An Automatic Diffraction Data Collection System with an Imaging Plate

BY ISAO TANAKA, MIN YAO, MAMORU SUZUKI AND KUNIO HIKICHI

Department of Polymer Science, Faculty of Science, Hokkaido University, Sapporo 060, Japan

AND TOMOYUKI MATSUMOTO, MICHIJI KOZASA AND CHUJI KATAYAMA

MAC Science Co. Ltd, 2-25-16 Nakanokami-cho, Hachioji, Tokyo 192, Japan

(Received 22 August 1989; accepted 20 March 1990)

Abstract

An automatic diffraction data collection system with an imaging plate has been developed for protein crystallography. The system works in a similar way to the conventional rotation camera method, but in full online mode. After exposure to the X-ray beam, the imaging plate (100 mm radius) rotates whilst a reading head scans across the plate to measure the stimulated luminescence in a record-player-like manner. During the next period of exposure, the image taken immediately before is processed in parallel. The system has been tested using both peptide and protein crystals and has been proven to work successfully.

Introduction

The imaging plate (IP), a flexible plastic plate coated with fine photostimulable phosphor crystals (BaFBr:Eu$^{2+}$) is now widely used in several scientific fields where quantitative detection of photons is necessary (Amemiya & Miyahara, 1988). Among other things it is especially suitable for diffraction data collection from protein single crystals. High sensitivity, wide dynamic range and a wide detection area with neither spatial distortion nor nonuniformity of response (Miyahara, Takahashi, Amemiya, Kamiya & Satow, 1986) are characteristic features of the IP. All these features are ideal for protein data collection, where the sample deteriorates relatively rapidly and minimization of exposure time in the X-ray beam is desirable.

While the IP has potential in new fields being explored at synchrotron facilities such as time-resolved crystallography and Laue crystallography (Hajdu et al., 1987), it should also find an application with normal X-ray sources in the laboratory as a convenient diffraction data collection system. In this paper we describe a new automatic diffraction data collection system using an IP, designed especially for protein crystallography. Diffraction data have been collected for both peptide and protein crystals, and the system has been shown to work successfully.

Hardware description

1. Configuration

A perspective view of the hardware system (called DIP100) with a schematic drawing is shown in Fig. 1. It adopts the normal-beam-with-flat-film geometry. A single disc-shaped IP with a radius of 100 mm is held perpendicular to the X-ray beam. The crystal spindle axis is in the horizontal plane and perpendicular to the X-ray beam. The crystal-plate distance is adjustable from 75 to 300 mm. Rotation or stationary patterns may be taken with this system.

Fig. 1. A perspective view of the DIP100 system with a schematic drawing showing A: graphite monochromator; B: double-hole collimator; C: goniometer; D: imaging plate disc (100 mm radius); E: servo motor; F: disc fixing mechanism; G: reading arm with erasing lamp; H: slide mechanism of reading arm; J: photomultiplier and amplifier; K: He-Ne laser; L: dark box of detector unit; M: slide mechanism of detector unit. The crystal-plate distance is variable from 75 to 300 mm. After exposure the reading arm moves onto the IP plate, and while IP is rotating, the energy of the emitted luminescence is measured.

© 1990 International Union of Crystallography
After an exposure to the X-ray beam, the IP rotates with an angular velocity of 390 rev min\(^{-1}\) min whilst a reading head scans across the plate to measure the stimulated luminescence in a record-player-like manner. An He–Ne laser is used for the excitation of the latent image and two photomultipliers, which are designed to cover the wide dynamic range of the IP, count the emitted photons. Any residual image on the IP is completely erased by the illumination of the intense visible light prior to the next exposure. The IP, the reading head and photomultipliers are enclosed in a box to shield them from natural light. The DIP100 is driven directly by a VME crate controlled by a Motorola 68000 processor, which is connected to an NEC PC9801 personal computer with 20 Mbyte hard disk.

Since the IP is similar to an ordinary X-ray film in the sense that it requires a scanning process after exposure, it has initially been applied as in camera methods where the IP replaces the X-ray film (Miyahara et al., 1986; Amemiya et al., 1988). However, in the DIP100 system, after an exposure to X-rays the image on the IP is read without dismounting the IP. Therefore, full online control of the data collection is possible.

2. Data correction

The advantage of the DIP100 readout system described above is that the readout mechanism becomes relatively simple. However a consequence of this mechanical simplicity is that a correction for the method of readout is required.

In the readout step, the IP rotates with a fixed angular velocity and photons are counted at fixed time intervals. The number of samplings per 360° rotation is 5120, resulting in a spatial interval in the annular direction of slightly less than 125 μm at the plate edge. Each count in the \((r, \theta)\) polar coordinate system is added to the nearest four \((X, Y)\) grid points on a Cartesian coordinate system. The Cartesian grid size is 125 × 125 μm with pixels in both \(X\) and \(Y\). The number of contributors to each Cartesian grid point varies locally, and therefore the data require a local sampling number correction. A further empirical correction of intensity for radius dependence is also made.

Fig. 2 shows the correction curve for radius dependence. The intensity on the IP, which was exposed uniformly to an \(^{55}\text{Fe}\) source mounted on the readout arm, was measured. The experimental curve was smoothed and used to correct for the \(r\) dependence of intensity. The abrupt drop at the film center is the effect of the pre-erasing by the laser beam at high sampling point density, and gives rise to an unusable area at the center of the IP of about 5 mm radius. In practice the correction factor for both sampling number and \(r\) dependence is calculated once for all grid points and used for data correction on a pixel-by-pixel basis.

3. Coordinate system

The Cartesian laboratory coordinate system is defined as shown in Fig. 3. The origin is on the intersection of the rotation axis with the primary X-ray beam. \(X\) is parallel to the primary X-ray beam, \(Z\) to the rotation axis and \(Y\) is chosen such that it is perpendicular to both axes and makes the Cartesian system right-handed. The ideal IP coordinate system \((X_s, Y_s)\) is defined on an ideal IP plane which is normal to the \(X\) axis. The plate is at a distance \(D\) from the crystal. \(X_f\) is parallel to \(Z\) and \(Y_f\) to \(Y\). Since the actual IP plane may be slightly inclined from this ideal plane, parameters \(w_1\) and \(w_2\) are introduced which represent twist and tilt angle of the IP plane respectively. The practical IP coordinate system is designated as \((X_a, Y_a)\) where \(X_a\) is the intersection of the \(XY\) plane with the IP plane and \(Y_a\) is the intersection of the \(XY\) plane with the IP plane. The intensity is first read onto the \((r, \theta)\) polar coordinate system and transferred to the \((X_m, Y_m)\) Cartesian coordinate system or the IP measurement.
system. The vector from the center of the rotation \((O_x)\) to the direct-beam position \((O_d)\) is \((X_o, Y_o)\). The relation between the ideal IP plane coordinate system \((X_f, Y_f)\) and the IP measurement system \((X_m, Y_m)\) is

\[
X_m = X_o + [(DX_f \cos w_2 \cos w_3 + DX_f \cos w_1 \cos w_2)
- DY_f \cos w_1 \sin w_2)
+ D \cos w_1 \cos w_2]^{-1}
\times (X_f \sin w_1 \cos w_2 + Y_f \cos w_1 \sin w_2)
\]

\[
Y_m = Y_o + [(DY_f \cos w_2 \sin w_3 + DY_f \cos w_2 \sin w_2)
+ D \cos w_2 \sin w_3]
\times (X_f \sin w_1 \cos w_2 + Y_f \cos w_1 \sin w_2)
+ D \cos w_2 \sin w_3]^{-1}
\]

\[
\cos \theta = \sin w_1 \sin w_2.
\] (1)

Any reciprocal-lattice point \((hkl)\) is projected onto the ideal IP system by the relation

\[
\left( \begin{array}{c} X_f \\ Y_f \\ 1 \end{array} \right) = C \Phi U B \left( \begin{array}{c} h \\ k \\ l \end{array} \right). \] (2)

Here, \(B\) is the orthogonalization matrix from the crystal coordinate system to the Cartesian crystal coordinate system (Busing & Levy, 1967). \(U\) is the crystal orientation matrix which rotates the Cartesian crystal coordinate system onto the laboratory coordinate system and is defined as

\[
U = \begin{pmatrix}
\cos p_3 & -\sin p_3 & 0 \\
\sin p_3 & \cos p_3 & 0 \\
0 & 0 & 1
\end{pmatrix}
\times \begin{pmatrix}
\cos p_2 & 0 & \sin p_2 \\
0 & 1 & 0 \\
-\sin p_2 & 0 & \cos p_2
\end{pmatrix}
\times \begin{pmatrix}
\cos p_1 & -\sin p_1 \\
\sin p_1 & \cos p_1
\end{pmatrix}
\] (3)

where \(p_1, p_2\) and \(p_3\) are the rotation angle in the \(X, Y\) and \(Z\) axes respectively.

\(\Phi\) represents the rotation matrix around the spindle axis to swing \((hkl)\) onto the Ewald sphere,

\[
\Phi = \begin{pmatrix}
\cos \varphi & -\sin \varphi & 0 \\
\sin \varphi & \cos \varphi & 0 \\
0 & 0 & 1
\end{pmatrix}. \] (4)

\(C\) is not written here in matrix form but it is a conversion equation from \((X_b, Y_b, Z_b)\) to the ideal IP plane system \((X_f, Y_f)\), where \((X_b, Y_b, Z_b)\) represents any reciprocal point on the Ewald sphere (Arndt & Wonacott, 1977). The relations between \((X_b, Y_b, Z_b)\) and \((X_f, Y_f)\) are

\[
X_f = \frac{Z_b - D}{1 + Y_b}
\]

\[
Y_f = \frac{Y_b - D}{1 + X_b}. \] (5)

### Software description

Fig. 4 shows the software system for data collection and processing. The program is written in Fortran apart from some basic routines where assembler language is used. The whole program system, named \(ELMS\), consists of four main programs and seven subprograms. The latter can be selected from a main programs menu or can also run as stand-alone programs. All other options are chosen from the menus of the main or subprogram. The system is aimed at real-time processing; during the time for one exposure, the image taken immediately before is processed in parallel.

1. \(ELMS\) and \(POPS\) program for offline data collection

The \(ELMS\) program was the first to be developed. Together with its subprograms \(ELMS\) covers almost all the necessary steps for diffraction data collection.

The \(POPS\) program has been developed for the post-processing of the IP data. The word 'post' means only that the processing is not done online (see \(ADAC\) below) but after data collection.

The \(DIC\) subprogram allows modification of control parameters for the program package.

The \(DIP\) subprogram is used for the manual operation of the DIP100 system. Two main commands, \(STIL\) or \(ROTT\), invoke a set of sequential operations for taking, respectively, one stationary or one rotation photograph, and displaying the image
on the PC98 screen. The IP data are transferred to the PC98 disk by entering command DFER explicitly.

The FIL subprogram is used to manipulate the IP image file stored on the PC98 disk. Nonuniformity correction, background-level estimation and peak search are the main commands supported.

The PEP subprogram allows the user to display all the control constants as well as the contents of intermediate (binary) files produced by the program package.

The PAR subprogram invokes one of the three refinement routines CCLS, RSLS(RGSLS) or SFTLS. The former two routines use the centroid coordinates of the spots from a set of stationary photographs, and the last uses those from a rotation photograph. The CCLS routine refines camera constants \( w_1, w_2, w_3 \) and \( D \). Usually the first three parameters are close to zero and only \( D \) (the crystal-to-plate distance) as read from the instrument is a required input. The quantity minimized is the sum of the squared distance between observed and calculated reflection positions

\[
M = \sum [(X_{\text{obs}} - X_{\text{cal}})^2 + (Y_{\text{obs}} - Y_{\text{cal}})^2].
\]

The RSLS(RGSLS) routine refines setting parameters. Refined parameters are \( p_1, p_2 \) and \( p_3 \) for RSLS and the cell constants in addition for RGSLS. RGSLS requests the crystal system, which is used to determine constraints on cell parameters during the refinement cycle. The quantity minimized is

\[
M = \sum (1 - R_{\text{cal}})^2,
\]

where \( R_{\text{cal}} \) represents the distance of each reciprocal-lattice point in the diffracting condition from the center of the Ewald sphere (measured in dimensionless reciprocal-lattice units). The SFTLS routine refines geometrical shift parameters \( a_{11}, a_{12} \) etc. as defined by

\[
X = a_{11} + a_{12}X_m + a_{13}Y_m + a_{14}X^2_m + a_{15}Y^2_m
\]

\[
Y = a_{21} + a_{22}X_m + a_{23}Y_m + a_{24}X^2_m + a_{25}Y^2_m
\]

for each rotation photograph by minimizing the discrepancy between observed and calculated coordinates. Since the routine expects data from a rotation, the specialized indexing routines IDX and GEN (see below) are used for indexing. Because the equation used is linear no iteration is necessary for the refinement itself, but the indexing cycle is repeated with new parameters until the program achieves internally fixed termination criteria.

The GEN subprogram generates reflections within a given rotation range, and displays them on the screen or creates a predicted coordinate file which is used in the INT subprogram. Special care has been taken for speedy calculation. The generation time for a relatively large unit cell, \( 100 \times 100 \times 100 \) Å with resolution limit 2 Å and rotation range 2° is about 1 min, sufficiently fast for practical use.

The INT subprogram integrates all the reflections in a file made with the GEN subprogram.

2. ATIX automatic indexing program

The ATIX program is designed for the automatic determination of the crystal orientation, whose cell dimensions are already known. The program requires spot coordinates from at least two stationary patterns 90° apart. These are automatically prepared with the ADAC program below. The program consists of five steps.

Step 1: The reciprocal coordinates (\( P_i \)) of the reflections on stationary patterns and the vectors between them (\( P_{ij} \)) are calculated. The shortest 100 vectors, whose frequency of occurrence is greater than a given value, are kept in a somewhat similar way to that of Kabsch (1988).

Step 2: All reciprocal-lattice vectors (up to 500) whose reciprocal length is within a given range are generated from the given cell parameters.

Step 3: The shortest \( N \) vectors (usually \( N = 20 \)) in step 1 are used for this step. The triangles formed by any pair of vectors (whose angle is closer to 90° than a given angle) are compared with those formed by the calculated vectors in step 2. The combinations for which all three edge-length differences are shorter than a given threshold are kept. The best 100 combinations in terms of sum of the edge-length differences are kept for the next step.

Step 4: For all 100 combinations, the setting parameters \( p_1, p_2 \) and \( p_3 \) are calculated by the double-reflection method (Busing & Levy, 1967) and refined by RSLS using only the shortest \( N \) of the \( P_{ij} \). The results are sorted based on the number of rejected reflections and the sum of the absolute shift angle of \( p_1, p_2 \) and \( p_3 \).

Step 5: The best solution in step 4 is subject to the refinement using all 100 \( P_{ij} \) 'reflections' followed by one using all \( P_i \) reflections. The refinement terminates after repeated use of RSLS, CCLS, RSLS, CCLS and RSLS. An optional switch invokes the refinement of cell parameters.

We recently noticed that the algorithm we described here is almost identical to one used by Tucker (1986). The only difference is that in step 3 we use the sum of three edge-length differences for the judgement of fit, whilst Tucker uses two lengths and one angle.

3. ADAC automatic data collection program

ADAC is an automatic data collection and online processing program. In the ADAC screen menu, one of the two functions ASTI or ADAC is chosen. ASTI is for taking a sequential set of stationary photo-
graphs and ADAC for a sequential set of rotation photographs.

ASTI takes stationary photographs, corrects for nonuniformity, estimates the background level and its standard deviation, displays the IP image on the screen, saves the graphic image, searches for peaks above threshold and saves these in a file.

ADAC takes rotation photographs, corrects for nonuniformity, estimates the background level and its standard deviation, displays the IP image on the screen, saves the graphic image, searches for peaks above threshold and saves these in a file.

The automatic reorientation routine is controlled by checking unindexable reflections and the r.m.s. errors between observed and calculated positions, all of which are output data from SFTLS.

Test data collection

In order to test the system, data from two crystals were collected. One was a peptide crystal (Tanaka, Kojima & Ashida, 1977) and the other an egg white lysozyme crystal. The Laue groups of both crystals are tetragonal $P4/mnm$. Therefore crystals were rotated by 45° for data collection with their c axes parallel to the spindle axis.

Fig. 5 shows schematically the egg white lysozyme diffraction data collection to a resolution limit of 2.3 Å using a graphite-monochromatized Cu Kα radiation from a rotating-anode X-ray generator (50 kV, 250 mA). In this case each exposure step covers 1.5° of rotation, and thus 30 oscillations give the required 45° of rotation. When the second exposure starts, the processing of the data from the first image also starts in parallel on the PC9801. One 'exposure' consists of erasing any residual image, the actual exposure, reading out and data transfer. The time required for the actual exposure depends on several factors such as crystal quality, desired resolution limit and the power of the X-ray source, but normally it is in the order of 10 to 30 min with a rotating-anode source. In the present case 25 min of exposure time was chosen and collection of an image takes about 30 min altogether. The processing of an image consists of data correction, reflection generation, and integration, which took 3.0-3.5 min respectively in the present case. Therefore, while exposing the IP for a second rotation range, the processing of the first image is finished. The time-limiting step is the exposure step. After 15 h, which is the sum of exposing steps, all the data set up to 2.3 Å were ready in the form of squared structure-factor amplitudes $I(hkl)$. The time the crystal was exposed to the beam was 12.5 h.

Figs. 6(a) and (b) show $R_{\text{merge}}$ of the peptide crystal data plotted against average intensity level and resolution. For those reflections with a strong intensity $R_{\text{merge}}$ is about 3 to 4%. When Cu Kα radiation is used for data collection, the maximum

![Figure 5](image-url)

Lysozyme Data Collection (2.3 Å) Timetable

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st. exposure</td>
<td>1st. processing</td>
</tr>
<tr>
<td>2nd. exposure</td>
<td>correction</td>
</tr>
<tr>
<td>last exposure</td>
<td>generation</td>
</tr>
<tr>
<td>last processing</td>
<td>integration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 min.)</td>
<td>correction</td>
</tr>
<tr>
<td>(25 min.)</td>
<td>generation</td>
</tr>
<tr>
<td>(3 min.)</td>
<td>integration</td>
</tr>
<tr>
<td>(1 min.)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. The timetable for 2.3 Å lysozyme diffraction data collection and real-time processing. The data collection conditions are given in Table 1.

![Figure 6](image-url)
Table 1. Lysozyme data collection (2-3 Å) summary

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>Rotating anode (50 kV, 250 mA)</td>
</tr>
<tr>
<td>Detector</td>
<td>Fuji imaging plate</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
</tr>
<tr>
<td>Collimator</td>
<td>0.3 mm Φ</td>
</tr>
<tr>
<td>Camera distance (mm)</td>
<td>100</td>
</tr>
<tr>
<td>Rotation range/film (°)</td>
<td>1-5</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>45</td>
</tr>
<tr>
<td>Total exposure time (h)</td>
<td>15</td>
</tr>
<tr>
<td>Merging R (full)</td>
<td>5.6% (10428)</td>
</tr>
<tr>
<td>(full + partial)</td>
<td>6.8% (30234)</td>
</tr>
</tbody>
</table>

resonance reached is 1.7 Å with this device. Up to this resolution, data were collected and processed automatically without any problem and the overall $R_{merge}$ was 3.7%.

For the data from the protein crystal, $R_{merge}$ was slightly worse, mainly because of the weaker diffraction intensity from the crystal. Fig. 6(c) shows the $R_{merge}$ plotted against resolution. All reflections including weak ones were used for the calculation. The scaling factors within 30 files were also satisfactorily constant. Table 1 is a summary of the data collection of the lysozyme crystal.

Concluding remarks

An automatic data collection system has been developed for protein crystallography using an imaging plate. The computer requirement is particularly simple and satisfactory results were obtained from test data collection on both peptide and lysozyme crystals. Although the DIP100 can be connected to much larger computer systems, the minimum computer requirement is one personal computer with 20 Mbyte hard disk. Because of the relatively simple software required for the DIP100 system, online processing is possible. The system is being used for actual structure analysis in several laboratories including the Protein Engineering Research Institute in Osaka and EMBL in Heidelberg. In one of the applications [cytochrome $c_2$ from *Rhodopsseudomonas viridis* (Miki, Saeda, Masaki, Kasai, Miki & Hayashi, 1986)] the data obtained by the present system were compared with the diffractometer data. The $R_{merge}$ between two data sets was 8.21% with 2-5 Å resolution (10739 reflections) and 5.86% with 4 Å resolution (3685 reflections). More recently we collected 2-3 Å resolution data from a single crystal of ribonuclease H mutant (E48Q) (Kanaya, Kohara, Miyagawa, Matsuaki, Morikawa & Ikehara, 1989). Starting from the refined coordinates of the native crystal, the crystallographic $R$ factor dropped to 24% after some 100 cycles of Hendrickson & Konnert least-squares refinement and before including solvent molecules. The resulting Fourier map clearly shows the regions where the molecule deviates from the input model, suggesting good quality of the data (manuscript in preparation).

The authors thank many students in Hokkaido University for their help in various stages of the program development, especially Mr Chiharu Yamazaki and Mr Shun-ichi Nagasawa. We thank Dr Kunio Miki for making unpublished results available. We also thank Mr Howard Terry at EMBL Hamburg outstation in DESY for reading the manuscript, and Dr Paul Tucker of EMBL Heidelberg for valuable comments. This work was supported in part by a Scientific Research Grant from the Ministry of Education.

References


