iiint_D P(y_Q, y'_P, \psi) dy_Q dy'_P d\psi$$

where D is the domain in the three dimensional (y_Q, y'_P, ψ) -space satisfying the inequalities $|\theta| \leq \pi/2$ and $y_{\min} > y_i$ simultaneously. The theoretical expressions for the cases $P = MN$ and $P = MC$ are to be obtained by substituting the appropriate expressions for $P(y_Q, y'_P, \psi)$ in (38). These integrals can be evaluated by the Monte Carlo method and for this it is convenient to make use of the transformation

$$y_Q = u/(1-u), \quad y'_P = v/(1-v), \quad (44)$$

and rewrite (43) as

$$P_{QAM} = \iiint_{D'} [P(y_Q, y'_P, \psi)]_{uv} \frac{du dv d\psi}{(1-u)^2 (1-v)^2} \quad (45)$$

where D' is the image of D under the transformation (44). The symbolic notation $[. . .]_{uv}$ implies that, during numerical computation of the quantity in square parentheses, y_Q and y'_P are to be replaced by $u/(1-u)$ and $v/(1-v)$ respectively.

* In a non-centrosymmetric crystal of space group $P1$, when there are many atoms in the P -group, the P atoms can take up either a centrosymmetric configuration or a non-centrosymmetric configuration. These two situations are symbolically denoted by MC and MN respectively.

5. Discussion of the theoretical results

From the theoretical expressions for P_{QAM} derived in § 4 it is seen that P_{QAM} , for given y , p and space-group category, is a function of k and σ_1^2 . In actual crystals the truncation limit y , is in the neighbourhood of 0.2 (Ponnuswamy and Parthasarathy 1977). The Bijvoet difference data of reflections for which $y(H)$ and $y(\bar{H})$ are close to this limit may

Table 1. Expressions for E_p for the space group categories 1, 3, 5 and 6.

Space group category	Expression for E_p
1	$[(\Sigma \cos \theta_i)^2 + (\Sigma \sin \theta_i)^2]^{1/2}$
3	$[(\Sigma \cos \theta_i \cos \phi_i)^2 + (\Sigma \cos \theta_i \sin \phi_i)^2]^{1/2}$
5	$[(\Sigma \cos \theta_i \cos \phi_i \cos \psi_i)^2 + (\Sigma \sin \theta_i \sin \phi_i \sin \psi_i)^2]^{1/2}$
6	$[(\Sigma \cos \theta_i \cos \phi_i \cos \psi_i)^2 + (\Sigma \cos \theta_i \cos \phi_i \sin \psi_i)^2]^{1/2}$

Note: Σ denotes the summation over the p atoms in the asymmetric unit.

Table 2. Values of P_{QAM} (in %) as a function of k , σ_1^2 and p for space group categories 1, 3, 5 and 6.

σ_1^2	SGC	k	1			3			5			6		
			p	1	2	3	1	2	3	1	2	3	1	2
0.05	0.1	58.0	57.4	56.4	57.5	56.3	56.5	56.2	57.6	58.5	53.1	57.7	58.0	
		0.8	55.2	54.8	54.0	55.1	53.9	54.2	53.9	54.6	56.2	53.3	54.9	56.0
0.10	0.1	63.2	61.7	61.2	62.1	60.7	60.7	60.5	62.4	63.0	59.2	62.2	62.5	
		0.8	59.4	59.0	58.3	59.0	57.6	57.1	57.8	59.2	60.4	57.1	59.0	60.0
0.15	0.1	67.1	64.7	64.6	65.8	63.8	63.9	63.9	66.2	66.2	61.9	66.2	65.6	
		0.8	63.3	62.3	61.4	62.5	60.6	59.7	61.1	62.5	63.4	59.8	62.3	62.9
0.20	0.1	70.5	68.2	67.5	68.2	66.6	66.4	66.5	68.8	69.1	64.5	68.4	68.5	
		0.8	67.0	65.6	64.6	65.4	63.2	63.1	63.9	65.5	66.7	62.1	65.2	65.9
0.30	0.1	76.4	73.5	72.5	73.0	71.0	71.5	72.1	73.6	73.9	68.6	72.8	73.5	
		0.8	73.5	70.3	69.7	70.6	67.5	67.9	69.0	70.2	71.6	65.8	70.0	70.5
0.40	0.1	81.7	77.7	77.1	77.4	75.0	75.5	76.4	77.8	78.0	71.7	76.8	77.8	
		0.8	79.3	75.0	74.0	74.4	72.1	71.3	73.0	74.5	75.6	68.5	74.1	74.4
0.50	0.1	86.3	81.1	80.4	80.6	78.6	79.7	79.7	81.2	82.4	74.2	80.0	81.0	
		0.8	85.0	78.8	77.4	78.6	76.0	75.1	76.8	77.7	78.9	71.1	76.8	78.1
0.60	0.1	90.8	83.8	83.4	83.5	81.8	83.4	83.1	84.6	85.5	76.4	82.9	84.5	
		0.8	90.8	81.8	80.4	81.6	79.3	78.5	80.1	80.6	82.0	72.6	79.5	81.3
0.80	0.1	98.0	88.1	88.8	88.1	87.6	88.2	88.5	89.5	90.3	79.4	87.3	90.0	
		0.8	99.1	84.1	86.6	85.2	84.8	84.4	85.0	85.5	86.9	73.9	83.6	85.8

Note: SGC denotes space group category.

Table 3. Values of the P_{QAM} (in %) as a function of k and σ_1^2 and P for the many-atom cases $P = MN$ and MC .

P	σ_1^2	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.8
	k									
MN	0.1	56.2	60.3	63.7	66.8	70.7	74.5	78.1	80.7	86.4
	0.8	53.7	57.8	60.4	63.1	68.0	71.2	75.1	78.0	83.6
MC	0.1	54.8	58.5	61.4	63.9	68.2	70.8	73.6	75.8	79.8
	0.8	53.2	56.3	59.1	61.3	65.4	68.2	69.9	71.7	73.8

Table 4. Comparison of the values of P_{QAM} obtained for truncated and untruncated data.

σ_1^2		P			
		1	2	MN	MC
0.05	UD	62	62	61	60
	TD	58	57	56	55
0.20	UD	76	72	72	70
	TD	71	68	67	64
0.40	UD	88	82	82	78
	TD	82	78	74	71
0.60	UD	96	89	88	83
	TD	91	84	81	76
0.80	UD	100	94	94	89
	TD	98	88	86	80

Note: The values of P_{QAM} given in the rows corresponding to TD (truncated data) are obtained from the present theory taking k to be 0.1 while those in the rows corresponding to UD (untruncated data) are the ones calculated from the theoretical results in Parthasarathy (1965).

not be very accurate. We shall therefore assume that reflections for which $y(H)$ and $y(\bar{H})$ are greater than 0.3 would be sufficiently accurate to yield useful Bijvoet difference data. Hence we shall obtain the values of P_{QAM} by taking y_i to be 0.3.

Values of P_{QAM} have been calculated for triclinic, monoclinic and orthorhombic crystals containing p ($= 1, 2$ and 3) anomalous scatterers per asymmetric unit for different values of k and σ_1^2 by evaluating the appropriate integrals (see (38) and (45)) by the Monte Carlo method (Demidovich and Maron 1975). The results obtained are given in table 2. The corresponding results obtained for the many-atom cases (i.e. $P = MN$ and $P = MC$) are given in table 3.

It would be appropriate to make a comparative study of the values of P_{QAM} obtained for truncated data (i.e. the present results) with the corresponding results obtained for

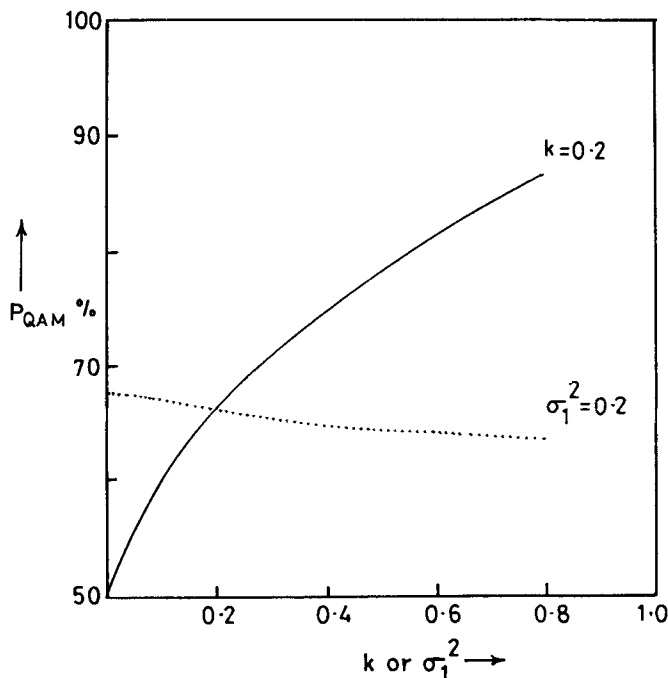


Figure 2. Variation of P_{QAM} (in %) as functions of the parameters k and σ_1^2 for the case $p=1$ in space group category 5. The solid curve denotes the variation of P_{QAM} with respect to σ_1^2 at a constant value of $k=0.2$ while the dotted curve denotes that with respect to k at a constant value of $\sigma_1^2=0.2$.

untruncated data (i.e. the results obtained for triclinic crystals by Parthasarathy 1965). The relevant values of P_{QAM} obtained for triclinic crystals are given in table 4. It is seen that, for given values of k and σ_1^2 , the value of P_{QAM} for the truncated data is always less than that for untruncated data and this is to be physically expected.

The variation of P_{QAM} as a function of σ_1^2 is shown in figure 2 for the case $p=1$ corresponding to the space group category 5 by taking the value of k to be 0.2. Similar trend is observed in the other space group categories for other values of p . It is seen from figure 2 that, for given values of p and k and given space group category, as the value of σ_1^2 increases the value of P_{QAM} systematically increases. It therefore follows that the largest percentage of reflections would have their phases correctly determined by the QAM if we choose, out of a series of isomorphous heavy atom derivatives of a given protein, the one for which the value of σ_1^2 is the largest. Such a choice of the heavy atoms for the derivative has another advantage, namely, that the positions of the anomalous scatterers (which are required for phase determination by the QAM) can be determined more easily if the value of σ_1^2 is larger.

The variation of P_{QAM} as a function of k is shown in figure 2 for the case $p=1$ corresponding to the space group category 5 by taking the value of σ_1^2 to be 0.2. The almost flat nature of the curve (see figure 2) shows that, for given values of p and σ_1^2 and for given space group category, as k increases the value of P_{QAM} decreases at an extremely slow rate. Thus it is seen that P_{QAM} is more sensitive to change in σ_1^2 than in k . This is understandable from the following consideration: It is obvious that P_{QAM} would

depend primarily on the probability distribution of $|\theta|$. Since the probability distribution of $|\theta|$ depends only on σ_1^2 but not on k (Parthasarathy, 1965), we expect P_{QAM} to be predominantly influenced by σ_1^2 . The influence of k on P_{QAM} arises from the measurability criterion $z_{min} > z_t$ which the inverse reflections have to satisfy. It is known that data truncation due to unobserved reflections has only a small effect on the measurability (Parthasarathy 1982). Hence we expect k to influence P_{QAM} to a much lesser extent only. We have therefore given the values of P_{QAM} only for $k=0.1$ and 0.8 in tables 2 and 3.

It would be interesting to obtain the values of P_{QAM} to be expected in crystals containing a few typical heavy atoms besides a few hundred light atoms and in the case of heavy-atom derivatives of small proteins.

Co, Fe and I are good anomalous scatterers for $CuK\alpha$ and Br is a good anomalous scatterer for $MoK\alpha$. The values of P_{QAM} expected for crystals of space group categories 1, 3, 5 or 6 containing p (i.e. $p=1, 2$) heavy atoms besides a few hundred carbon atoms per asymmetric unit are given in table 5. These values were obtained by bilinear interpolation from the results in table 2 by making use of the average values of k and σ_1^2 corresponding to $CuK\alpha$ data for which $(\sin\theta/\lambda) < 0.4 \text{ \AA}^{-1}$. It is seen from table 5 that, even in crystals containing about 300 atoms per asymmetric unit, P_{QAM} value could be as high as 70% if a proper heavy atom species is chosen to prepare the derivative.

Table 5. Values of P_{QAM} (in %) expected in crystals belonging to the space group categories 1, 3, 5 and 6 and containing p ($=1$ or 2) anomalous scatterers and q ($=100, 300$ or 500) normal scatterers per asymmetric unit.

Atom		Space group category				Atom		Space group category							
$\langle k \rangle$	q	p	$\langle \sigma_1^2 \rangle$	1	3	5	6	$\langle k \rangle$	q	p	$\langle \sigma_1^2 \rangle$	1	3	5	6
Fe 0.24	100	1	24.5	72	70	68	66	Br*	100	1	45.3	84	79	78	73
		2	39.3	76	74	77	76			2	62.1	84	82	85	83
	300	1	9.8	62	61	60	58		300	1	21.9	71	69	67	65
		2	17.8	66	65	67	67			2	35.7	75	73	76	75
	500	1	6.1	58	58	57	55		500	1	14.4	66	65	63	61
		2	11.5	62	61	63	63			2	25.1	71	69	71	70
Co 0.29	100	1	23.2	71	69	67	65	I	100	1	65.8	93	84	84	76
		2	37.7	76	73	76	75			0.21	2	79.2	87	86	88
	300	1	9.2	61	60	59	58		300	1	39.5	76	72	71	68
		2	16.8	65	64	66	66			2	56.4	82	80	83	81
	500	1	5.7	58	57	56	55		500	1	28.3	75	72	71	67
		2	10.8	61	60	62	62			2	43.9	79	76	78	77

*Values for $MoK\alpha$ radiation.

Ba, Sm, Gd, Pt, Au, Hg, Pb and U are the heavy atoms generally used to prepare derivatives of proteins for facilitating structure determination. We shall consider situations where the following typical conditions are satisfied: (i) The asymmetric unit consists of about 1000 to 2000 normal scatterers besides 1 or 2 anomalous scatterers.

Table 6. Values of P_{QAM} (in %) expected in crystals belonging to the space group categories 1, 3, 5 and 6 and containing p ($=1, 2$ or 3) anomalous scatterers and q ($=1000, 1500$ or 2000) normal scatterers per asymmetric unit.

Atom				Space group category				Atom				Space group category			
$\langle k \rangle$	q	p	$\langle \sigma_1^2 \rangle$	1	3	5	6	$\langle k \rangle$	q	p	$\langle \sigma_1^2 \rangle$	1	3	5	6
Ba 0·19	1000	1	10·5	63	62	60	59	Au 0·12	1000	1	19·5	70	68	66	64
		2	19·0	67	65	68	67			2	32·5	74	72	74	74
		3	26·0	70	69	72	71			3	41·8	78	76	79	78
	1500	1	7·3	60	59	58	57		1500	1	13·9	66	65	63	61
		2	13·6	63	62	65	64			2	24·3	70	68	71	70
		3	19·0	66	66	65	68			3	32·5	74	72	75	74
	2000	1	5·6	58	58	56	55		2000	1	10·8	64	63	61	59
		2	10·5	62	60	62	62			2	19·5	68	66	68	68
		3	15·0	64	63	66	65			3	26·6	71	70	72	72
Sm 0·27	1000	1	11·3	63	62	61	59	Hg 0·12	1000	1	19·9	70	68	66	64
		2	20·3	68	66	68	68			2	33·1	75	72	75	74
		3	27·6	71	69	72	72			3	42·4	78	76	79	78
	1500	1	7·9	60	59	58	57		1500	1	14·3	66	65	63	61
		2	14·6	64	63	65	65			2	24·9	71	69	71	70
		3	20·3	67	66	69	68			3	33·1	74	73	75	75
	2000	1	6·0	58	58	56	55		2000	1	11·1	64	63	61	60
		2	11·3	62	61	62	62			2	19·9	68	66	69	68
		3	16·1	64	63	66	66			3	27·1	71	70	72	72
Gd 0·27	1000	1	10·6	63	62	60	59	Pb 0·13	1000	1	20·8	71	68	67	65
		2	19·1	67	65	67	67			2	34·2	75	73	75	74
		3	26·1	70	69	71	71			3	43·8	78	77	79	79
	1500	1	7·3	59	59	57	56		1500	1	14·9	67	66	64	62
		2	13·6	63	62	64	64			2	25·8	71	69	71	71
		3	19·1	66	65	68	67			3	34·2	74	73	76	75
	2000	1	5·6	58	57	56	55		2000	1	11·6	64	63	61	60
		2	10·6	61	60	62	62			2	20·8	68	67	69	69
		3	15·1	64	63	66	65			3	28·1	71	70	73	72
Pt 0·11	1000	1	18·9	70	68	66	64	U 0·18	1000	1	24·3	73	70	68	66
		2	31·7	74	72	74	73			2	39·0	77	74	77	76
		3	41·0	77	76	78	78			3	48·9	80	79	82	80
	1500	1	13·5	66	65	63	61		1500	1	17·7	68	66	65	63
		2	23·7	70	68	70	70			2	30·0	73	70	73	72
		3	31·7	73	72	74	74			3	39·0	76	74	77	77
	2000	1	10·5	63	62	60	59		2000	1	13·9	66	64	63	61
		2	18·9	67	66	68	68			2	24·3	70	68	70	70
		3	25·9	70	69	72	71			3	32·5	73	72	75	74

(ii) The composition of the q atoms is such that 80% of these are carbons, 10% are nitrogens and 10% are oxygens. (iii) The crystals are such that data corresponding to 2 Å resolution can be collected. The values of P_{QAM} for a few typical macromolecular situations are given in table 6. It is seen from table 6 that, in small proteins containing about 1000 to 1500 atoms, a value of P_{QAM} more than 0.7 can, in principle, be obtained by a proper choice of the heavy atom derivative. For example, for a crystal belonging to the space group category 5 and containing 1500 non-hydrogen atoms besides 3 uranium atoms per asymmetric unit, the value of P_{QAM} is 0.76. It may finally be noted that, if a number of heavy atom derivatives of a protein are available, the value of P_{QAM} will be the largest for the derivative for which the value of σ_1^2 is the largest and hence QAM will be the most effective for that derivative.

6. Verification of the theoretical results

The theoretical results derived in this paper were tested by using calculated data from two crystals (the hydrogens were excluded) and the details are given in table 7. The Bijvoet differences, normalized structure amplitudes of the inverse reflections and the values of $|\theta|$ were computed for the reflections in the range $(1/a_{\text{min}}) \leq (\sin \theta / \lambda) \leq 0.6$ where a_{min} is the minimum cell dimension of the crystal concerned*. The values of P_{QAM} for the two crystals computed from these results were taken to be the "observed values" of P_{QAM} . The mean values of σ_1^2 and k for these crystals were used to obtain the theoretical values of P_{QAM} from the results in table 2. Since the mean values of k for these examples are about 0.09, we have used the results corresponding to $k=0.1$ and computed the theoretical values of P_{QAM} by linear interpolation by treating σ_1^2 as the variable. It is seen from table 7 that the theoretical values of P_{QAM} agree reasonably well with the corresponding observed values.

Table 7. Details of the structures used for verifying the theoretical results.

Structure*	Asymmetric unit	space group	SGC	p	$\langle k \rangle$	$\langle \sigma_1^2 \rangle$	P_{QAM}	
							Observed (%)	Theoretical (%)
1	$\text{C}_4\text{N}_2\text{O}_4\text{Cl}$	P222	5	1	0.088	0.562	79	82
2	$\text{C}_{32}\text{N}_2\text{O}_{10}\text{Cl}_2$	P1	1	2	0.086	0.402	76	78

* Structure (1) is derived from that of Shamala and Venkatesan (1972) by excluding the hydrogen atoms. Structure (2) is similarly derived from that of Par Roques *et al* (1980). An over-all temperature factor of 2.7 Å² was used for structure (1) and 4.5 Å² was used for structure (2). These are the average values of the isotropic equivalents for the atoms in the original structures.

* For the theoretical reasons for omitting reflections for which $(\sin \theta / \lambda) < 1/a_{\text{min}}$ in statistical calculations see Wilson (1948).

Acknowledgement

One of the authors (SP) thanks the CSIR for financial support.

References

- Dale D, Hodgkin D C and Venkatesan K 1963 *Crystallography and Crystal Perfection* (ed.) G N Ramachandran (London: Academic press) p. 237
- Demidovich B P and Maron I A 1981 *Computational Mathematics* (Moscow: MIR)
- Foster P and Hargreaves A 1963 *Acta Crystallogr.* **16** 1124, 1133
- Geurtz T J H 1963 *Application of anomalous scattering methods in the structure determination of cytosine hydrochloride* Ph.D. Thesis, Rijks Universiteit, Utrecht
- Hall S R and Maslen E N 1965 *Acta Crystallogr.* **18** 265
- International Tables for X-ray Crystallography* Vol I 1969 (England: The Kynoch Press)
- Par Roques R, Rossi J C, Declerco E T J and German E T G 1980 *Acta Crystallogr.* **B36** 1589
- Parthasarathy S 1965 *Acta Crystallogr.* **18** 1028
- Parthasarathy S 1982 *Crystallographic Statistics* (eds) S Ramaseshan, M F Richardson and A J C Wilson (Bangalore: Indian Academy of Sciences) p. 133
- Parthasarathy S and Srinivasan R 1964 *Acta Crystallogr.* **17** 1400
- Parthasarathy S and Ponnuswamy M N 1976 *Acta Crystallogr.* **A32** 302
- Peerdemann A F and Bijvoet J M 1956 *Acta Crystallogr.* **9** 1012
- Ponnuswamy M N and Parthasarathy S 1977 *Acta Crystallogr.* **A33** 298
- Ramachandran G N and Raman S 1956 *Curr. Sci.* **25** 348
- Ramachandran G N and Srinivasan R 1970 *Fourier Methods in Crystallography* (New York: Wiley)
- Raman S 1959 *Z. Kristallogr.* **111** 301
- Shamala N and Venkatesan K 1972 *Cryst. Struct. Commun.* **1** 227
- Srinivasan R and Parthasarathy S 1976 *Some Statistical Applications in X-ray Crystallography* (New York: Pergamon press)
- Stout G H and Jenson L H 1968 *X-ray Structure Determination* (New York: The Macmillan Company)
- Velmurugan D and Parthasarathy S 1984 *Acta Crystallogr.* **40** 548
- Yow-lam Oh and Maslen E N 1966 *Acta Crystallogr.* **20** 852