DATA COLLECTION FOR PROTEIN STRUCTURE INVESTIGATION WITH SYNTEX DIFFRACTOMETER

YU. V. NEKRASOV

Institute of Crystallography, USSR Academy of Sciences, Moscow

The protein crystals are a hard object for investigation. They give a lot of weak reflections and are unstable.

The main experimental problem is the accuracy of data. The solving of this problem is radically improved with the use of such instrumentation as synchrotron radiation and area detectors.

The single-channel diffractometers with sealed X-ray tubes are also used. Half of all the protein structures described in 1988 were solved with such devices.

The Syntex $P2_1$ is a 4-circle single-crystal X-ray diffractometer. Our device is 17 years old.

We started with the improving of its collimation system and software. The work over software was going parallelly with conducting the structural experiments.

Let us remember a scheme of experiment. It includes the preliminary stage (in manual mode) and the data collection (in automatic mode).

Scheme of experiment.

Determination of the orientation matrix.	Search for reflections. Centering of reflections. Indexing. Calculation of matrix. Refinement of matrix.
Selection of the data collection parameters.	Measurement of the integrated intensity. Analysis of the peak profiles.
Procedure for the absorption correction.	Measurement of the "Absorption curve".

Data collection.	Measurement of the integrated intensity of reflections.

1. Preliminary stage

1.1. *Search for reflections*

We use a random search by biaxial synchronous scans along ω and ϕ, with transitions along X. The area of search is automatically divided into 3 ... 4 standard (ϕ, χ) zones for facility of indexing of found reflections (Nekrasov, 1988).

The optimum 2θ area is:

$$2\theta = \sin^{-1}(n\lambda / 2.d), \quad n = 7 \ldots 10$$

The scan rate is:

$$V < 60 \frac{\delta}{n^2 (\sqrt{I + B} + \sqrt{B})^2} I^2 \qquad (1.1)$$

where I is the peak intensity, counts/sec.

B is the background intensity, counts/sec.

δ is the value of scan step, °.

n is the value of fluctuations of (I + B) and B expressed in terms of their standard deviations.

The practical values of the parameters are given in tables (Nekrasov and Sosfenov, 1989).

Boundaries of optimum 2θ area of search, °.

d, A°	Mo Kα	Cu Kα
10	28.8 ... 41.7	61.2 ... 100.9
20	14.3 ... 20.5	31.3 ... 45.3
50	5.7 ... 8.2	12.4 ... 17.7
100	–	6.2 ... 8.8
200	–	3.1 ... 4.4
500	–	1.3 ... 1.8

The scan rate, °/min, was calculated from (1.1); the intensity given in the table was divided for calculation by factor 5 (that is to find the reflections at the level 0.2 of their peak intensity),

$\delta = 0.25°$, $n = 1$

B, counts / sec	I, counts / sec: 25	50	100	250
0	75	150	300	750
1	50	113	246	661
5	32	80	193	566
10	23	63	161	504

1.2. *Centering of reflections*

It is the most important program of the preliminary stage. The main peculiarities of our program are the following.

1. Optimization of the scan rate:

$$V = (I / \beta I_{max})\, V_{max}$$

where I is the peak intensity,

I_{max} is the upper limit one for using the V_{max},

β is the axis weight ($\beta_{2\theta} = \beta_{\omega} = 1$, $\beta_{\chi} = 2 \sin \theta$).

2. Optimization of the scan interval. The scan is finished when a condition has been fulfilled:

$$\Sigma N < mT\alpha I$$

or after exhaustion of given scan steps number. In this expression,

Σ N is the running sum of counts over m steps,

$T = (V_{max} / V)\, T_0$ is the duration of a step,

T_0 is the nominal one,

α iş the given level of the peak intensity: 0.65.

3. Symmetrical scan of the peak, from "center" to edges:

$$A_0 = \frac{A_{01} \Sigma N_1 + A_{02} \Sigma N_2}{\Sigma N_1 + \Sigma N_2}$$

where A_{01} and A_{02} are the "centers" of left and right halfs of peak,

ΣN_1 and ΣN_2 are the integrated intensities of them,

A_0 is the "center" of peak.

4. Quick execution of the preliminary iterations:

$$I_{max} \text{ (prelim.)} = (0.1 \ldots 0.2)\, I_{max}$$

1.3. *Indexing of found reflections*

We use the firm "Autoindexing" program with some modifications like those introduced by Clegg (1984). If the reflections are properly distributed and centered their indexing is not a hard task.

1.4. *Refinement of orientation matrix*

The original matrix for protein crystals has not a high accuracy. One uses it for selection of special matrix reflections. We select 3 reflections with long, orthogonal reciprocal vectors. They are multipled and centered along the following scheme:

$$\begin{aligned} H\,K\,L\,(\psi_1) \rightarrow \quad & -H-K-L \quad (\psi_1) \\ & H\;\;K\;\;L \quad (\psi_1 + 180) \\ & -H-K-L \quad (\psi_1 + 180) \end{aligned}$$

It is desirable among selected reflections to have one with χ angle $= 90\ (+\ 10)^\circ$.

A useful routine is "Calculation of indices of reflections with the given angles".

1.5. *Test for quality of crystal and orientation matrix. Selection of data collection parameters*

A simple but not sufficient indicator is the lattice parameters if there are equal axes and/or known angles.

The more important indicator is the position of reflection profile with regard to the scan range calculated. That is tested by measurement of strong reflections spread in the reciprocal space. Among them there must be a reflection with χ angle = 90 ($\pm$ 10)° that allows the azimuthal rotations. Measurement of such reflection gives valuable information about both the anisotropy of the mosaic spread in the crystal and the quality of the orientation matrix.

We use at this stage the azimuthal rotations with measurement of the background at the ends of the incomplete scan range. The ψ angle corresponding to the largest sum of backgrounds shows the most wide profile of the peak examined. The ψ angle corresponding to the largest difference of backgrounds shows the largest inaccuracy of the matrix.

The scan range is selected so that 95 ... 98% of integrated intensity of the widest peaks was registered. The interval between background measurement points is usually double the scan range. If a crystal has a very large cell, the background is measured in the interstices.

The detector aperture is selected by repeated measurements of the integrated intensity of strong reflections with varied aperture.

1.6. *Procedure for absorption correction*

We use the azimuthal rotations method following North, Phillips & Mathews (1968).

1.7. *Some auxiliary programs*

"The Manual Mode". It is a modification of the main program engaged with the keyboard. The intensity measurement acquires a single character, and the results are printed, not verified and not stored. This mode is also used when necessary for any checking in the course of data collection.

"Fast estimation of intensity": a standard scan in the range 0.50° with rate 0.25°/sec. The indices and intensity are printed out if the latter exceeds a given threshold.

2. Data collection

Our program has some peculiarities: a more full optimization of measurements; data collection in preset regions of the reciprocal space; versa-

tile monitoring of the course of experiment; regular information on the course and quality of the experiment.

2.1. *Optimization of measurement of peak intensity*

The optimization includes: (1) selection of scan speed for the first (estimating) measurement of intensity; (2) selection of scan speed for the second (main) measurement, as usual; (3) monitoring of the pace of the data collection.

This process is carried out by regular recalculation of the Data Collection Variable Parameters (DCVP) on the basis of comparison of the current values of the Optimization Parameters (OP) with their target values.

The DCVP are:

t_1 is the duration of the first measurement of intensity,

t_2 is the duration of the second measurement,

t_{min} is the minimum duration of measurement,

t_{max} is the maximum one,

$(I/\sigma)_{min}$ is the low accuracy limit of the first measurement for execution of the second measurement.

The OP are:

q_{01} is the relative quantity of reflections measured with statistical accuracy $(I/\sigma)_{01}$ or better.

q_{02} is the one measured with statistical accuracy $(I/\sigma)_{02}$ or better, $(I/\sigma)_{02} > (I/\sigma)_{01}$.

$\varepsilon_{01} = \Sigma\sigma/\Sigma I$ is the "mean relative statistical error" calculated over the reflections of the q_{01} group.

P_{min} is the minimum pace of data collection, reflection per hour.

P_{max} is the maximum one.

The DCVP are recalculated after measurement of current group consisting of n reflections according to the results over the last m groups; as

usual, t_2 is calculated for each reflection when the condition $(I/\sigma)_1 > (I/\sigma)_{min}$ is fulfilled).

The formulae are:

$$t_1 = MAX\ (q_{01}/q_1,\ q_{02}/q_2,\ \varepsilon_1/\varepsilon_{01})^2\ t'_1$$

$$t_2 = [(I/\sigma)^2_{max}/(I/\sigma)^2_1 - 1]\ t_1$$

$$t_{min} = (P/P_{max})\ t'_{min}$$

$$t_{max} = (P/P_{min})\ t'_{max}$$

$$(I/\sigma)_{min} = (I/\sigma)'_{min} + [1 - MAX\ (q_{02}/q_2,\ \varepsilon_1/\varepsilon_{01})^2]$$

There are the current values of parameters in the above formulae:

$$q_1 = \Sigma\ w\ n_1\ /\ \Sigma\ w$$

$$q_2 = \Sigma\ w\ n_2\ /\ \Sigma\ w$$

$$(I/\sigma)'_{min} = \Sigma\ w\ (I/\sigma)_{min}\ /\ \Sigma\ w$$

$$\varepsilon_1 = \Sigma\ w\ \varepsilon_1\ /\ \Sigma\ w$$

$$t_1' = \Sigma\ w\ t_1\ /\ \Sigma\ \psi$$

$$t'_{max} = \Sigma\ w\ t_{max}\ /\ \Sigma\ w$$

$$t'_{min} = \Sigma\ w\ t_{min}\ /\ \Sigma\ w$$

$$P = \Sigma\ w\ P'\ /\ \Sigma\ w$$

$$P' = n\ /\ (E_m - E_{m-1})$$

n_1, n_2 are the current values of n corresponding to values q_{01}, q_{02}. E is the exposure, hour.

The sum is taken over m groups, w is the weight of group.

The parameter q_{01} is controlling the duration of the first measurement. The parameters q_{02} and ε_{01} are controlling both the duration of the first measurement and the low accuracy limit for execution of the second measurement.

This algorithm ensures the following possibilities:

1) to obtain the data of the given statistical quality without unnecessary loss of the productivity of the diffractometer and

2) to plan an experiment along four lines:

a) with the resulting mean statistical error equal to the set one for regulated quantity of reflections,

b) with the regulated numbers of reflections measured with statistical accuracy equal to the set one or better,

c) and (d): the same, but the pace has priority.

The efficiency of optimization depends on the homogeneity of the intensity of reflections. Therefore the data collection must be fulfilled in spherical layers in reciprocal space, by inverted passage of the rows and with allowance for absences.

2.2. *Optimization of measurements of background*

We can vary the position and the method of background measurement. The position can be at sides of the peak, independent of scan range, or in the interstices. The method can be up to the reception of a given number of counts or in time equal to 1 or 1/2 of the scan time.

The measured surface of the background can be smoothed off-line by least squares (for the peripheral layers) and the individual corrections are calculated for each reflection.

2.3. *Regions of measurements*

The program allows one to assign four regions of reciprocal space with automatic transition from one region to another. They can be spherical layers, planes, rows or their combinations.

User gives the individual parameters for each region: the scan range, the method of background measurement and the minimum pace.

2.4. *Monitoring of experiment*

The monitoring covers eight lines:

- the radiation decay of crystal, by measurement of the check reflections,

- the stability of the orientation of crystal in the capillary by profile analysis for each non-zero reflection. When there is a small shift of peak, the addition steps can be done (Syntex); the integrated intensity is computed as the largest sum over the given number of steps in the scan range.

- the sudden changes in the orientation of crystal and also the overall effi-

ciency of the instrument, by analysis of the statistical accuracy of measurements (see 2.1),

- the misalignment of the backgrounds is verified for each reflection; the high background can be replaced by the low one,

- excessive intensity, in three ways: from the number of counts in each scan step, from the counting rate in each step (Syntex) and from the number of counts registered at each 0.001 sec. The last control allows work without beam stop.

2.5. *Information on the course and quality of the measurements*

The information of three types is printed: the special messages, the row of check reflections and the row of current reflections. Moreover, at any time one can switch on full printing for any number of reflections.

The special messages are: start of a new region of measurements (the main parameters), recentering of the matrix reflections, cause of stoppage, etc.

Row of check reflections: deviation of intensity from the initial value; number of reflections with background disbalance; HKL of current reflection; number of measured reflections and exposure.

Row of current reflections: ε_1, q_1, q_2 (see 2.1) accumulated for the current region of measurements; HKL and intensity of strongest reflection, maximum background, mean intensity and mean background – in measured group of reflections; HKL of current reflection; number of measured reflections and exposure. This information is printed at given time interval and also at the end of each region of measurements.

3. Concluding remarks

The described software allows data collection for complicated objects practically without intervention of the operator.

We investigated hundreds of protein and tens of organic and mineral crystals, among them there were the crystals of a virus with the lattice parameters 480 A°. The statistical accuracy of data was very close to the given one in all cases when the crystals allowed this at a reasonable time.

Abstract

The author's software and practical recommendations for investigations with a single-crystal diffractometer are briefly considered. An algorithm is proposed for optimized data collection. The algorithm ensures the planned quality of data without unnecessary loss of the productivity of the diffractometer. The monitoring of experiment and the information about quality of measurements are also discussed.

REFERENCES

1. Clegg,W. (1984). J. Appl. Cryst *17*, 334-336.
2. Nekrasov, Yu. V. (1988). Soviet Phys. Crystallogr. *33*, 471-473.
3. Nekrasov, Yu. V. and Sosfenov, N. I. (1989). In: *Methods of Structure Analysis.* Moscow, Nauka, pp. 140-153 (in Russian).
4. Syntex. The $P2_1$ operation manual, 1974.