Acta Cryst. (1979). A35, 146–157

On the Treatment of Protein Data Measured on the Oscillation Camera

BY KEITH WILSON AND DAVID YEATES*

Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford, England

(*Received* 25 *April* 1978; *accepted* 26 *July* 1978)

Abstract

The method used in Oxford for the collection and reduction of protein data from the Arndt–Wonacott oscillation camera [Arndt, Champness, Phizackerley & Wonacott (1973). *J. Appl. Cryst.* **6**, 457–463] is described. The photographs are digitized with an Optronics Photoscan P-1000 drum scanner operated in an off-line mode, and the magnetic tapes are subsequently processed on an ICL 1906A computer. Results are presented for the protein phosphorylase b. The estimation of standard deviations for both fully and partially recorded observations is discussed, and the intensities for these two categories of term are compared.

1. Introduction

The oscillation camera has re-emerged as a powerful tool in the measurement of X-ray intensities. The revival of the instrument has been largely because of its efficiency in recording the large number of data required in protein structure determination, and has been entirely dependent upon the development of high-speed microdensitometers for the automatic processing of the photographs. The advantages of the method are discussed by Milledge (1966) and Arndt (1968), and a recent collection of papers (Arndt & Wonacott, 1977) provides an excellent summary of the current state of the art.

A number of protein structures have now been determined using this method, *inter alia* tyrosyl t-RNA synthetase (Irwin, Nyborg, Reid & Blow, 1976), glyceraldehyde 3-phosphate dehydrogenase (Biesecker, Harris, Thierry, Walker & Wonacott, 1977), pyruvate kinase (Levine, Muirhead, Stammers & Stuart, 1977) and phosphorylase b (Johnson, Weber, Wild, Wilson & Yeates, 1978). These studies give the impression that the quality of the data is at least comparable with that obtained using a diffractometer. We consider it timely to describe in detail the steps involved in the processing of data which have led to a successful structure determination of phosphorylase b using the Arndt–Wonacott oscillation camera (Arndt *et al.*, 1973). We discuss the current state of the system, attempt to point out those weaknesses of which we are aware and propose remedies for some of these weaknesses. We hope this will lead to a constructive and critical comparison with systems developed in other laboratories and to the development of an improved system in the future.

The methods for estimation of standard deviations, the integration and utilization of all data, and the consideration of the partially recorded terms differ from the work of previous authors.

2. Data collection

Data from the protein phosphorylase b (Johnson *et al.*, 1978) provide the results quoted below. The crystalline form of the protein is summarized in Table 1 and the conditions for data collection are in Table 2.

Table 1. *The crystalline form of phosphorylase b*

Space group	$P4_32_12$
Crystalline habit	Tetragonal prisms elongated along **c**
Cell dimensions	$a = 128{\cdot}5$ Å, $c = 115{\cdot}9$ Å
Unit-cell volume	$1{\cdot}94 \times 10^6$ Å^3
Daltons of protein in the unit cell	800 000
Average size of crystals used in the collection of data	$1{\cdot}5 \times 0{\cdot}8 \times 0{\cdot}8$ mm

Table 2. *Recording conditions for the native data from tetragonal phosphorylase b* (*see Table* 1)

Camera	Arndt–Wonacott screenless oscillation camera (Enraf–Nonius, Delft)
X-ray source	Elliot GX3 rotating anode operated at 1·6 kW
Radiation	Cu $K\alpha$, $\lambda = 1{\cdot}54182$ Å, Ni filter
Collimation	Standard, 0·3 mm beam at crystal
Detector	Ilford Industrial *G* X-ray film; Kodak Kodirex X-ray film
Film cassettes	Flat plate, maximum radius for data 60 mm
Crystal-to-film distance	96·0 mm
Maximum resolution	2·86 Å
Total φ for data set	45° (this gives four equivalent observations of each unique term)
$\Delta\varphi$ for each exposure	1·25°
Number of films per data set	50 two-film packs
Exposure time for each photograph	7500 s deg^{-1}; 2·6 h (10 oscillations) per exposure
Total time for 80 000 reflections	125 h

* Present address: Department of Social and Community Medicine, 8 Keble Road, Oxford, England.

0567-7394/79/010146-12$01.00

The exposure time per photograph was chosen to be the maximum in order to obtain sufficient blackening of the film for weak reflections, but consistent with providing a reasonable number of oscillation photographs per crystal. Irradiation for 24 h gave a loss of intensity through radiation damage of less than 25% for most crystals. This allowed the collection of some nine oscillation photographs per crystal with data corresponding to spacings greater than 3 Å.

We use a semi-empirical method to estimate the appropriate oscillation angle for each photograph. Recorded reflections fall into one of three categories: (i) fully recorded; (ii) partially recorded at the start or end of the oscillation range; (iii) overlapping terms. For the most efficient collection of data we wish to maximize the number of fully recorded terms without introducing a significant number of overlapping reflections. It is not possible to estimate reliably the intensities of the latter and they are lost from the data. For data to a particular maximum Bragg angle, crystal-to-film distance, reciprocal-cell dimensions, mosaic spread and crystal orientation the numbers of the terms in the three categories are calculated for a number of oscillation angles around the expected optimum. (The coordinate generation procedure is described in § 4.)

The density of the projection of the reciprocal net onto the film will vary with the value of the oscillation angle, φ, and care must be taken to consider the most densely populated range of φ. Results for data extending to 3 Å for phosphorylase b are shown in Fig. 1.

The crystals used were consistently larger than the beam in all three dimensions. We believe that the systematic errors thus introduced are more than compensated by the reduction in random error in using large crystals. The problem is discussed further in § 11.

Crystals are aligned on the camera with a major direct axis coincident with the oscillation axis by recording 'still' photographs for two major zones. For convenience the angle of rotation, φ, is defined as zero when one of the major zones is perpendicular to the X-ray beam.

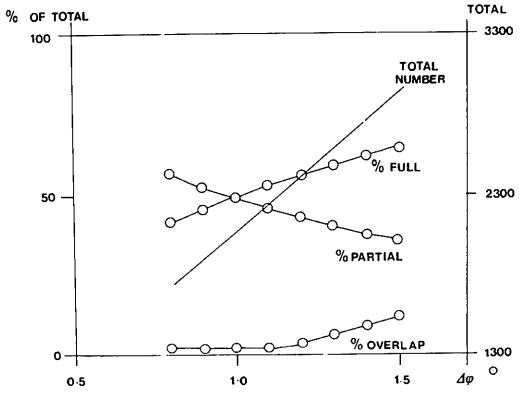

Fig. 1. The percentage and number of fully recorded, partially recorded and overlapping reflections for phosphorylase b at 3 Å assuming the mosaic spread to be 0·35″.

Photographs are now taken which will allow accurate determination of the crystal orientation (§ 4). We have used two methods for this purpose. In the first a small-angle oscillation photograph (usually 1°) is recorded about $\varphi = 0{\cdot}0°$, *i.e.* $-0{\cdot}5$ to $+0{\cdot}5°$, and a second at $\varphi = 90{\cdot}0°$. Each oscillation may be split up into two films, *e.g.* for the first range $-0{\cdot}5$ to $0{\cdot}0°$ and $0{\cdot}0$ to $+0{\cdot}5°$: this simplifies the estimation of the degree of partial recording. In the second method we record 'still' photographs at $\varphi = 0°$ and $\varphi = 90°$. The results of the two methods are compared in subsequent sections.

Three fiducial spots are marked on each orientation (and data) film. We have encountered some problems at this point and use the following procedure.

(1) We ensure that the film cassettes are aligned as accurately as possible on the camera so that the fiducial spots are in the expected positions with respect to the main beam and the oscillation axis by mechanical adjustment of the cassette supports.

(2) The direct beam is recorded on each exposure to allow estimation of the mis-setting of the cassette, and hence fiducial-spot, positions.

(3) Films are packed tightly into the cassettes. With two-film packs of Industrial *G* the films may move within the cassettes between the recording of data and the marking of fiducial spots, which is carried out on a separate instrument.*

Residual errors in (1) are largely overcome by (2). (3) however prevents accurate measurement of the position of partially recorded reflections relative to the origin for the determination of crystal orientation.

Data photographs are recorded with sequential, non-overlapping, oscillation ranges, which allows summation of partially recorded terms from the end of one exposure with those from the start of the next. This gives the maximum number of data per crystal. The first photograph is repeated at the end of data collection for estimation of a radiation-damage correction.

Data have been recorded on both Ilford Industrial *G* and Kodak Kodirex X-ray film. We have found no difference in the quality of data recorded and note that Kodirex is a faster film.

3. The optical microdensitometer

Data films are digitized using an Optronics Photoscan P-1000 drum scanner, operated in an off-line mode. The binary image of the photograph is written directly

* The referee has indicated that a slight modification to the cassette (the fitting of stop screws through the rim of the cover) prevents movement of the cover with respect to the cassette body and thus of the film within the cassette.

onto a magnetic tape. The integration and reduction of the data are subsequently carried out on an ICL 1906A computer.

The reflections on the photographs are roughly 1 mm in diameter. The films are scanned on a 100 μm grid. The optical-density values in the range 0·0–2·0 optical-density units are output as integers in the range 0–255.

The combined non-linear response of the Photoscan and the X-ray film is calibrated periodically. The latter is the dominant factor in the non-linearity (Nockolds & Kretsinger, 1970). We have recorded the direct beam with a tube collimator for a series of short exposures to produce a scale for estimation of non-linearity and calculate a point-by-point correction curve for the optical densities to be applied during the integration procedure (Fig. 2).

Microdensitometry is reviewed by A. J. Wonacott and R. M. Burnett in Arndt & Wonacott (1977, pp. 119–138).

4. Orientation of a crystal and prediction of reflections

This section of the data processing is the most important since the determination of reliable values for the integrated intensities depends upon knowing precise values for the unit-cell parameters, and the crystal orientation. We need to know where a particular diffraction maximum will occur, and if the corresponding reciprocal-lattice 'point' has passed completely through the Ewald sphere. Small changes in a given orientation or in reciprocal-lattice parameters will not affect the overall pattern on a particular photograph but will affect the distribution of partially recorded reflections, and their degree of recording. Thus identification of a set of partially recorded reflections affords a way of determining the crystal mis-setting, and variations in the unit-cell parameters, provided the degree of recording can be estimated with some certainty.

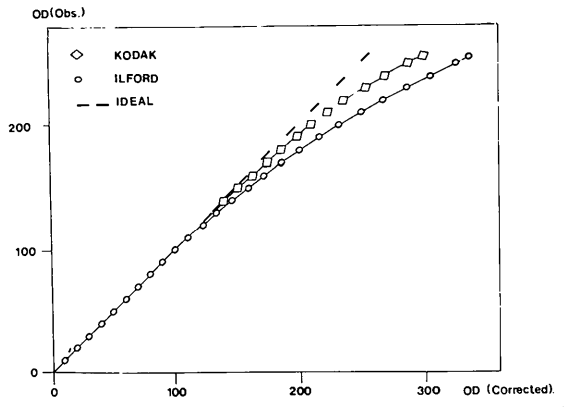

Fig. 2. Point-by-point correction curve for the optical-density values output by the Optronics Photoscan densitometer using Kodak Kodirex and Ilford Industrial *G* X-ray film.

Best estimates for the parameters are obtained by an iterative least-squares minimization process on,

$$(D_{obs} - D_{calc})^2, \tag{4.1}$$

where

$$D_{obs} = 1 + g.e\,(\mathbf{H})$$
$$D_{calc} = (1 + 2X_L + X_L^2 + Y_L^2 + Z_L^2)^{1/2} \tag{4.2}$$

are the observed and calculated distances of reciprocal-lattice point $P(\mathbf{H})$ from the centre of the Ewald sphere; g is a factor depending on the observed degree of recording; $e(\mathbf{H})$ is the radius of a spherical approximation to the reflecting domain $c(\mathbf{H})$ given by

$$e(\mathbf{H}) = \frac{\gamma}{2}.\sigma, \tag{4.3}$$

where σ is the distance of $P(\mathbf{H})$ from the origin of reciprocal space, and γ is the reflecting range of the crystal and depends upon the collimation conditions and mosaic spread. (X_L,Y_L,Z_L) are the coordinates of $P(\mathbf{H})$ in a reference frame fixed in the camera (Fig. 3), given as,

$$\mathbf{X}_L = [\varphi].[S].[W].\mathbf{H}^T, \tag{4.4}$$

where $[\varphi]$ describes the rotation relative to some datum, $[S]$ describes the mis-setting of the crystal by three orthogonal angles δX, δY, δZ, and $[W]$ transforms from the crystallographic to an orthogonal coordinate system.

The observed partially recorded reflections are normally obtained from 'setting' photographs. These may be either small-angle oscillation photographs or 'still' photographs, taken at $\varphi = 0°$ and $\varphi = 90°$ to reduce correlation between the parameters, but in addition can include observations from data photographs. For success in the procedure we find that it is important to have good initial estimates for the mis-setting angles, and that for small-angle oscillation photographs the observed degree of recording is estimated with care. When still photographs are used we find it is satisfactory to assume that all reflections

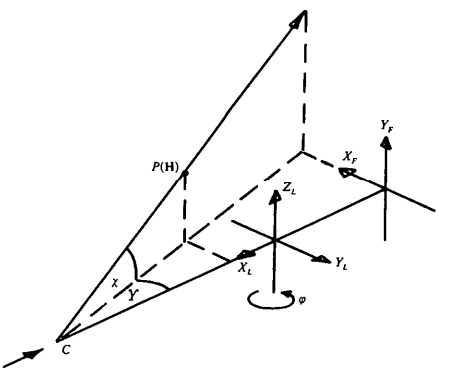

Fig. 3. The reference frame for the coordinates of a reflection on the oscillation camera.

are 50% partially recorded. Experience has shown that a satisfactorily accurate matrix is more easily defined when using 'still' photographs.

For a given angular range φ_i to φ_f the film coordinates (X_F,Y_F) for reflections expected to appear on the corresponding photograph are generated as,

$$(X_F,Y_F) = \mathscr{F}(Y,\chi,S)_{X,Y} \qquad (4.5)$$

where S is the crystal-to-film distance and the form of $\mathscr{F}$ depends upon the geometry of the film detector. Each reflection occurring is flagged as to whether it is fully recorded, partially recorded at φ_i or φ_f, or ignored due to overlapping. A plot of each predicted film is available for comparison with the corresponding data photograph, as described in Arndt & Wonacott (1977, p. 90). We find this particularly useful for checking the orientation procedure by comparing individual reflections. Any errors are best detected here since they will not otherwise be apparent until the agreement between equivalent reflections is analysed later (§§ 6 and 8). The positions of the fiducial marks recorded on the data photographs can also be checked since these are occasionally displaced due to film slippage in the cassette.

The procedures for crystal orientation used in three other laboratories are discussed by J. Nyborg and A. J. Wonacott in Arndt & Wonacott (1977, pp. 139–151) and by A. Jones, K. Bartels and P. Schwager in Arndt & Wonacott (1977, pp. 105–117).

5. Integration

The fiducial marks on each photograph are used to generate an initial approximation to the transformation from the predicted film coordinates to coordinates in the scanner reference frame. Using this to define approximate positions for a set of fully recorded reflections, their centroids are used to determine a precise transformation:

$$(X_0,Y_0) = \zeta(X_F,Y_F). \qquad (5.1)$$

The program automatically defines a mean density level for the photograph and then selects a specified number of terms, preferably close to the maximum Bragg angle, above this level for the determination. The centroid of a diffraction maximum is calculated as

$$X = \frac{\sum_i X_i \bar{O}_i}{\sum_i \bar{O}_i}; \quad Y = \frac{\sum_i Y_i \bar{O}_i}{\sum_i \bar{O}_i}, \qquad (5.2)$$

where $\bar{O}_i$ is the optical density for point (X_i,Y_i) above a bias level.

The integrated intensity of a reflection is determined as

$$I(\mathbf{H}) = \sum_{i=1}^{N_p} O'_i - k \sum_{j=1}^{4} \sum_{k=1}^{N_b} O'_{jk}, \qquad (5.3)$$

where O'_i and O'_{jk} are contributions from the peak and background sampling areas N_p and N_b respectively. The integration box we favour is shown in Fig. 4. O' denotes an optical density value O corrected for non-linearity in response of the film/scanner system. The size of the diffraction maxima on the film will vary for different values of (Y,χ) and it is necessary to vary the size of the integration box to accommodate this effect.

This oblique-incidence factor can be accounted for by writing the variable box parameter as

$$n = n_0/f(Y,\chi). \qquad (5.4)$$

The background sampling areas are kept constant, and the background-to-peak scale factor, k, varied as

$$k(Y,\chi) = \frac{N_p(Y,\chi)}{4N_b}. \qquad (5.5)$$

The angular correction function $f(Y,\chi)$ depends on the geometry of the film and for a flat-plate cassette is given by

$$f(Y,\chi) = \cos Y \cos \chi. \qquad (5.6)$$

Reflections are rejected if they exceed the saturation level on the film, or if any of the four background values differ significantly from their mean. All reflections having a negative $I(\mathbf{H})$ and not rejected as a background failure are retained. Their proper treatment has been investigated elsewhere (French & Wilson, 1978).

6. Inter-film scale determination for packs of two films

The program *FILMPACK* performs three operations: initial estimation of the standard deviations within each photograph, calculation of the film-to-film scale factor, and application of the scale factor and other corrections before merging equivalent terms from the two films. The program accepts data from precession or

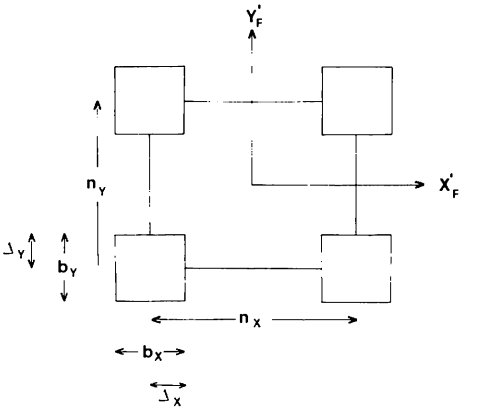

Fig. 4. The box shape for the integration of the terms (n = reflection, b = background).

oscillation photographs. Only the latter will be described here.

First, all equivalent terms from both photographs are sorted together. The standard deviations of the intensities are estimated independently for the first and second film in the pack from the expression:

$$\sigma_i = \left[\sum_{j=1}^{N} \frac{(I_i - \bar{I})^2}{N-1} \right]^{1/2}, \qquad (6.1)$$

where I_i is the observed intensity of the ith member of a set of N fully recorded equivalent terms, $\bar{I}$ is the unweighted mean intensity for this set, σ_i is the estimated standard deviation for the ith member of the set. The estimates for each set of equivalents are summed in ten ranges of the mean intensity for the set. We determine the constants A and B for the empirical expression (6.2*a*) by minimizing the sum of the squares of the discrepancies between the left- and right-hand sides of the equations for the ten ranges:

$$\sigma_{\text{est}} = A + BI, \qquad (6.2a)$$

where σ_{est} is the estimated standard deviation, I is the mean intensity of the set and A and B are constants to be determined. This expression was used as a totally empirical observation of the variation of σ with I, and gives apparently good agreement with the observed errors (Fig. 5).

Wilson (1977) has suggested that it would be more appropriate to estimate the variance rather than the standard deviation from an expression of this form and we have investigated the modification of the equation to

$$\sigma^2_{\text{est}} = A + BI. \qquad (6.2b)$$

We have not obtained a satisfactory concordance between the observed and estimated variance using (6.2*b*) (Archibald & Wilson, 1978), and at present still use (6.2*a*). All results described in the paper refer to the use of (6.2*a*). We are continuing our study of the possible use of (6.2*b*) or a related expression as it would resolve the problems discussed in § 9.

Expression (6.2*a*) provides an estimated standard deviation for each individual reflection from the mean observed intensity of the set of equivalent terms. Errors have been estimated from (6.2) for both fully and partially recorded terms in the same manner.

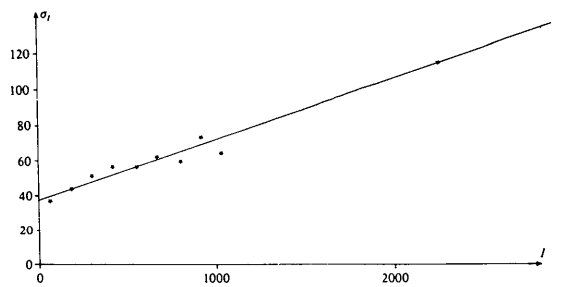

Fig. 5. The estimated standard deviation as a function of intensity. ★ Observed values (6.1); the straight line represents the best fit of (6.2*a*) to these observations.

For this empirical estimation of errors to be valid there must be a significant number of equivalent terms on each photograph. We have considered both truly equivalent and Friedel-related terms to be equivalent for this purpose, for both native proteins and isomorphous derivatives. When there are insufficient numbers of equivalent terms we assume default values of A and B derived from our experience with previous photographs.

The estimated standard deviations provide weighting factors in the determination of the film-to-film scale factor. The intensities on the first and second film (I_1 and I_2 respectively) in a flat cassette on an oscillation camera are related by:

$$I_1 = I_2 \exp(At \sec 2\theta), \qquad (6.3)$$

where At is the film constant. The factor $\exp(\sec 2\theta)$ allows for the oblique-incidence absorption by the first film (Cox & Shaw, 1930; Whittaker, 1953).

We evaluate At from:

$$At = \frac{\sum_{i=1}^{N} \left\{ \frac{1}{\sigma_r^2} \left[\log_e \left(\frac{\langle I_{1i} \rangle}{\langle I_{2i} \rangle} \right) \right] / \sec 2\theta \right\}}{N}, \qquad (6.4)$$

where $\langle I_{1i} \rangle$ and $\langle I_{2i} \rangle$ are the mean intensities for a set of equivalent terms recorded on the first and second films respectively, σ_r is the estimated standard deviation of $\log_e(\langle I_{1i} \rangle / \langle I_{2i} \rangle)$ and N is the number of sets of equivalents. Terms are only included in the summation if the mean intensity for the equivalents is greater than six standard deviations for both the first and the second film.

After determination of the film constant, At, the following operations are carried out.

(i) The intensities for the second film are corrected for the oblique-incidence absorption by the first film (equation 6.3).

(ii) All fully recorded equivalents from the first and second films are summed to give a mean weighted intensity; observations more than four standard deviations from this mean are rejected. This eliminates spurious intensities (*e.g.* particles of dust on the film). This number and the monitor of the rejected terms provide a useful diagnostic for a poorly determined crystal-orientation matrix.

(iii) The mean weighted intensity for terms with *identical* observed indices on the first and second film is calculated. Equivalent terms within each film are not averaged.

(iv) The Lorentz–polarization and oblique-incidence (Cox & Shaw, 1930; Whittaker, 1953) corrections are applied to the data.

7. Radiation damage

Protein crystals in general show appreciable deterioration of the diffraction pattern on exposure to X-rays. A number of models have been proposed for the effects of irradiation, the earliest being that of Blake & Phillips (1962). In all such models the intensity of a reflection decreases with increases in both exposure time and scattering angle.

We have used two models for the effects of irradiation: (i) that of Blake and Phillips:

$$I_0 = I_t\left\{\exp\left[\frac{-\mu_2+\mu_3}{t}\right] + \exp(-\mu_3 t)[1{\cdot}0 - \exp(-\mu_2 t)\exp(-DS)]\right\}^{-1}; \quad (7.1)$$

and (ii) a 'linear-quadratic' expression (linear in t and quadratic in S):

$$I_0 = I_t[1{\cdot}0 + t(A + BS + CS^2)], \quad (7.2)$$

where I_0 is the intensity at time zero, I_t is the intensity at time t, A,B,C are empirical constants, μ_2,μ_3,D are the physical constants for the decay process described by Blake & Phillips, and S is $(\sin\theta)^2/\lambda^2$ in (7.1) and $\sin\theta/\lambda$ in (7.2). The number of an exposure in a sequence of photographs is taken as a measure of the time t.

Expression (7.1) is based on a physical model for the decay. Expression (7.2) is a purely empirical model based on observations of the general form of the decay for protein crystals.

To assess the radiation damage we repeat the first exposure for a crystal at the end of the data collection (§ 3). If such a photograph is not available it is sometimes possible to evaluate a correction from comparison of equivalent terms recorded on different exposures in the sequence: this is only feasible for high-symmetry space groups.

The program *FILMRAD* sorts together equivalent terms from the photographs and minimizes φ:

$$\varphi = \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{1}{\sigma_j^2}\left(I_{t_j} - \frac{\bar{I}_0}{C_{t_j}}\right)^2, \quad (7.3)$$

where N is the total number of sets of equivalent terms; I_{t_j} is the intensity of a term j in a set of n equivalent terms; intensity j is recorded at time t; σ_j is the standard deviation of I_{t_j}; C_{t_j} is the current estimate of the correction to be applied to the intensity I_{t_j}; C_{t_j} is computed from (7.1) or (7.2) as appropriate; and $\bar{I}_0$ is the current estimate of the mean intensity for this set of equivalent terms corrected to time $t = 0$. The program is given initial estimates of the constants in (7.1) and (7.2) and iteratively minimizes the expression φ in (7.3) by the method of least squares. (7.3) is formulated so as to segregate the calculated contributions $(\bar{I}_0, C_{t_j})$ into the right-hand term and the observations (I_{t_j}) into the left-hand term.

Typical results are shown in Fig. 6. For many crystals both models gave a satisfactory fit to the observed decay [curve (*a*)]: by default we use the Blake–Phillips expression as this is based on a theoretical model. For more than 25% of the crystals we were unable to obtain a satisfactory fit to the observed decay: indeed for several crystals the variation of intensity with time showed an increase rather than the expected decrease [curve (*b*)]. For such crystals we presume that the variation of intensity with time and scattering angle includes significant contributions from phenomena other than radiation damage.

In summary such a method for the estimation or radiation damage is not totally satisfactory. We discuss the problem further in § 8.

8. Reduction to the unique data

The program *SSM* (scale–sort–merge) determines inter-film scales and averages the equivalent terms to give a unique set of data. The operation of the program, especially where it relates to photographic data, is described in the following subsections. Two problems which have arisen in the reduction will be described in §§ 9 and 10.

8.1. *Inter-film scale determination*

Equivalent reflections from the various films are sorted together, retaining the original indices, the unique indices, the intensity and standard deviation, and the number of the film. Scale factors between the films are determined by the method of Fox & Holmes (1966). A significant reduction in the number of

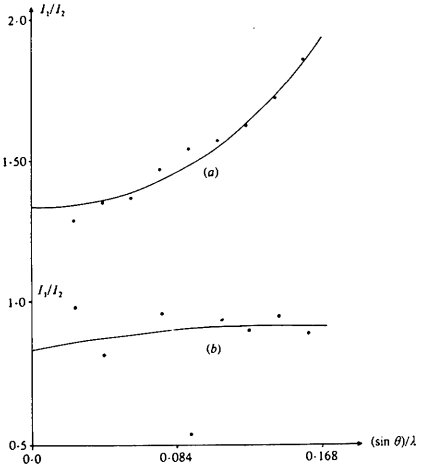

Fig. 6. The radiation-damage correction estimated from (7.1). Solid curve: best fit of (7.1). Curve (*a*) represents a most satisfactory fit of model to observations, (★ experimental observations), and curve (*b*) a most unsatisfactory agreement (● experimental observations).

iterations required is achieved by estimation of the initial scales from the mean intensity on each photograph. The scale determination is iterated until the shifts in the scale factors are less than 0·02%, which has generally required 10–15 cycles for a set of fifty photographs. All data are included in the scale determination and the equations of Fox & Holmes are weighted by the inverse of the variance of the observations estimated as in § 6.

We originally expected that the photographs from one crystal would at this point require roughly the same scale factor if a satisfactory radiation-damage correction had been applied and scales which increased with film number for those for which a correction had not been previously applied. This is not consistently so. There is often a significant random variation in scales for the sequential exposures from a crystal. For photographs corrected for radiation damage (§ 7) a residual trend in the scale factors may be apparent. For crystals where no correction was applied the trend may be in the opposite direction to that expected: that is, the scale decreases with the number of the exposures in the sequence.

These observations are related to the problems in satisfactory determination of a radiation-damage correction described in § 7, and perhaps to the lack of an absorption correction. We conclude that the variation of the intensity of the diffraction pattern with time is a complex phenomenon. For some crystals the change obeys the decay law proposed by Blake & Phillips reasonably well, and a correction based on their model gives a satisfactory set of inter-film scales. The empirical 'linear-quadratic' correction (7.2) often produces equally satisfactory results, however, and it is probable that the detailed form of the correction is not vital, provided it includes approximately the right dependence upon time and scattering angle.

The explanation of the complex variation lies in the other phenomena which may contribute to the change in intensity with time. Such phenomena may include, *inter alia*, the following.

(1) Evaporation of liquid from, or distillation of liquid to, the immediate surroundings of the crystal during collection of data for proteins where the crystals are sealed in glass capillary tubes containing some liquid: the change in the volume of liquid will affect the absorption properties of the system; evaporation/distillation may also cause swelling/shrinkage of the unit cell with consequent intensity changes mainly in the low-angle reflections.

(2) Fluctuation of the intensity of the X-ray source: the data were collected on an Elliot GX3 rotating anode and while we have no positive evidence for fluctuation it cannot be entirely excluded: this may produce both random and/or systematic variations in the scales for a sequence of exposures.

(3) Variation in the thickness of the photographic emulsion (unlikely) and/or in the developing conditions for sequential films: this should introduce primarily random errors.

All such effects will make realistic modelling of intensity change with time by equations such as (7.1) or (7.2) difficult, if not impossible. Other laboratories (P. R. Evans, private communication; R. Wierenga, private communication) have reported the determination of both relative scales and temperature factors by an algorithm analogous to that of Fox & Holmes. While we have no experience with such a scheme at present, it would seem to provide a useful solution to the problems described above. The relative temperature factor between films would hopefully take up residual radiation-damage effects, differences in absorption dependent upon scattering angle and any other relative variations of intensity with scattering angle.

8.2. *Estimated standard deviations*

The inter-film scale factors are applied to the observations and the partially recorded fractions are summed to give complete intensities (see § 8.3). The intensities are analysed to provide improved estimates of the standard deviations by the following method.

For all sets of equivalent terms, including those from both fully, and summations of partially, recorded observations, we tabulate the distribution of the following expression in ranges of the mean intensity:

$$\frac{F_i^2 - \bar{F}^2}{\sigma_F}, \tag{8.1}$$

where F_i^2 is the square of the modulus of the ith of a set of equivalent structure factors, $\bar{F}^2$ is the mean-square modulus for the set, and σ_F is the standard deviation of F_i^2. If σ_F^2 is a true estimate of the variance then the distribution should be invariant with $\bar{F}^2$ and for each range should itself have a mean of zero and a standard deviation of unity.

We modify the standard deviation according to the expression (Dodson, 1976)

$$\sigma_{\text{new}} = A(\sigma_{\text{old}}^2 + B^2 \bar{F}^2)^{1/2}, \tag{8.2}$$

where σ_{new} is the modified standard deviation, σ_{old} is the original value from (6.2), and A and B are constants input by the user. The constants A and B are empirically varied by manual intervention until the distribution approximately possesses the expected properties. An example is provided in Fig. 7.

8.3. *Output of the unique data*

The mean weighted intensity and its standard deviation are computed for the unique terms: data from different crystals and different films are all averaged for the individual data sets. Negative and positive mean intensities are all accepted as meaningful observations and output by the program. The treatment of such

intensity distributions is described elsewhere (French & Wilson, 1978).

9. Partially recorded terms: standard deviations

The following symbols will be used throughout this section:

e.s.d.	estimated standard deviation;
P	the fraction to which a reflection is partially recorded;
$1-P$	the fraction of partial recording of the same term on an adjacent exposure;
I_F	the intensity of a fully recorded term;
I_P, I_{1-P}	the intensities of terms recorded by fractions P, $1-P$;
σ_F	the e.s.d. of I_F;
σ_P	the e.s.d. of I_P;
σ_S	the e.s.d. of $I_P + I_{1-P}$.

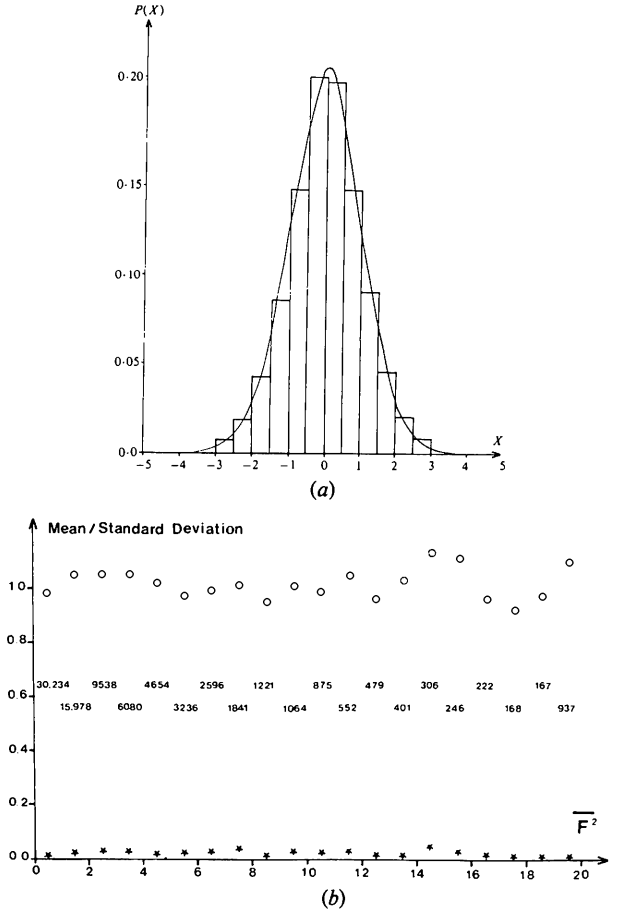

Fig. 7. (*a*) A plot of $(F^2 - \bar{F}^2)/\sigma_{F^2}$ for the native phosphorylase b data after application of (8.2). The histogram represents the observed data, the curve a theoretical Gaussian distribution of errors. (*b*) The mean (★) and standard deviation (○) of the distribution of (8.2) as a function of $\bar{F}^2$ for the same data. Also shown is the number of terms in each range.

Table 3. *The standard deviations estimated from* (6.2*a*) *for fully and partially recorded terms on the same photographs*

The standard errors are given by: $\sigma = 29{\cdot}45 + 0{\cdot}0142I$, and the maximum intensity is 7204. σ_{P50} represents the standard deviation of the sum of the two halves for a term partially recorded by 50% and so forth (*i.e.* σ_S in text).

I	σ_F	σ_{P50}	σ_{P25}	σ_{P10}
0	29·45	41·65	41·65	41·65
100	30·87	42·65	42·66	42·66
1000	43·65	51·69	51·93	52·31
5000	100·65	91·85	95·22	100·25
7500	135·95	116·96	122·87	131·56

During integration the same 'box size' is used for the partially and fully recorded terms (§ 5), and in the estimation of the e.s.d.'s (§ 6) the e.s.d. for both classes is computed from the same empirical expression (6.2) in the present program system. Those reflections (near the oscillation axis) which are split between more than two films are rejected from the data.

The problem discussed in this section is dependent upon this method of integration. A practical means of avoiding the problem would be to modify the box for each partially recorded term during integration and include in the summation only those optical-density points which correspond to the partial reflection. We reject this approach as we consider it to be fraught with the danger of introducing systematic error into the intensities because it is critically dependent upon the estimated shape and extent of each partially recorded term. (It also poses a relatively large computing problem.)

We now return to a discussion of the inadequacy of the method used and describe a valid alternative.

Consider two sequential exposures with identical values of the constants A and B in (6.2). σ_F on either exposure is given by:

$$\sigma_F = A + BI_F \tag{9.1}$$

and σ_P on the first exposure and σ_{1-P} on the second by:

$$\sigma_P = A + BI_F P \tag{9.2}$$

$$\sigma_{1-P} = A + BI_F(1-P). \tag{9.3}$$

On summing the two fractionally recorded parts to give the total intensity, and then by the conventional summation of variance we obtain:

$$\sigma_S = \{(A + BI_F P)^2 + [A + BI_F(1-P)]^2\}^{1/2}. \tag{9.4}$$

In Table 3 we compare σ_F and σ_S for typical values of A, B and intensity for data from phosphorylase b. For the smallest terms the e.s.d.'s have the expected values: for an intensity of zero σ_S is $\sqrt{2}\sigma_F$. As the intensity increases σ_F increases more rapidly than σ_S, as expected. However, for the top 20% of the data σ_S is actually less than σ_F: this is not the expected result. As

the intensity increases σ_F and σ_S should converge, becoming equal for an infinitely large intensity when the 'empty' background region of the fractionally recorded terms contributes a negligible proportion of the error.

A solution to the problem is the estimation of the fraction of partial recording for each term and the output of this value with the intensity during integration. Then for the fraction of the integration box which corresponds to the reflection we have

$$\sigma_{P_1} = (A + BI_P/P)\,P^{1/2} \tag{9.5}$$

and for the remainder of this box – the empty region:

$$\sigma_{P_2} = (A)(1-P)^{1/2}. \tag{9.6}$$

For the rest of the same partially recorded by $1-P$ on the adjacent exposure we have:

$$\sigma_{(1-P)_1} = [A + BI_{(1-P)}/(1-P)](1-P)^{1/2} \tag{9.7}$$

$$\sigma_{(1-P)_2} = [A](P)^{1/2}, \tag{9.8}$$

which gives an e.s.d. for the sum of the fractional intensities:

$$\sigma_S = \left[\left(A + \frac{BI_P}{P}\right)^2 P + \left(A + \frac{BI_{(1-P)}}{1-P}\right)^2 (1-P) + A^2 P + A^2(1-P)\right]^{1/2}. \tag{9.9}$$

Therefore

$$\sigma_S = [(A + BI_F)^2 + A^2]^{1/2}. \tag{9.10}$$

Expression (9.10) shows the effect of the present method of integration on the e.s.d. For partially recorded spots with the same box size as a fully recorded term the variance is increased by the term A^2 in (9.10), that is by the error inherent in the empty box.

Expressions (9.5)–(9.9) illustrate the means of implementation for the method. In the computation of σ_P and σ_{1-P} we utilize the estimate of P as well as the intensity I_P.

An alternative solution to the problem is given by the use of expression (6.2*b*) for the estimation of variance rather than (6.2*a*) to give standard deviation. We then find that:

$$\sigma_F^2 = A + BI$$

$$\sigma_S^2 = 2A + BI$$

and, as should be the case, the variance for a partially recorded reflection of zero intensity is twice that of a fully recorded term, the variances converging on identical values as I tends to infinity. Further studies on this problem will be published elsewhere (Archibald & Wilson, 1978).

10. Partially recorded terms: intensities

We have compared the square of the moduli, F^2, for fully recorded terms with those for the sum of fractionally recorded terms. The latter class are systematically larger than the former, and the percentage discrepancy is approximately invariant with F^2 (Table 4). A similar phenomenon has been noted

Table 4. *Comparison of intensity of fully and partially recorded terms as a function of the mean intensity for the data set for phosphorylase b with glucose* 1*-phosphate*

The 3 Å information includes the 6 Å terms. The data are divided into range of $F^2 = 250$. The overestimation is considerably less at 6 Å resolution. F_F^2 is the intensity of a fully recorded term, F_P^2 that of a partially recorded term, Δ the mean value of $F_P^2 - F_F^2$, and $\%\Delta$ the percentage value of Δ, *i.e.* $(F_P^2 - F_F^2)/\bar{F}^2$, where $\bar{F}^2$ is the mean of the fully and partially recorded terms.

	3 Å data					6 Å data				
	NREF	F_F^2	F_P^2	Δ	$\%\Delta$	NREF	F_F^2	F_P^2	Δ	$\%\Delta$
1	2364	106	104	−2	−1·9	303	111	107	−4	−3·6
2	1536	349	379	30	8·3	221	370	379	9	2·4
3	1070	598	638	40	6·5	158	626	632	6	1·0
4	767	843	896	53	6·1	127	868	891	23	2·6
5	610	1093	1142	49	4·4	105	1106	1122	16	1·4
6	473	1328	1408	80	5·8	74	1358	1380	22	1·6
7	338	1592	1662	70	4·3	68	1623	1638	15	0·9
8	260	1831	1899	66	3·5	47	1858	1872	14	0·8
9	234	2074	2180	106	5·0	47	2120	2120	0	0·0
10	175	2323	2421	98	4·1	43	2366	2373	7	0·3
11	120	2549	2679	130	5·0	35	2601	2653	52	2·0
12	97	2803	2926	123	4·3	14	2882	2923	41	1·4
13	79	3031	3211	180	5·8	20	3134	3075	−59	−1·9
14	49	3310	3411	101	3·0	13	3355	3314	−41	−1·3
15	39	3541	3703	172	4·7	11	3649	3642	−7	−0·2
16	42	3840	3939	99	2·5	9	3942	3799	−143	−3·7
17	43	4073	4188	115	2·8	16	4110	4161	51	1·2
18	34	4240	4509	269	6·1	11	4384	4358	−26	−0·6
19	33	4550	4692	142	3·1	4	4500	4702	202	4·4
20	153	6298	6578	280	4·3	47	6594	6594	0	0·0

previously by Irwin *et al.* (1976), whose analysis was based on variation with F rather than F^2. The comparison of squares rather than moduli provides a more rational analysis as the squared moduli are directly related to the observations (intensities). This partly explains the difference between our results and those of Irwin *et al.* who observed that the percentage overestimation decreased with the modulus, F. Conversion of the information in Table 4 to F's rather than F^2 produces a similar decrease.

A second difference between the present results and those of Irwin *et al.* is the relatively large discrepancy for very small moduli in the latter. No such tendency is apparent in our data and we explain this by the inclusion of all data in the set of observations, however small the observed intensity: no cut-off is applied for low intensities and the introduction of a sampling bias is avoided.

Three possible causes of the discrepancy phenomenon are as follows.

(1) Inaccuracy in the determination of the crystal orientation matrix. For terms flagged as fully recorded but which are in reality partially recorded this will lead to an underestimate of the 'fully' recorded terms, and hence an apparent overestimate of the partially recorded ones.

(2) Underestimation of the crystal reflecting range, γ, in the prediction of those terms which reflect in a given oscillation range. The effect of this is similar to that of (1).

(3) The use of Ni-filtered Cu $K\alpha$ rather than monochromatic radiation, which gives a visibly higher background within as compared to without the lunes of data.

Effect (3) is probably not important. In Fig. 8 we show the integration box for a term which is fractionally recorded by 50%, that is an average partially recorded term. Half the box size for the peak corresponds to a region where there is no reflection. This 'no-reflection' half is outside the lune, as are two of the background boxes. The 'reflection' half of the term lies within the lune as do the remaining two backgrounds. Taken together the four backgrounds are expected to provide a reasonable approximation to the true background level for such an average term. They may not provide the ideal value for each individual partially recorded term, but should not systematically affect the integrated intensities if a statistically significant number of terms is considered.

In Table 4 the discrepancy between the fully and the sum of partially recorded terms is given for the same set of integrated phosphorylase photographs to limiting resolutions of 3 and 6 Å. The discrepancy is considerably smaller for the data at low Bragg angle. This is entirely consistent with explanations (1) and (2) above as both are likely to affect high-angle more than low-angle reflections – particularly for (1) if the estimate of the crystal-to-film distance is in error. It is not simple to separate these two causes. To avoid (2), the value of γ must be increased during orientation and integration. To obtain an improved crystal-orientation matrix it is necessary, *inter alia*, to include data close to the maximum Bragg angle in the definition of the matrix.

11. Absorption

Three approaches suggest themselves for the treatment of absorption of X-rays by the crystal and its immediate environment: (i) no absorption correction; (ii) experimental absorption curves for crystals which are larger than the beam in each dimension (Schwager, Bartels & Huber, 1973); and (iii) conventional experimental (or theoretical) corrections for crystals which are small enough for the crystal to see all of the X-ray source all of the time. We have not applied an absorption correction, as discussed below. For extension of the data to 2 Å we intend to use method (ii), for which an excellent discussion is provided by K. Bartels in Arndt & Wonacott (1977).

For phosphorylase b data to a resolution of 3·0 Å we have consistently used crystals which are larger than the beam in each dimension and have not applied an absorption correction. Hence the random errors of observation in the structure-factor moduli have been reduced at the expense of the introduction of some systematic error. We justify this approach on two grounds.

Firstly the morphology of the crystals is reasonably constant and the absorption might be expected to be similar for all such crystals. Thus all the data should be in error in roughly the same systematic manner.

Secondly the technique of multiple isomorphous replacement (MIR) has been used to provide an experimental solution to the phase problem. In MIR the data for the isomorphous derivatives are scaled to the structure-factor moduli of the native protein before

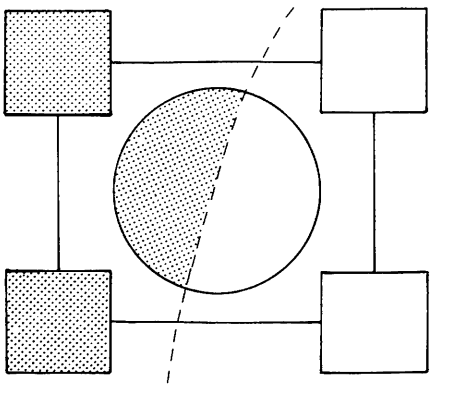

Fig. 8. The integration box for a term partially recorded by 50%. The dashed line represents the edge of the 'lune'. The dotted half of the term is recorded and lies inside the lune as do the two dotted backgrounds. The remaining two backgrounds and the 'unrecorded' half of the spot lie outside the lune.

calculations of the phases. In this laboratory, as in several others, the data are scaled in blocks of reciprocal space to maximize the significance of the isomorphous differences throughout the set of data. Matthews & Czerwinski (1975) describe a typical procedure.

This 'local' scaling should effectively minimize systematic differences between the parent and derivative moduli, whence the random errors become of prime importance. This approach will not affect the systematic errors in the parent moduli which will be important if, for example, these are to be compared with calculated terms.

In Fig. 9 we present scale factors F^2_{PH}/F^2_P, for a representative phosphorylase derivative before and after 'local' scaling. That the scale factors are restricted to the range 0·8–1·2 supports the assumption that the systematic errors are not too severe. The form of variation of scale factor with φ, ζ^2 and l is what might be expected for data collected on crystals which exceed the size of the beam and no absorption correction has been applied. A direct relationship between absorption and scale factor is complex as the unique data for which the scale factors are determined are a merged series of observations from several crystals and each term contains contributions from a variety of regions of the sphere of reflections. The improvement in the scale factors after local scaling is evident in Fig. 9.

12. Discussion

A method of treatment for data from the oscillation camera has been described and results have been quoted for the application to phosphorylase b. The method has led to a satisfactory electron density synthesis for phosphorylase b at 3 Å resolution (Johnson *et al.*, 1978).

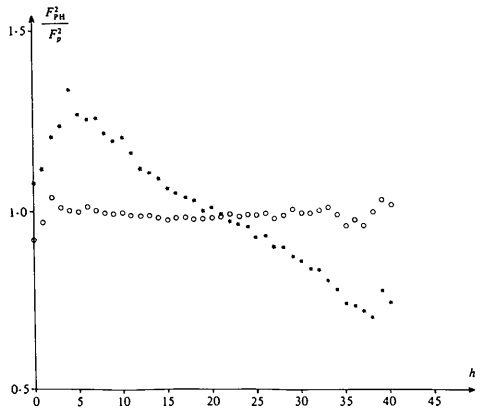

Fig. 9. Scale factors F^2_{PH}/F^2_P before (★) and after (○) local scaling for the maltotriose–phosphorylase data. The scale as a function of index h is presented as it showed the largest variation.

Table 5. *Summary of the data sets collected for phosphorylase b*

The numbers in parentheses refer to the discrepancy between the mean intensity and observed intensity for one of a group of equivalent terms resulting in the rejection of that term. The merging R is defined as $R = [\sum_i(I_i - \bar{I})/\sum I_i] \times 100$ where I is the mean intensity and I_i the intensity of the ith number of a group of equivalent terms.

Data set	Total terms	Unique terms	Number rejected	Number negative	R (%)
Native	81536	19480	1328 (3)	744	10·24
EMTS	80432	36641	149 (5)	2261	9·76
PT	72638	33518	370 (4)	1329	11·80
G3	66287	18271	526 (4)	486	8·74
G1P	68460	18164	245 (4)	624	8·31
ATP	75046	19061	255 (4)	794	9·84

Table 6. *The time required for collection and processing of a 3 Å data set for phosphorylase b using the Arndt–Wonacott oscillation camera compared with that for a single-counter diffractometer*

The diffractometer times are estimated by extrapolation of experience of 6 Å resolution (Johnson *et al.*, 1974). For both methods the times quoted are optimal and ignore problems such as unexpected deterioration of crystals during irradiation.

	Camera	Diffractometer
Data collection	1 week	10 weeks
Scanning of films	24 h	–
Integration		
Real time	4 weeks	1 week
c.p.u. time	4 h	30 min
Reduction to unique data set		
Real time	1 week	1 week
c.p.u. time	2 h	2 h
Total real time	6 weeks	12 weeks

The quality of the data is summarized in Table 5. We believe they compare favourably with data collected by diffractometer techniques. The merging R factor (defined in Table 5) for data sets collected on a four-circle diffractometer to 6 Å resolution was typically 5·2% (Johnson, Madsen, Mosley & Wilson, 1974) and for the present data, 5·0%.

The times needed for the collection and processing of a 3 Å data set for phosphorylase b on the oscillation camera and a single-counter diffractometer are compared in Table 6. The minimum time involved in collection of a camera set is seven days: we only attained this minimum for three of the seven data sets collected. We estimate the time required on a single-counter diffractometer for data of comparable quality as ten weeks. Roughly 24 h are required to scan each photographic data set on the Optronics densitometer: this procedure can, however, be run more or less concurrently with data collection. Thus the oscillation camera is almost an order of magnitude faster than a single-counter diffractometer for the collection of data.

We also note that we have estimated the time for two equivalent observations of each intensity for the diffractometer (*i.e.* one Friedel pair) whilst the camera produces of necessity four equivalent terms in the measurement of the unique data set.

The subsequent processing of the data is considerably slower for the photographic system. The shortest time in which we have integrated a complete set of photographs is roughly twenty days on an ICL 1906A computer. This time is, however, largely limited by the turn-around on the multi-user machine, and the investment of personal time was considerably less than twenty days.

The combined collection and processing time is thus more than halved for the change over from diffractometer to camera. This rate is continually improving in favour of the camera as our data-reduction system is enhanced.

A rate-limiting step in our photographic system has been in the determination of the crystal-orientation matrix. Determination of the matrix from small-angle oscillation photographs has proved time consuming and furthermore the results described in § 10 suggest that the matrices obtained are not entirely satisfactory. The difficulty in determining a matrix is remedied by the use of 'still' rather than small-angle oscillation photographs. To improve the accuracy of the matrix for the high-angle data it is also necessary to include spots close to the resolution limit of the data in the determination of the matrix. This reduces the discrepancy between the fully and the sum of partially recorded intensities. Improved estimation of the crystal-orientation matrix is currently under investigation (Stura, 1978).

The computation of satisfactory radiation-damage corrections and inter-film scales are other points of weakness in our system. The estimation of both relative temperature and scale factors during the inter-film scale determination will overcome these problems. A further weakness is in the methods used to estimate errors in the data (§§ 6 and 9). The solution to this problem is again currently under investigation.

We are grateful for financial support from Imperial Chemical Industries (KW) and the Medical Research Council (DY). We are deeply indebted to Dr L. N. Johnson for many helpful discussions and to Professor D. C. Phillips for encouragement of the project. The excellent facilities and helpful cooperation provided by the Oxford University Computing Service have been of inestimable value to the study.

References

ARCHIBALD, I. & WILSON, K. S. (1978). Unpublished results.

ARNDT, U. W. (1968). *Acta Cryst.* B**24**, 1355–1357.

ARNDT, U. W., CHAMPNESS, J. N., PHIZACKERLEY, R. P. & WONACOTT, A. J. (1973). *J. Appl. Cryst.* **6**, 457–463.

ARNDT, U. W. & WONACOTT, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

BIESECKER, G., HARRIS, J. I., THIERRY, J.-C., WALKER, J. E. & WONACOTT, A. J. (1977). *Nature (London)*, **266**, 328–333.

BLAKE, C. C. F. & PHILLIPS, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: International Atomic Energy Agency.

COX, E. G. & SHAW, W. F. B. (1930). *Proc. R. Soc. London*, **127**, 71.

DODSON, E. J. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 204–211. Copenhagen: Munksgaard.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

FRENCH, S. & WILSON, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

IRWIN, M. J., NYBORG, U., REID, B. R. & BLOW, D. M. (1976). *J. Mol. Biol.* **105**, 577–586.

JOHNSON, L. N., MADSEN, N. B., MOSLEY, J. & WILSON, K. S. (1974). *J. Mol. Biol.* **90**, 703–717.

JOHNSON, L. N., WEBER, I. T., WILD, D. L., WILSON, K. S. & YEATES, D. G. R. (1978). *FEBS (Fed. Eur. Biochem. Soc.) Proc. Meet.* **11**, 185–194.

LEVINE, M., MUIRHEAD, H., STAMMERS, D. K. & STUART, D. I. (1977). *Nature (London)*, **271**, 626–630.

MATTHEWS, B. W. & CZERWINSKI, E. W. (1975). *Acta Cryst.* A**31**, 480–487.

MILLEDGE, H. J. (1966). *Acta Cryst.* **21**, A220.

NOCKOLDS, C. E. & KRETSINGER, R. H. (1970). *J. Phys. E*, **3**, 842–846.

SCHWAGER, P., BARTELS, K. & HUBER, R. (1973). *Acta Cryst.* A**29**, 291–295.

STURA, E. A. (1978). Private communication.

WHITTAKER, E. J. W. (1953). *Acta Cryst.* **6**, 218.

WILSON, A. J. C. (1977). Private communication.