is, if there are only two data sets, identical with the more special condition (2) given by Abrahams & Keve (1971); the sum over i in (A2) can be transformed into $(\delta m_h)^2$. Note that minimization of s in (A2) by simply putting $\partial s/\partial g_i = 0$, $i = 1, \ldots, n$, often does not converge, *cf.* Sparks (1970). The arrays in Fig. 3 were approximated by straight lines $\delta m_{\text{exper}} = a + b\delta m_{\text{expect}}$ and the values of a and b are given in Table 2 together with their e.s.d.'s.

The values of Q, σ_Q and s_Q in Table 2 and the plot given in Fig. 3(*a*) for the data before the χ^2 test show clearly one or both data sets to deviate from a normal distribution in accordance with the large Q values 1·14 and 1·13 given in Table 1. The Q values 0·96 and 0·98 for the two data sets after the χ^2 test may indicate slightly overestimated variances leading to a slope of the normal probability plot less than unity, which is in accordance with the numbers in Table 2 and with the plot in Fig. 3(*b*).

This work was made possible by a grant from the Deutsche Forschungsgemeinschaft and by the support of Professor Th. Hahn. The program for the least-squares refinement was written by E. Nowack and the drawings were prepared by R. Becker. The criticism of a referee of *Acta Cryst.* on the first version of this paper encouraged the author to develop the relations (9) to (13) and (21).

References

ABRAHAMS, S. C. (1964). *Acta Cryst.* **17**, 1327-1328.
ABRAHAMS, S. C. (1969). *Acta Cryst.* A**25**, 165-173.
ABRAHAMS, S. C. (1974). *Acta Cryst.* B**30**, 261-268.
ABRAHAMS, S. C., HAMILTON, W. C. & MATHIESON, A. MCL. (1970). *Acta Cryst.* A**26**, 1-18.
ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157-165.
ALTE DA VEIGA, L. M., ANDRADE, L. R. & GONSCHOREK, W. (1982). *Z. Kristallogr.* **160**, 171-178.
BLESSING, R. H., COPPENS, P. & BECKER, P. (1974). *J. Appl. Cryst.* **7**, 488-492.
DUNITZ, J. D. (1979). *X-ray Analysis and the Structure of Organic Molecules.* Ithaca: Cornell Univ. Press.
FLACK, H. D. (1984). *Methods and Applications in Crystallographic Computing*, edited by T. ASHIDA & S. HALL, pp. 53-67. Oxford Univ. Press.
FLACK, H. D., VINCENT, M. G. & VINCENT, J. A. (1980). *Acta Cryst.* A**36**, 495-496.
FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* A**34**, 517-525.
GONSCHOREK, W. (1981). *Port. Phys.* **12**, 49-60.
GONSCHOREK, W. (1982*a*). *Z. Kristallogr.* **158**, 197-204.
GONSCHOREK, W. (1982*b*). *Z. Kristallogr.* **160**, 187-203.
GONSCHOREK, W. & FELD, R. (1982). *Z. Kristallogr.* **161**, 1-5.
HAHN, TH. (1960). *Z. Kristallogr.* **113**, 403-429.
HAMILTON, W. C. & ABRAHAMS, S. C. (1970). *Acta Cryst.* A**26**, 18-24.
HAMILTON, W. C., ROLLETT, J. S. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129-130.
HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A**29**, 510-513.
IBERS, J. A. & HAMILTON, W. C. (1964). *Acta Cryst.* **17**, 781-782.
International Tables for X-ray Crystallography (1974). Vol. IV, edited by J. A. IBERS & W. C. HAMILTON, pp. 293-310. Birmingham: Kynoch Press.
JEFFERY, J. W. (1964). *Acta Cryst.* **17**, 1329.
JEFFERY, J. W. & ROSE, K. M. (1964). *Acta Cryst.* **17**, 343-350.
LEHMANN, M. S. (1975). *J. Appl. Cryst.* **8**, 619-622.
MCCANDLISH, L. E., STOUT, G. H. & ANDREWS, L. C. (1975). *Acta Cryst.* A**31**, 245-249.
NIELSEN, K. (1977). *Acta Cryst.* A**33**, 1009-1010.
REES, B. (1976). *Acta Cryst.* A**32**, 483-488.
REES, B. (1977*a*). *Isr. J. Chem.* **16**, 180-186.
REES, B. (1977*b*). *Isr. J. Chem.* **16**, 154-158.
REES, B. (1978). *Acta Cryst.* A**34**, 254-256.
ROLLETT, J. S. (1984). *Methods and Applications in Crystallographic Computing*, edited by T. ASHIDA & S. HALL, pp. 163-175. Oxford Univ. Press.
SCHULZ, H. (1971). *Acta Cryst.* A**27**, 540-544.
SHOEMAKER, D. P. (1968). *Acta Cryst.* A**24**, 136-142.
SPARKS, R. A. (1970). *Crystallographic Computing*, edited by F. R. AHMED, S. R. HALL & C. P. HUBER, pp. 282-284. Copenhagen: Munksgaard.
WAERDEN, B. L. VAN DER (1971). *Mathematische Statistik.* Berlin: Springer.
WILSON, A. J. C. (1976). *Acta Cryst.* A**32**, 781-783.
WILSON, A. J. C. (1979). *Acta Cryst.* A**35**, 122-130.

Acta Cryst. (1985). A**41**, 195-203

Optimal Strategy of Surveying Reciprocal Space in Protein Crystallography

BY S. V. NIKONOV AND YU. N. CHIRGADZE

Institute of Protein Research, USSR Academy of Sciences, 142292 *Pushchino, Moscow Region, USSR*

(*Received* 25 *May* 1984; *accepted* 6 *November* 1984)

Abstract

A general approach to the search for an optimal strategy of surveying reciprocal space is considered. It consists in determining a set of photographs that give the maximum number of structure factors in the minimum exposure time. Such a set of photographs satisfies the required strategy if the following initial conditions are fulfilled: the given completeness of the data set, the resolution limit, the crystal setting and the technical conditions of the method used for the data-set collection. A general algorithm is applied to the screened precession and screenless rotation-oscillation methods used in protein crystallography.

0108-7673/85/020195-09$01.50

Table 1. *Data collection statistics*

Protein and reference	Space group and cell parameters (Å)	Method of data collection	Formula of R factor	Compound	Resolution (Å)	Number of observations	Number of independent reflexions	R factor (%)
Beef liver catalase (Murthy *et al.*, 1981)	$P3_221$ $a = 104{\cdot}0$ $c = 104{\cdot}0$	Rotation	$\frac{\sum\sum\lvert F^2_{h_i} - \overline{F^2_h}\rvert}{\sum\sum \overline{F^2_h}}$	Native	2·5	168 293	40 624	10·3
				Mersalyl	2·5	119 902	32 538	9·2
				$PtCl_4$	2·5	91 303	31 742	8·5
Tokyo/3/67 neuraminidaze (Varghese, Laver & Colman, 1983)	$I422$ $a = 139{\cdot}6$ $c = 191{\cdot}0$	Rotation	$\frac{\sum\sum\lvert F^2_{h_i} - \overline{F^2_h}\rvert}{\sum\sum F^2_{h_i}}$	Native	3·1	45 266	10 953	11·9
				Diamino-dinitroplatinum	3·1	50 108	10 484	11·9
				Mercury phenyl glyoxal	3·1	28 325	8606	10·3
				Sodium tungsten	3·1	24 329	8363	11·8
				Sialic acid	3·1	26 138	9313	11·4
				Lactose	3·1	28 401	8667	11·3
L-arabinose-binding protein (Gilland & Quiocho, 1981)	$P2_12_12_1$ $a = 55{\cdot}5$ $b = 71{\cdot}8$ $c = 77{\cdot}8$	Diffractometer	$\frac{\sum\sum\lvert F^2_{h_i} - \overline{F^2_h}\rvert}{\sum\sum \overline{F^2_h}}$	Native	2·4	37 737	12 645	3·4
				CdI_2	2·4	19 228	10 885	3·5
				pCMBS	2·4	32 313	12 508	5·4
Aspartate carbamoyl transferase (Honzatko *et al.*, 1982)	$P321$ $a = 122{\cdot}1$ $c = 142{\cdot}1$	Rotation	$\frac{\sum\sum\lvert F^2_{h_i} - F^2_{h_j}\rvert}{\sum\sum\lvert F^2_{h_i} + F^2_{h_j}\rvert}$	Native	2·8	146 156	24 369	9·13
				MSA	2·8	74 777	22 525	9·41
				$UO_2(NO_3)_2$	2·8	50 666	22 277	10·51
				$KAu(CN)_2$	2·8	69 171	22 599	9·83
Cytochrome c' (Weber *et al.*, 1980)	$P2_12_12_1$ $a = 56{\cdot}6$ $b = 71{\cdot}8$ $c = 75{\cdot}6$	Multiwire area detector	*	Native	2·5	81 380	18 513	6·2
				K_2PtCl_6	2·5	71 072	10 881	7·4
				K_2HgI_4	2·5	67 994	14 082	6·8
Glyceraldehyde 3-phosphate dehydrogenase (Leslie & Wonacott, 1983)	$P2_12_12_1$ $a = 132{\cdot}7$ $b = 125{\cdot}8$ $c = 98{\cdot}9$	Rotation	*	Native	2·5	72 135	27 783	5·6
				$PtCl_4$	2·5	50 449	17 449	11·0
				HgI_4	2·5	75 118	23 698	9·4

Note: $F^2_{h_i}$ is the scaled intensity from the ith film for reflexions h. $\overline{F^2_h}$ is the weighted mean value for all observations of reflexion h.
* Formula for R factor was not reported by the authors of cited papers.

The proposed method allows also quantitative evaluation of different strategies of surveying. The usual intuitive approach is shown to overestimate the number of photographs. The optimal strategy reduces the total exposure time by up to 50%, which leads to a decrease in the necessary number of single crystals.

Introduction

Protein molecules crystallize in large unit cells and are subject to radiation damage. Therefore it is important to use the most efficient method for collecting three-dimensional structure factor data, *i.e.* the given number of structure factors should be measured in the minimum time (Arndt, 1968; Blundell & Johnson, 1976; Schutt & Winkler, 1977). Two approaches to this problem are available. The first is the development of new techniques, for example, to improve the quality of the X-ray beam and X-ray detectors. The second concerns the optimum strategy of surveying reciprocal space. In the following discussion we shall consider only the second approach.

Schemes of surveying reciprocal space can be classified as zero-, one-, two- or three-dimensional depending on whether they survey a point, a line, a plane or a volume element of this space (Arndt, 1968). Single-counter diffractometers, screened precession and oscillation cameras are widely used in protein crystallography. These apparatus use zero- two- and three-dimensional schemes, respectively.

The zero-dimensional scheme provides wide possibilities for selecting a way of surveying reciprocal space. In this case the optimum strategy reduces the time of measurements of the intensity of the individual diffraction maximum (Blundell & Johnson, 1976). However, the strategy of surveying is much more important for the two- and three-dimensional schemes because of recording a large number of symmetry-related reflexions. Therefore, the number of spots measured on the photographs exceeds the number of points in the asymmetric part of the reciprocal lattice (Honzatko *et al.*, 1982; Murthy, Garavito, Johnson & Rossmann, 1980; Weber, Paramokos, Bode, Huber, Kato & Laskowski, 1981).

It should be noted that the measurement of a large number of equivalent reflexions does not ensure the same increase of the structure-factor accuracy (Table 1). For example, the sets of structure factors of heavy-atom derivatives of beef liver catalase (Murthy, Reid,

Sicigano, Tanaka & Rossmann, 1981) have a smaller number of equivalent reflexions, in comparison with the native crystal, but the values of R factors of these sets are lower. Therefore one can decrease the exposure time and obtain the required accuracy of structure factors even if the number of equivalent reflexions is smaller.

The number of ways of surveying reciprocal space increases essentially if one surveys only a part of the whole asymmetric unit. In this case, a loss of the quality of the Fourier synthesis is inevitable. However, there are many examples of successful determination of three-dimensional protein structures when only 70–90% of the asymmetric unit was surveyed (Donald, Musick & Rossmann, 1979; Stammers & Muirhead, 1977; Weber *et al.*, 1981).

The ways of searching for the optimum strategy of surveying reciprocal space have been discussed by Schutt & Winkler (1977), who showed that about 90% of the reciprocal-lattice points may be recorded by the oscillation method for Laue group $m3$ if the crystal is rotated only within limits of about 20°. The same results have been obtained for the screened precession method (Nikonov & Chirgadze, 1981). In this paper a general approach is proposed to search for the optimum strategy of surveying reciprocal space. This approach has been applied to the screened precession and oscillation methods of data collection.

Principles of optimal planning of surveying reciprocal space

A series of contiguous photographs should be taken to collect a three-dimensional data set of structure factors. To select the optimum data-collection strategy, it is necessary to know, at least, two parameters of the structure-factor set: resolution limit, $d_{\min}$, and data-set completeness, η. The η value can be determined as the ratio of the number of the structure factors measured in the given set and the number of points in the asymmetric unit of reciprocal space. Three factors limit the number of ways of surveying reciprocal space: (1) the geometry of the data-collection methods and technical parameters of the camera; (2) the necessity of bringing the photographs to a common scale; (3) the type of crystal setting.

The first is connected mainly with the possibility of recording any photograph that should be taken to collect the three-dimensional data set. The second determines the required minimum number of common reflections, N_0^c, on different photographs (Hamilton, Rollett & Sparks, 1965; Rossman, Leslie, Abdel-Meguid & Tsukihara, 1979).

The best crystal setting will correspond to the minimum number of required photographs. In the screened precession method, the unique axis of the highest Laue symmetry must be set along the camera spindle axis (Azaroff, 1954; Buerger, 1964). In the rotation method the same setting has been shown to be appropriate for crystals producing all Laue groups (Wonacott, 1977) except the cubic one. In the case of the cubic $m3$ Laue group the $[10\bar{1}]$ direction should be set along the camera spindle axis (Schutt & Winkler, 1977). However, in practice for some reason the optimum crystal setting is often impossible. The planning of a survey of reciprocal space with different crystal settings allows the optimum strategy of data collection to be found.

The problem of the optimum way of data collection consists in the search for a method of surveying the reciprocal space that allows one to obtain the required set of photographs with the minimum total exposure time. It is possible to solve this problem in the following way. Let us find a countable *initial set M of photographs*, where M is the number of all photographs to be recorded at the given conditions. To create this set, it is necessary to take into account the symmetry and setting of the crystal, the data set resolution and the technical parameters of the camera. Then we can select a countable subset, m, of photographs that satisfies the condition of optimization

$$T(m) = \min, \quad \text{if } \eta \geq \eta_0 \text{ and } N_i^c \geq N_0^c, \qquad (1)$$

where $T(m)$ is the total exposure time of the set m of photographs, η_0 is the given completeness of the data set and N_i^c is the number of common reflections on the ith photograph.

A correct solution of this problem presents considerable computational difficulties. Indeed, the structure-factor set consists of a large number of independent reflexions, which can be measured from a large number of photographs. However, the number of structure factors added to the set by every subsequent photograph depends on the completeness of this set, as some reflexions may be recorded on few photographs of the initial set.

We propose to search for the solution of this problem using the following approximation. We will create the required optimum set of photographs by a contiguous selection of photographs with maximum efficiency from the initial set of photographs. The efficiency of a photograph can be determined as

$$\varepsilon_i(m) = \Delta N_i(m)/T_i, \qquad (2)$$

where $\Delta N_i(m)$ is the number of structure factors added by the ith photograph to the m set of photographs, T_i is the exposure time of the ith photograph.

The number of added structure factors from the ith photograph is determined as

$$\Delta N_i(m) = N_i - N_i^s - N_i^c(m), \qquad (3)$$

where N_i is the total number of reflexions on the ith photograph, N_i^s is the number of symmetry-related reflexions, $N_i^c(m)$ is the common number of symmetry-independent reflexions on the ith photograph

that has been recorded on the other photographs, already inserted into the data set.

The exposure time of any photograph can be evaluated from the Lorentz factor of reflexions that are surveyed in the most unfavourable conditions. These are to use the most 'rapid' reciprocal-lattice points passing through the Ewald sphere with the maximum velocity.

The logical flowchart for creating the optimum set of photographs is shown in Fig. 1. This algorithm is valid for any method of data collection but the details of the calculation are very specific. We have realized this algorithm for both the screened precession and oscillation methods.

Optimization of surveying reciprocal space in the screened precession method

The optimization of surveying reciprocal space in the screened precession method was discussed by us earlier (Nikonov & Chirgadze, 1981). It was shown

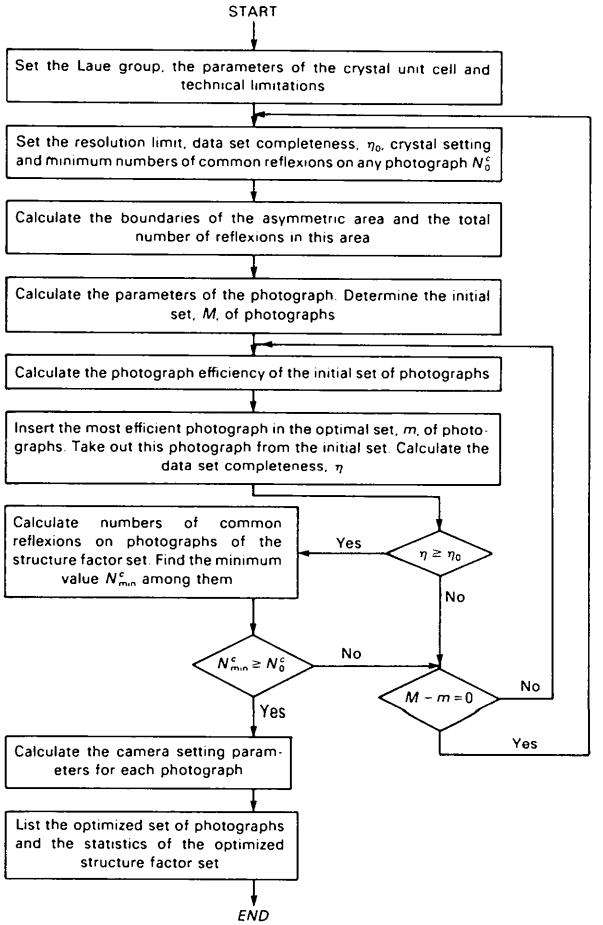

Fig. 1. Flowchart of the algorithm for an optimizing survey of reciprocal space. M is the number of photoglraphs in the initial set, m is the number of photographs in the data set, η is the completeness of the data set.

that the parameters of an initial set of photographs can be calculated in the spherical area of reciprocal space with the radius $d^*_{min} = \lambda / d_{min}$ using the following relations:

$$\mu_i = \sin^{-1}\frac{d^*_{min}}{2} - \sin^{-1}\frac{\zeta_i}{d^*_{min}}, \tag{4}$$

$$\nu_i = \sin^{-1}\frac{d_{min}}{2} + \sin^{-1}\frac{\zeta_i}{d^*_{min}}, \tag{5}$$

$$\xi_{i(max)} = d^*_{min}\cos\left(\sin^{-1}\frac{\zeta_i}{d^*_{min}}\right), \tag{6}$$

$$\xi_{i(min)} = 2\frac{\zeta_i}{d^*_{min}}\cos\left(\sin^{-1}\frac{d^*_{min}}{2}\right), \tag{7}$$

where λ is the wavelength of X-rays; μ_i is the precession angle; ν_i is the half-angle of Laue cone, $\xi_{i(max)}$ and $\xi_{i(min)}$ are the outer and inner radii of the exposure zone on a photograph; $\zeta_i = m_j \Delta\zeta_j$ is the distance of the ith plane from the reciprocal-lattice origin; m_j is the number of planes in a plane family; $\Delta\zeta_j$ is the interplanar spacing in the jth family.

The values (4)-(7) are the functions of only two variables d^*_{min} and $\Delta\zeta_j$. At the given crystal setting the spacing $\Delta\zeta_j$ determines the type of the plane family. To select the plane family, one must take into account the symmetry of reciprocal space, the population of points on the plane and the technical parameters of the precession camera. Sometimes it is impossible to obtain the required completeness of the data set at a given resolution. In this case the d^*_{min} value should be increased or the η value decreased.

The calculation of the $\Delta N_i(m)$ value using the screened precession method is connected with considerable difficulties as the planes of different families intersect, which leads to multiple recording of some groups of reflexions. The situation may be simplified by using only one crystal setting. In this case, all reflexions are arranged in parallel rows. Then it is possible to find the non-integer indices h_t, k_t, l_t of the row ends situated on the sphere of the d^*_{min} radius from the expression

$$\mathbf{d}_{min} = h_t\mathbf{a}^* + k_t\mathbf{b}^* + l_t\mathbf{c}^*, \tag{8}$$

where $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$ are axial reciprocal-lattice vectors.

In this expression only one index is independent and the two others are determined by the crystal orientation and the row position. Denoting the independent index as A_j we can determine the length of a row as

$$L_j = (A_{j2} - A_{j1})t^*_j, \tag{9}$$

where A_{j1}, A_{j2} are values of the index A_j at the sphere surface, t^*_j is the reciprocal-lattice translation along the row direction.

The total number of reflexions that belong to the jth row and are recorded from the ith photograph

(Fig. 2) is expressed as

$$N_{ij} = \text{Integer}\left\{\frac{L_j - [L_j^2 - 4(\xi^2_{i(\max)} - \xi^2_{i(\min)})]^{1/2}}{t_j^*}\right\}. \quad (10)$$

Integer here means only the integer part of the expression.

Hence, the number of reflexions added to the jth row of the structure-factor set by the ith photograph is determined as

$$\Delta N_{ij}(m) = N_{ij} - N^s_{ij} - N^c_j(m), \quad (11)$$

where N^s_{ij} is the number of symmetry-related reflexions of the jth row recorded on the ith photograph, $N^c_j(m)$ is the number of symmetry-independent common reflexions of the jth row inserted earlier in the structure-factor set. Then

$$\Delta N_i(m) = \sum_{j=1}^{n_a} \Delta N_{ij}(m), \quad (12)$$

where n_a is the number of symmetry-independent rows on the photographs of the structure-factor set.

Now let us evaluate the exposure time of any photograph. It was shown (Nikonov & Chirgadze, 1981) that in the precession method the Lorentz factor of the most 'rapid' reciprocal-lattice points is proportional to the area of the exposure zone of a photograph. Then

$$T_i \simeq \xi^2_{i(\max)} - \xi^2_{i(\min)}, \quad (13)$$

and the efficiency of the ith photograph is

$$\varepsilon_i(m) = \frac{\sum_{j=1}^{n_a} \Delta N_{ij}(m)}{\xi^2_{i(\max)} - \xi^2_{i(\min)}}. \quad (14)$$

The possibilities of the screened precession method can be improved considerably by recording two or more reciprocal-lattice levels on one photograph. The conditions of such a recording were determined in a previous paper (Nikonov & Chirgadze, 1981).

Optimization of surveying reciprocal space in the rotation–oscillation method

In the rotation method the recording elements of reciprocal space are similar to one another at the given resolution and crystal setting. These elements differ only by the values of oscillation angles $\Delta\varphi_i$ and the population of reciprocal-lattice points in their volumes. A large value of the oscillation angle leads to a decrease of the relative number of partially recorded reflexions. On the other hand, the maximum permissible range of $\Delta\varphi_i$ must secure the minimum number of overlapping reflexions and the optimum signal-to-noise ratio (Wonacott, 1977). Sometimes the value of $\Delta\varphi_i$ is limited by the radiation damage of the crystal and the collimation conditions.

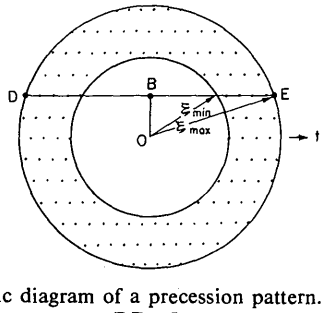

Fig. 2. Schematic diagram of a precession pattern. In this figure $DE = L_j$.

In order to obtain the maximum number of structure factors for one setting of the crystal, it is necessary to rotate the crystal within the angle (Wonacott, 1977)

$$\varphi_{\max} = 2\pi/n \quad (15)$$

provided the n-fold axis of the crystal symmetry is aligned parallel to the rotation axis. Then the number of non-overlapping photographs M in the initial set is determined from the equation

$$\sum_{i=1}^{M} \Delta\varphi_i = \varphi_{\max}. \quad (16)$$

We have suggested that symmetry-related reflexions can be recorded only on one or two photographs, no more. Indeed, the symmetric region of a recording element contains two identical parts; one of them enters the Ewald sphere and the other leaves it. If the rotating area $\varphi \leq \varphi_{\max}$, the whole exposed volume of the reciprocal space cannot exceed the double volume of the asymmetric area. Therefore, if a reciprocal-lattice point is assumed to have three or more symmetry equivalents in the exposed zone, then at least the $(n-2)$ unexposed symmetry-related parts of the reciprocal space must contain $(n-3)$ or fewer symmetry-related points. However, this is impossible. The deduced regularity permits us to represent the distribution of reflexions between photographs of the initial set as a square symmetric matrix with the element a_{ij}. Diagonal elements of this matrix correspond to the number of reflexions, the symmetry-related equivalents of which are recorded on the ith photograph, and non-diagonal elements correspond to the number of common reflexions, the equivalents of which are recorded on the ith and jth photographs.

According to Wonacott (1977), the reciprocal-lattice point is located on the surface of the Ewald sphere if the rotation angle of the crystal is determined as

$$\varphi_h = \cos^{-1}(-d^{*2}/2\xi) - \tau, \quad (17)$$

where

$$d^{*2} = x_0^2 + y_0^2 + z_0^2,$$
$$\xi = (x_0^2 + y_0^2)^{1/2},$$
$$\tau = \tan^{-1}(y_0/x_0),$$

and values of x_0, y_0 and z_0 can be found from the

equations

$$x_0 = ha^*_{x_0} + kb^*_{x_0} + lc^*_{x_0},$$
$$y_0 = ha^*_{y_0} + kb^*_{y_0} + lc^*_{y_0},$$
$$z_0 = ha^*_{z_0} + kb^*_{z_0} + lc^*_{z_0},$$

where $a^*_{x_0}$, $a^*_{y_0}$, $a^*_{x_0}$; $b^*_{x_0}$, $b^*_{y_0}$, $b^*_{z_0}$; $c^*_{x_0}$, $c^*_{y_0}$, $c^*_{z_0}$ are projections of the reciprocal-lattice vectors $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$ on the axes $\mathbf{x}_0$, $\mathbf{y}_0$ and $\mathbf{z}_0$. We define a standard orientation of the crystal at $\varphi = 0$, when the z_0 axis coincides with the rotation axis and the x_0 axis coincides with the primary X-ray beam.

Two solutions for φ_h obtained from (17) correspond to the two rotation angles at which the reciprocal-lattice point intersects the sphere of reflexions. The reflexion is recorded on the ith photograph if

$$\varphi_{i0} < \varphi_{hi} \leq \varphi_{i1}, \tag{18}$$

where φ_{i0} and φ_{i1} are initial and final values of the rotation angle of the ith photograph. One should search for the distribution of the reflexions for the whole of reciprocal space of the given resolution although only some of the reciprocal-lattice points will answer the condition $0 \leq \varphi_{hi} \leq \varphi_{\max}$. This problem can be solved with a multiplication of asymmetric points by symmetry operations.

The matrix of the distribution of reflexions among the initial set of photographs allows us to calculate the number of structure factors added to the data set by any given photograph. The number of structure factors for the jth photograph is

$$\Delta N_j(m) = \sum_{k=1}^{M} a_{jk} - \sum_{l \neq j}^{m} a_{jl}, \tag{19}$$

where M and m are the numbers of photographs in the initial set and the optimized structure-factor set, respectively; a_{jk} and a_{jl} are elements of the symmetric matrix.

The value of the Lorentz factor in the rotation method is independent of the parameters of a photograph. Therefore the velocity of the points passing through the sphere is the same for the corresponding points of different photographs if the angular velocity of the crystal is not changed. In this case, the exposure time is proportional to the oscillation angle, $\Delta\varphi_j$, and the efficiency of photographs is determined as

$$\varepsilon_j(m) = \Delta N_j(m)/\Delta\varphi_j. \tag{20}$$

In some special cases the strategy of surveying reciprocal space can provide a recording with overlaps between contiguous photographs. Such a strategy is needed when the lifetime of the crystal is comparable with the exposure time of a photograph, or when it is desirable to obtain the maximum number of fully-recorded reflexions. The corresponding initial set will be similar to the initial set determined by (16) if each region of the overlap is considered as a separate photograph. Therefore the efficiency of photographs is calculated by the same formula (20). But each photograph of the structure-factor set will contain two or three photographs of the initial set.

Results

The algorithm for surveying reciprocal space has been tested during the planning of the data collection of different protein crystals using the screened precession and rotation methods. The Nova-3/12 computer with 64 k byte core memory has been used.

A compilation of the structure-factor set following the optimization procedure shows that the efficiency of a photograph decreases as the set completeness increases, being independent of the Laue group and dimensions of the unit cell of the crystal. At the same time the number of redundancy measurements for some reflexions increases. Let us call the ratio of the number of reflexions on the photograph to the number of structure factors in the data set the *redundancy of*

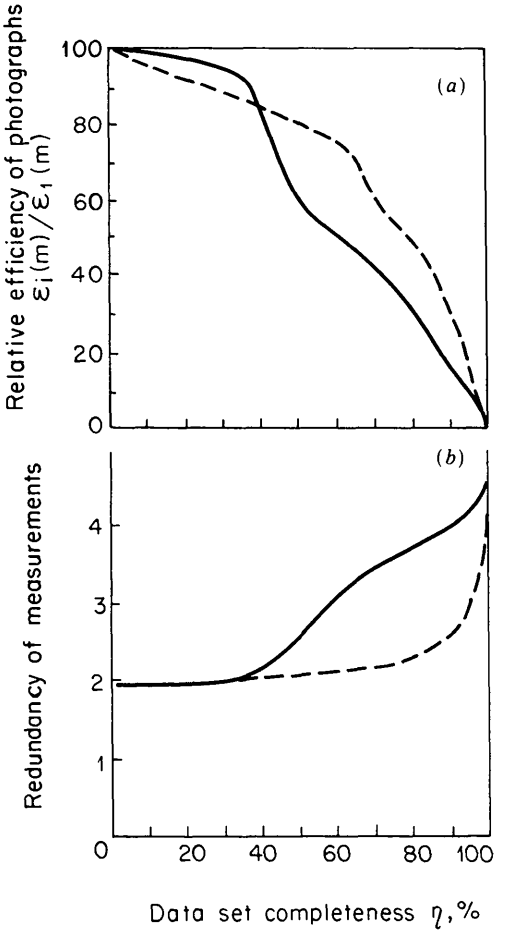

Fig. 3. (*a*) Relative efficiency of photographs $\varepsilon_i(m)/\varepsilon_1(m)$ and (*b*) redundancy of measurements *versus* the completeness of the data set for γ-crystallin IIIb. Solid line, precession method; dotted line, rotation method.

measurements. The ratio of the efficiency of the *i*th photograph inserted into the data set to the efficiency of the first photograph of the data set will be called the *relative efficiency of the ith photograph.* The dependences of these values on the completeness of the structure-factor set are shown in Fig. 3. The curves are calculated for the eye-lens protein γ-crystallin IIIb ($P2_12_12_1$, $a=58{\cdot}7$, $b=69{\cdot}5$, $c=116{\cdot}9$ Å) at 3 Å resolution for the screened precession and rotation methods. As can be seen, these dependences are less pronounced if the completeness of the data set is less than 40%. This means that the connections between separate photographs are weak. In this case the redundancy of the measurements can be explained by the symmetry of an individual photograph. If the value of η is increased, the number of repeatedly recorded reflexions will increase and the connections between photographs will become stronger. Hence, the extension of the completeness of the structure-factor set can be used to bring the data to a common scale instead of using a perpendicular setting of the crystal (Colman, Jansonius & Matthews, 1972; Matthews, Fenna, Bolognesi, Schmid & Olson, 1979). The rapid increase of the redundancy of measurements at $\eta>40\%$ in the screened precession method can be explained by the insertion of high-symmetry *hnl* planes into the data set of γ-crystallin IIIb, whereas at $\eta<40\%$ the data set contains only the $h,h-n,l$ planes.

The relationship between the completeness of the data set and the ratio of the total exposure time of the optimized set of photographs to that of the initial set $T(m)/T(M)$ is shown in Fig. 4. The plots have been calculated for γ-crystallin IIIb at different values of d_{min} using only the rotation method. It is clear that all the curves are close to one another if $d_{min}>2$ Å. The curve at $d_{min}=1{\cdot}5$ Å can be explained by the presence of a blind area at this resolution (Wonacott, 1977). For the rotation method, the asymmetric area has been shown to be exposed twice when the crystal rotates within the angle φ_{max}. Therefore the curves shown in Fig. 4 begin to change their curvature essentially if $50\leq T(m)/T(M)\leq 60\%$. The corresponding values of η are from 75 to 90%. *This feature is typical of all Laue groups and is independent of the dimensions of the crystal unit cell.*

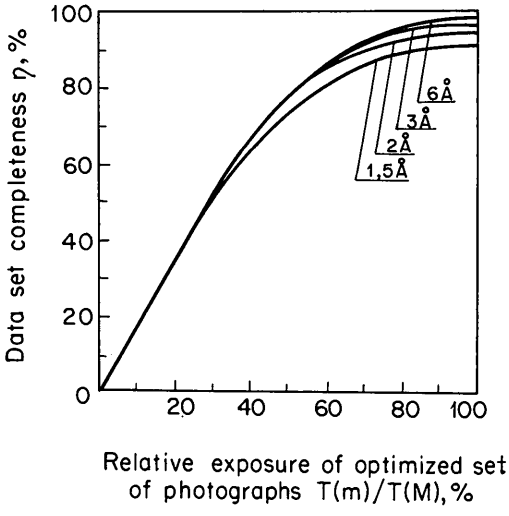

Fig. 4. Relationship between the completeness and the total relative exposure time $T(m)/T(M)$ of the data set for γ-crystallin IIIb in the rotation method. $T(m)$, the total exposure time of the optimized set of photographs; $T(M)$, the total time of the initial set of photographs.

Taking into account the fact that in many cases $\eta=70$–90% secures the reliability of the interpretable Fourier syntheses (Donald *et al.*, 1979; Stammers & Muirhead, 1971; Howard, Xuong & Salemme, 1981; Leslie & Wonacott, 1983), we can suppose that $\eta=85\%$ is close to the optimal value. It should be noted that the *a priori* η value can be obtained in the optimum way only with indispensable photographs of the initial set. Therefore, the intuitive selection of photographs for the data set very often leads to an increase of the exposure time. For instance, the completeness $\eta=82.2\%$ has been obtained for the glyceraldehyde 3-phosphate dehydrogenase crystal (Leslie & Wonacott, 1983), belonging to the Laue group *mmm*, by a crystal rotation through an angle of 62°, whereas the selection of only optimized indispensable photographs for such η value lowers this angle to 49°.

A comparison of some intuitive and optimized approaches to the collection of the protein data set by the rotation method is presented in Table 2. In the case of the optimized approach the exposed zone may contain a few separate parts. The number of these parts depends on the crystal symmetry and the resolution limit. If $d_{min}>2{\cdot}5$ Å the total value of the exposed zone does not exceed 50–60% of its maximum value. This means that the exposure time is shortened by about 50%. It should be noted that the completeness $\eta=85\%$ is enough to bring the photographs to a common scale, even if the photographs have not been overlapped in their angle ranges.

Analogous results have been obtained using the screened precession method. A comparison of the intuitive and optimized approaches is given in Table 3. In this case the given η value for the optimal set ranged from 85 to 95% as follows from reference data. The optimization is seen to reduce the number of photographs to 30%. It should be noted that the reduction is much smaller for high-symmetry Laue groups and at a high resolution, because the planning of the data collection is simplified in these cases. The omission of zero-level planes belonging to families with high Miller indices is a peculiarity of the optimized approach in the screened precession method.

Therefore, the optimum strategy of surveying reciprocal space allows us to decrease the exposure

Table 2. *Comparison of non-optimized and optimized data sets for protein crystals obtained by rotation method*

Protein and reference	Laue group	Resolution limit (Å)	Maximal rotation angle (°)	Measured data sets (see references)			Optimized data sets			Exposure time gain, $\frac{(\varphi_t-\varphi_r)}{\varphi_t}$ (%)
				Completeness η (%)	Rotation angle range (°)	Total rotation area, φ_t (°)	Completeness, η (%)	Rotation angle range (°)	Total rotation area, φ_r (°)	
Apo-D-glyceraldehyde 3-phosphate dehydrogenase (Murthy *et al.*, 1980)	1	3·0	180	100	0-99 (b^*) 0-117 (c^*)	216	85	0-36 67-89 115-146	89	58·8
Methionyl-tRNA synthetase (Zelwer, Risler & Brunie, 1982)	$2/m$	2·5	90	96·8	0-90	90	85	0-18 43-74	47	47·8
Glyceraldehyde 3-phosphate dehydrogenase (Leslie & Wonacott, 1983)	*mmm*	2·5	90	82·2	0-62	62	85	18-69	51	17·7
Pyruvate kinase (Stuart, Levine Muirhead & Stammers, 1979)	*mmm*	2·6	90	91	0-47 (b^*) 0-24 (a^*)	71	85	12-45 73-90	50	29.6
Phosphorylase b (Johnson, Stura, Wilson, Sansom & Weber, 1979)	$4/mmm$	3·0	45	97·5	0-45	45	85	9-34	25	44·4
Beef liver catalase (Murthy *et al.*, (1981)	$\bar{3}\,m$	2.5	30	96.8	0-30	30	85	13-30	17	43.3
Thermolysin (Holmes & Matthews (1982)	$6/mmm$	1.6	30	92.0	0-30	30	85	0-9 14-27	22	26·7

Note: If not otherwise indicated the c axis of the crystal has been used as rotation axis. The oscillation angle $\Delta\varphi_t$ is calculated to avoid overlapping of reflexions. The number of common reflexions of a photograph exceeds 150 for all optimized sets.

Table 3. *Comparison of non-optimized and optimized data sets for protein crystals obtained by screened precession method*

Protein and reference	Laue group	Resolution limit (Å)	Completeness of each set (%)	Numbers of photographs in family planes								Optimized set						Gain in the number of photographs, $\frac{(N_r-N_o)}{N_r}$ (%)
				Measured data (see references)														
				$h\,n\,l$	$n\,k\,l$	$h,n-n,l$	$2k+n,k,l$	$3k+n,k,l$	$h,2h-n,l$	$h\,k\,n$	Total, N_r	$h\,n\,l$	$n\,k\,l$	$h,h-n,l$	$2k+n,k,l$	$3k+n,k,l$	Total, N_o	
Pyruvate kinase (Stammers & Muirhead 1977)	*mmm*	3·1	85	0-4	0-6	0-7	0-10	—	0-11	—	43	0-5	0-7	0-9	1-4	1-2	30	30·2
γ-Crystallin IIIb (Chirgadze, Sergeev, Fomenkova, Nikonov & Lunin, 1980)	$4/mmm$	4.8	95	0-1	—	0-2	0-2	0-5	—	—	14	0-1	—	0-2	0-3	1-2	11	21·4
Lysozyme from bacteriophage T4 (Remington *et al.*, 1978)	$\bar{3}\,m$	2·4	92	0-5	—	0-5	—	—	—	—	12	0-5	—	0-5	—	—	12	0
Bacteriochlorophyll a-protein (Matthews *et al.*, 1979)	$6/m$	2·8	90	0-8	—	0-9	—	—	—	0-3	23	0-7	—	0-11	—	—	20	13·0
Thermolysin (Colman *et al.*, 1972)	$6/mmm$	2.4	93	0-8	—	0-8	—	—	—	0-3	22	0-9	—	0-7	—	—	18	22·2

Note: The completeness of optimized sets is given in column 4.

time without a deterioration of the quality of the structure-factor set. The largest effect is obtained for crystals with a large unit cell, crystals having weak reflexions and those suffering from radiation damage.

We wish to thank Dr V. Lunin for valuable discussions.

References

ARNDT, U. W. (1968). *Acta Cryst.* **B24**, 1355-1357.

AZAROFF, L. V. (1954). *Rev. Sci. Instrum.* **25**, 928-929.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* New York, London, San Francisco: Academic Press.

BUERGER, M. J. (1964). *The Precession Method in X-ray Crystallography.* New York: John Wiley.

CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P., NIKONOV, S. V. & LUNIN, V. YU. (1980). *Dokl. Akad. Nauk SSSR*, **250**, 762-765. English trans.: *Proc. USSR Acad. Sci.* **250**, 42-45.

COLMAN, P. M., JANSONIUS, J. N. & MATTHEWS, B. W. (1972). *J. Mol. Biol.* **70**, 701-724.

DONALD, W., MUSICK, L. & ROSSMANN, M. G. (1979). *J. Biol. Chem.* **254**, 7611-7620.

GILLAND, G. L. & QUIOCHO, F. A. (1981). *J. Mol. Biol.* **146**, 341-362.

HAMILTON, W. C., ROLLETT, J. C. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129-130.

HOLMES, M. A. & MATTHEWS, B. W. (1982). *J. Mol. Biol.* **160**, 623-639.

HONZATKO, R. B., CRAWFORD, J. L., MONAKO, H. L., LADNER, J. E., EWARDS, B. F. P., EVANS, D. R., WARREN, S. G., WILEY, D. C., LADNER, R. C. & LIPSCOMB, W. N. (1982). *J. Mol. Biol.* **160**, 219-263.

JOHNSON, L. N., STURA, E. A., WILSON, K. S., SANSOM, M. S. P. & WEBER, I. T. (1979). *J. Mol. Biol.* **134**, 639-653.

LESLIE, A. G. W. & WONACOTT, A. J. (1983). *J. Mol. Biol.* **165**, 375-391.

MATTHEWS, B. W., FENNA, R. E., BOLOGNESI, M. C., SCHMID, M. F. & OLSON, J. M. (1979). *J. Mol. Biol.* **131**, 259-285.

MURTHY, M. R. N., GARAVITO, R. M., JOHNSON, J. E. & ROSSMANN, M. G. (1980). *J. Mol. Biol.* **138**, 859-872.

MURTHY, M. R. N., REID, T. J. III, SICIGNANO, A., TANAKA, N. & ROSSMANN, M. G. (1981). *J. Mol. Biol.* **152**, 465-499.

NIKONOV, S. V. & CHIRGADZE, YU. N. (1981). *Kristallografiya*, **26**, 687-694 English trans.: *Sov. Phys. Crystallogr.* **26**, 390-394.

REMINGTON, S. J., ANDERSON, W. F., OWEN, J., TEN EYCK, L. F., GRAINGER, C. T. & MATTHEWS, B. W. (1978). *J. Mol. Biol.* **118**, 81-98.

ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570-581.

SCHUTT, C. & WINKLER, F. K. (1977). In *The Rotation Method in Crystallography.* Amsterdam: North-Holland.

STAMMERS, D. K. & MUIRHEAD, H. (1977). *J. Mol. Biol.* **112**, 309-316.

STUART, D. I., LEVINE, M., MUIRHEAD, H. & STAMMERS, D. K. (1979). *J. Mol. Biol.* **134**, 109-142.

VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). *Nature (London)*, **303**, 35-40.

WEBER, E., PARAMOKOS, E., BODE, W., HUBER, R., KATO, I. & LASKOWSKI, M. JR (1981). *J. Mol. Biol.* **149**, 109-123.

WEBER, P. C., BARTSH, R. G., CUSANOVICH, M. A., HAMLIN, R. C., HOWARD, A., JORDAN, S. R., KAMEN, M. D., MEYER, T. E., WEATHERFORD, D. W., XUONG, N. & SALEMME, F. R. (1980). *Nature (London)*, **286**, 302-304.

WEBER, P. C., HOWARD, A., XUONG, N. & SALEMME, F. R. (1981). *J. Mol. Biol.* **153**, 399-424.

WONACOTT, A. J. (1977). In *The Rotation Method in Crystallography.* Amsterdam: North-Holland.

ZELWER, C., RISLER, J. L. & BRUNIE, S. (1982). *J. Mol. Biol.* **155**, 63-81.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

Acta Cryst. (1985). A**41**, 203-204

Revised space-group frequencies for organic compounds.*

By JERRY DONOHUE, *Department of Chemistry and Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, Pennsylvania* 19104, *USA*

(*Received* 5 *June* 1984; *accepted* 25 *September* 1984)

Abstract

In a recent paper Mighell, Himes & Rodgers [*Acta Cryst.* (1983), A**39**, 737-740] reported the space-group frequency for nearly 30 000 organic compounds in the *NBS Crystal Data Identification File* [(1982). JCPDS, Swarthmore, PA]. When the frequencies tabulated by Mighell *et al.* for those space groups that do not contain any operations of the second kind are revised somewhat different results are obtained. (Numbers given by Mighell *et al.* are given in parentheses.): 75% (75%) of compounds fall into only five space groups: $P2_1/c$, 29·2% (36.0%), $P2_12_12_1$, 18·8% (11·6%), $P\bar{1}$, 11·1% (13·7%), $P2_1$, 10·9% (6.7%) and $C2/c$, 5·4% (6·6%); 12(16) space groups account for 87% (90%) of the compounds.

* Supported in part by Research Grant No. GM20611 from the USPHS, national Institute for General Medical Sciences, and in part by the National Science Foundation, MRL Program, under Grant No. DMR-8216718.

Making use of the *NBS Crystal Data Identification File* (1982), Mighell, Himes & Rodgers (1983) calculated the

0108-7673/85/020203-02$01.50