Acta Cryst. (1968). B24, 289

Rapid measurement of large numbers of reflection intensities for proteins

By NGUYEN-HUU XUONG, *Department of Physics, and* JOSEPH KRAUT, OLIVER SEELY, STEPHAN T. FREER and CHRISTINE S. WRIGHT, *Department of Chemistry, University of California, San Diego, California, U.S.A.*

(*Received* 11 *September* 1967)

Data collection rates for protein crystallography can be greatly increased by combining modern digital image-processing procedures with simple photographic techniques that do not require the use of layer-line screens. A comparison of oscillation and precession photography for this purpose shows the latter to be somewhat more efficient. Intensities for 1698 reflections from chymotrypsinogen, obtained by screenless precession photography, have been found to agree with corresponding diffractometer measurements to within 6%, on the average. It is estimated that the proposed method will yield a set of high-resolution protein data in about one-tenth the time required by our present automatic diffractometer. Other advantages are also suggested.

One of the greatest obstacles to more rapid progress in protein crystallography remains the enormous number of reflection intensities that must be measured. Consider, for example, a protein of molecular weight 25,000 for which four heavy atom derivatives are available. In such a case (a fairly typical one), the five crystal species give approximately 7×10^4 unique reflection forms (hkl or $\bar{h}\bar{k}\bar{l}$ and all symmetry-related reflections) within a Bragg limit of 2·25 Å. If, as would certainly be desirable, Friedel-related reflections are to be measured separately (Matthews, 1966), and if two replications are required for each unique intensity, then we are faced with the necessity of having to make about 3×10^5 intensity measurements.

How long can one expect such a task to take? Obviously the answer depends upon X-ray tube intensities, crystal size and the equipment available. In this laboratory, recent experience with a Hilger–Watts automatic diffractometer suggests that a rate of about $1{\cdot}5 \times 10^4$ useful measurements per month can be sustained. Thus, at least 20 months would be required to obtain the desired data for this one protein with presently available commercial equipment.

The object of this note is to point out that greatly increased data acquisition rates can be realized by combining very simple photographic techniques with modern digital image-processing methods, and to present the results of some preliminary tests of the proposed procedure.

First, let us briefly consider the question of how the photograph can be photometered most efficiently. Obviously, the principal requirement to be met by the photometering instrument is that it be able to process a film in a time that is very short compared with the time required to expose the film, since this will enable one photometer (an expensive device) to handle the outputs of several X-ray cameras (relatively cheap devices). Fortunately, such instrumentation has already been highly developed for applications in several other disciplines, such as high-energy physics (Alder, Fernback & Rotenberg, 1966) and space-exploration (Nathan, 1966), and may be readily adapted for our purposes. An example is described by Abrahamsson (1966). A somewhat similar instrument is also commercially available from Technical Operations, Inc., Burlington, Massachusetts. One of these has been interfaced to an IBM-1131 computer in this laboratory, to provide a system that is able, in 5 minutes, to scan an entire 12 cm × 12 cm film on a 0·2 mm × 0·2 mm grid and store the complete digitized film-image on the computer's magnetic disk. The time required for the computer to process the image and obtain a set of integrated intensity measurements is variable, depending upon the program, but for the procedure we have adopted, it is about one second per reflection. Much shorter times would be possible if the photometer were connected to a computer with a large high-speed memory, but even with the present system only about 20 days of IBM-1131 time are needed to obtain a million integrated intensity measurements, including allowance for input-output operations.

Clearly, the essential technical problem is not to increase photometering speeds, but rather, to optimize photographic methods. This aspect of the overall question of efficient data acquisition is the subject of the remainder of the present note.

The most important thing to recognize, when considering photographic methods of data collection, is that layer-line screens are no longer necessary when indexing of reflections is done by computer. The sole purpose of a layer-line screen, in either Weissenberg (oscillation) or precession photography, is to produce an easily interpretable pattern of spots by limiting the reflections that are allowed to register on the film to those lying in a single plane of the reciprocal lattice. With an appropriate program, however, the computer does not require this simplification, and so all reciprocal lattice points passing through the sphere of reflection may be recorded. If a sufficiently small region of reciprocal space is swept out, overlap of reflections is kept within admissible bounds, and background due to incoherent and air scattering is not too large. On the other hand, a sufficiently large region of reciprocal space must be recorded so as not to lose too great a proportion of the data by partial reflection, and to minimize the number of film-sets (and scale constants) required.

In order to compare the efficiency of precession and normal-beam oscillation geometry for layer-line-screenless photography, a pair of FORTRAN programs have been written to index films obtainable with either kind of camera, given any camera setting, crystal lattice or crystal orientation. The programs allow for splitting or extension of spots, and discard all reflections which are either too close to others, (1 mm in the test case), or may be in danger of partial recording owing to edge effects. In Table 1, these categories are represented by rows 2 and 3 respectively. It should be noted in passing that overall reflection splitting on the precession camera can be minimized by setting the casette holder forward by about 10 mm.

Table 1 shows the results obtained with these programs for the case of a crystal of chymotrypsinogen ($P2_12_12_1$, $a=52{\cdot}0$, $b=63{\cdot}9$, $c=77{\cdot}1$ Å) mounted with **a** parallel to

Table 1. *Efficiencies of precession and oscillation cameras without layer-line-screens*

	Precession angle					Oscillation angle				
	1·0°	1·5°	2·0°	2·5°	3·0°	±1·0°	±1·5°	±2·0°	±2·5°	±3·0°
1. Reciprocal lattice points registered on film	1528	2356	3188	3952	4732	1002	1486	2002	2558	2990
2. Rejected because of overlap effects	0	0	127	1300	2716	0	0	192	720	1270
3. Rejected because of edge effects	1020	974	903	506	628	794	830	844	790	728
4. Measurable intensities remaining	562	1382	2158	2166	1386	208	656	966	1048	992
5. Efficiency (ratio of line 4 to line 1)	35%	59%	68%	52%	29%	21%	44%	48%	41%	33%

the X-ray beam, and moved through the indicated precession or oscillation angle. All reflections out to a minimum Bragg spacing of 2·25 Å were indexed. These results show that precession geometry is somewhat superior to oscillation geometry both in the fraction of reflections that are measurable and in the number of these recorded per film. Since exposure time is approximately proportional to the total volume of reciprocal space swept out, or equivalently to the total number of reflections registered on the film, regardless of camera geometry, the precession camera will produce data faster; and since each additional film requires an additional scale constant to be evaluated, the precession camera will yield a full three-dimensional set of data with fewer scale constants.

As a preliminary experimental check, a number of precession films, actually taken without layer-line screens, have been successfully indexed for three different proteins being studied in this laboratory. It has been found that spots may deviate as much as 0·3 mm from their theoretical positions, but this presents no serious problem since the deviation is systematic and may be compensated for by the program. In addition, one set of films, for chymotrypsinogen, has been processed to yield integrated intensities of 1698 reflections. These were found to agree with the corresponding diffractometer measurements to within 6 %, on the average.

Finally, let us consider a possible, though not optimum, strategy for collecting a full set of intensity data to 2·25 Å by the method advocated here. In the case of chymotrypsinogen, again, we see from Table 1 that a set of 2° precession films will yield intensity measurements for about 2100 reciprocal lattice points. A sequence of 30 nonoverlapping exposures, at suitable intervals of the spindle dial, will then record $6{\cdot}3 \times 10^4$ different reciprocal lattice points. This represents 57 % of all $1{\cdot}1 \times 10^5$ reciprocal lattice points under consideration. Three such sequences of exposures, taken for example in sets about three different crystal axes, would then leave only $(1-0{\cdot}57)^3$ or 8 % of the reciprocal lattice points unrecorded. These may either be ignored, or measured individually with a diffractometer. Note also that each reciprocal lattice point will have been recorded, on the average, $(6{\cdot}3 \times 10^4 \times 3)/(1{\cdot}1 \times 10^5) = 1{\cdot}7$ times, providing a replication number of about 7, on the average, for each unique intensity. We have found, in practice, that exposure times of 8 hours, with 2° precessions, produce spot densities equivalent to those obtained with 40-hour exposures in the conventional 21° precession with layer-line screen. Thus, the total camera time required to obtain the desired 90 exposures should be about one month. Since an automated photometer can easily handle the output of several cameras, it appears that not more than 2 months (instead of 20) should be adequate to accomplish the data acquisition task set out at the beginning of this communication, and in the process almost 4 times as many replications of each unique intensity will be obtained as well. It is also worth noting that the capital outlay required for equipment will be less than half that for an automatic diffractometer, that all intensity data will be obtained with crystals that have been exposed to X-rays for a maximum of 8 hours, and that absorption errors will be minimized because of the small precession angle.

References

ABRAHAMSSON, S. (1966). *J. Sci. Instrum.* **43**, 931.

ALDER, B., FERNBACK, S. & ROTENBERG, M. (1966). *Methods in Computational Physics*, Vol. 5. New York: Academic Press.

MATTHEWS, B.W. (1966). *Acta Cryst.* **20**, 230.

NATHAN, R. (1966). *Digital Video-Data Handling*, Technical Report 32877, Jet Propulsion Laboratory.

Acta Cryst. (1968). B**24**, 290

A graphical aid to indexing of reflections on 3- or 4-circle diffractometers.

By HÅKON HOPE, *Department of Chemistry, University of California, Davis, California* 95616, *U.S.A.*

(*Received* 9 *October* 1967)

A method is given whereby two- and three-dimensional representations of the reciprocal lattice of a crystal can be derived from measured diffractometer angular coordinates $\varphi, \chi, 2\theta$.

Busing & Levy (1967) have recently presented equations for angle calculations related to the operation of 3- and 4-circle diffractometers. They indicated numerical methods for initial determination of orientation parameters and cell dimensions. In connection with these initial steps a technique using a graphical interpretation of some of the procedures has been found to be useful, particularly in cases where accurate reciprocal cell dimensions have not been