Protein, DNA, and Virus Crystallography with a Focused Imaging Proportional Counter


Copyright © 1986 by the American Association for the Advancement of Science
Protein, DNA, and Virus Crystallography with a Focused Imaging Proportional Counter


A set of programs has been developed for rapid collection of x-ray intensity data from protein and virus crystals with a commercially available two-dimensional focused geometry electronic detector. The detector is compact and portable, with unusually high spatial resolution comparable to that used in oscillation photography. It has allowed x-ray data collection on weakly diffracting crystals with large unit cells, as well as more conventional “dichroism-height” crystals. The quality of the data is compared with that from oscillation photography and automated diffraction in the range of unit cells from 96.3 to 383.2 angstroms. Isomorphous and anomalous difference Pattersons, based on detector data, are shown for a variable surface glycoprotein mercury derivative and for a repressor-DNA bromine derivative, which has been solved at 7 angstroms with detector data only.

The most efficient method of collecting x-ray diffraction data is to record at one time with a photon counting detector as many as possible of the simultaneously occurring x-ray reflections. Conventional proportional counters have no spatial resolution, and thus conventional diffractometers collect only one or a few reflections at once. Film is an efficient but imperfect area detector, and widely used in the unit cell range above 100 Å (1). After the pioneering work by Xuong (2, 3) and by Arndt (4), several groups have developed data collection systems using two-dimensional photon counting electronic detectors with sufficient spatial resolution to assign every diffracted photon to the correct reflection (4, 5).

We have written software for rapid data collection with a novel commercially available detector that is portable enough to be moved from one x-ray generator to another in the laboratory. The detector generates a 512 by 512 pixel image of all simultaneously occurring diffraction. Its total angular subtense is 60 degrees when set at 11 cm from the crystal. A 5 minutes of arc oscillation exposure of tobacco bushy stunt virus (TBSV) (Fig. 1) demonstrates a novel feature of this instrument—high spatial resolution. The 512 by 512 pixel image has about 200-μm resolution at the detector face. This makes it roughly equivalent to digital film scanning with a 200-μm raster, except that the counter suffers none of the spot-edge effects (6) that limit optical density raster-scanning. Thus, the only condition to be met is that adjacent reflections be spatially resolved. Using Franks double mirror focusing optics (7) on an Elliot GX-6 rotating anode, unit cell dimensions in the 100 to 180 Å range arc readily resolved at crystal detector distances of 10 cm, and larger cells are resolved at 18 cm. Figure 1 shows diffraction from crystals of TBSV (a = 385 Å). The spots within the marked circle (Fig. 1A) at 2.6 Å resolution (specimen to detector distance 18 cm) can be seen on an expanded scale (Fig. 1B) to form a resolved, centered lattice.

The x-ray detector is a focused geometry imaging proportional counter, which is part of a data collection system (Nicolet-Xenotronics, Madison, WI). The counter is a sealed, xenon-filled unit, with a concave Be window (11.5 cm in diameter). The curvature reduces parallax at the edges of the detector by making the electric field lines more parallel to the path of the incident photons in the interaction zone near the front window. The signal is collected on a multiview anode and read capacitively on cathodes of more finely spaced wires (one set of wires in x and one in y). The detector and position-determining circuit combine a coarsely segmented cathode (2 cm) with capacitive charge division within a segment to produce resolution of 0.20-nm full width at half height. High resolution is achieved in both the x and y dimensions because electrons created in the initial ionization event are allowed to diffuse in the drift space, creating a relatively large (3 to 4 mm) cloud of ions. The distribution of the electrons is recorded on a number of wires from which the centroid and therefore the photon position can be calculated to a fraction of the actual wire spacing. The current instrument has six cathode segments in each direction and 12-bit analog to digital converters. The electronics have a dead time of 5 μsec. The current limiting factor for data collection rate is the 18-μsec time required by the data collection microcomputer to record and map (see below) each event. This characteristic permits acquisition rates up to 60 kHz at 50% dead-time loss (the unapped mode rate is 100 kHz). Our experience indicates that the data rate is limited by the power available from our x-ray source (Elliot GX-6).

R. M. Durbin, J. Moult, P. Metcalfe, D. Freymann, M. Blum, J. E. Anderson, S. C. Harrison, D. C. Wiley, Department of Biochemistry and Molecular Biology and Committee on Higher Degrees in Biophysics, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138. R. Burns, Nicolet-Xenotronics Corp., Verona Road, Madison, WI 53711.

*Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.
rotating anode, 100-μm focusing cup, Franks double mirror collimation optics). During 12 months of operation, the sealed detector has never failed, and no dead spots have appeared. The detector is used on a rotation oscillation camera (Supper, Natick, MA) with a vertical rotation axis. The detector can therefore be swung around in the horizontal plane to collect high angle data. Crystal to detector distance is variable: 12 cm was used for lysozyme, variable surface glycoprotein (VSG), influenza virus hemagglutinin (HA), and repressor-DNA, and 18 cm for TBSV (8). The data acquisition computer (CADMUS 9730) includes 1.5 megabytes of memory, a 65-megabyte Winchester disk, and a nine-track tape drive. Output is displayed on a color raster graphics terminal.

The direct output from the electronic detector is an x and y coordinate of the position of incidence of each photon. Several factors distort this output. The two most significant are anode modulation, caused by the presence every 1 mm of anode wires, which locally attract more charge, and pin cushion distortion, created by the curvature of the detector face. The first of these is corrected at the time of data collection by a mapping function from the detector position outputs (X₀, Y₀) to the 512 by 512 pixels (Xₙ, Yₙ) in the computer memory image array. The mapping function is produced with an ⁵⁵Fe source to create an even flood field. The pixel edges and size are defined [in detector output coordinates (X₀, Y₀)] so that each pixel contains 1/512 of the total intensity when the detector is evenly illuminated. Residual effects of anode modulation not equalized by this approach are compensated by taking background from the same pixels as spot intensities (see below). This correction is sufficient for the accurate integration of reflections, which is a local operation. To locate reflections accurately, we have introduced a correction for positional “pin cushion” distortion. This correction is an interpolation between curves “detector space” and “true space,” using data from the image of a brass calibration plate with a regular array of holes that fixes to the front of the detector. This two-stage correction preserves the minimal amount of positional information by dividing the axes into bins that each have an equal probability of acquiring a positionally random event. These corrections are required for the very high spatial resolution and require about 1 to 2 hours of detector time every week or two.

The crystal is first approximately aligned by a series (two to four) of “still” frames collected by the detector and analyzed on the detector graphics terminal by an interactive program developed by one of us (P.M.). This requires 5 to 10 minutes, during which crystal quality can also be assessed visually from the data displayed on the computer screen. Data are collected as a series of consecutive oscillation frames (a 5 minutes of arc angle oscillation exposure in our current practice). Each frame, a 512 by 512 pixel image, is written to magnetic tape after it is collected. A 2400-foot tape holds just over 100 frames at 1600 bytes per inch. Our strategy is to determine the precise crystal setting parameters directly from an analysis of the same data frames that are used to integrate the reflections. The method is based on techniques for oscillation-film data processing (9). This differs from the “electronic stationary picture method” (3) in which crystals are carefully aligned first and frames processed as they are collected.

Consecutive small oscillation (5 minutes of arc) frames and a film scanning like strategy offer several advantages:

1) The signal-to-noise ratio is increased relative to oscillation film methods, because spots only need to be integrated on frames in which they appear. (A typical spot may span two to five frames depending upon beam divergence, crystal mosaicity, and position relative to the oscillation axis). On an oscillation photograph, a spot is often recorded only during 1/10 of the oscillation range, while 9/10 of the time that region of the film records background scatter.

2) Crystal exposure time is not wasted in finding accurate setting parameters, since the same data are used both for setting information and intensity data.

3) The reflections are examined in three dimensions (x, y, z) on the detector face and frame number, which indexes the oscillation angle (φ), because a series of frames are analyzed at one time. Examination of a three-dimensional box (x, y, z, φ) around a reflection allows its x, y, z centroid and its rocking curve width, γ, to be determined. These are equivalent to knowing partial-spot
intensity ratios in oscillation film post-refinement, from which precise crystal alignment parameters and spot shapes can be calculated by established methods (9).

4) A knowledge of the spot position and width in three dimensions allows integration to be limited to pixels on frames containing the spot and allows background to be collected from corresponding pixels on frames “before” and “after” the reflection itself. This method of background substraction minimizes systematic errors due to pixel size variation and other nonuniformities due to anode modulation.

Our program to process the detector output is written in Fortran for a VAX 11/780 with a VMS operating system (Fig. 2). To initiate processing, a small sample (50 to 100) of intense reflections is located, usually from frames at the beginning and end of the φ range on the tape. The centroids (x, y, φ) and φ widths of the spots are determined by examining the reflections in three dimensions. These observations, along with estimates of the unit cell parameters, camera parameters, and approximate crystal setting angles, are used to predict the indices of this sample of reflections.

Three groups of parameters are then refined based on the observations and on the condition that indices be integral. (i) The unit cell dimensions and crystal alignment angles (together with the wavelength, which is known) determine the rotation angle φ (or frame number) at which a reflection appears. (ii) These, together with camera parameters (crystal to detector distance, camera tilt, and rotation), determine where on the image plane a reflection appears. (iii) Finally, the width of the rotation profile is determined by the beam divergence, dispersion, and crystal mosaicity.

The first two groups are refined together by a modified least-squares system. The errors in frame number (rotation angle) are weighted at least a thousand times more strongly than those in the detector face x, y position. This weighting ensures that the crystal parameters, which affect the rotation angle, effectively refine independently of the camera parameters, which do not. Ill determinacy of the least-squares equations due to interactions between crystal and camera parameters is thus prevented. The weighting is justified because the φ angle for spots is well measured and independent of residual distortions that may affect x, y position on the detector face. In detail, we proceed as follows. Profiles of the intense reflections chosen for the refinement are first used to refine the rocking curve and to improve the accuracy of φ (centroid) for each spot. A rocking curve of the form

\[ I = I_0 \left(1 - \sin \left((\pi/2)(\beta - \beta_0)/\gamma\right)\right) \]

is fit by a least-squares calculation, minimizing

\[ \sum_{i, spots \rightarrow frames} (I_{\beta i} - F_{\beta i})^2 \]

Here, I is intensity, integrated on the (x, y) plane, a function of φ, \(I_0\) is a scale factor, \(F_{\beta i}\) is the integrated intensity on frame f from spot i, \(\beta_i\) is the calculated intensity on frame f from spot i, and \(\gamma\) is the half width of the rocking curve.

The definition of the angle β can be found in (9). The parameter γ is assumed to have a form (9)

\[ \gamma = \gamma_0 + \gamma_1 \tan \theta_i \]

where \(\theta_i\) is the Bragg angle for reflection i, \(\gamma_0\) is a parameter to be refined, and \(\gamma_1\) is the dispersion of γ, given by \(\Delta\lambda/2\lambda = 0.0013\) for the CuKα doublet.

The rocking curve is used at this point to obtain a better estimate for the φ angle of
the reflection centroid, and it is used later in the program to determine the range of frames appropriate for spot integrations. Having determined \( \gamma \), we refine the other two parameters of the model together, minimizing

\[
\sum_{i} \left( w_{i} (x_{i} - x_{0})^{2} + w_{i} (y_{i} - y_{0})^{2} + w_{i} (\phi_{i} - \phi_{0})^{2} \right)
\]

where \( x_{i}, y_{i}, \phi_{i} \) are coordinates of the observed centroid of spot \( i \), \( x_{0}, y_{0}, \phi_{0} \) are calculated coordinates of centroid of spot \( i \), and \( w_{i} \) are weights.

We choose \( w_{i} = 0.001 \) (see above). After these two refinement steps, the observed \( \chi^{2} \) is usually about 0.2 frame (1 minute of arc) and 70 \( \mu \)m in each direction in the image plane.

A list is then produced from the refined parameters of all \( b, k, l \) values and \( x, y, \phi \) positions of reflections expected in the data.

A fast algorithm has been developed to do this, which minimizes floating point multiplication and division. The Ewald sphere (E sphere) is rotated around the \( x \)-axis by \( -\phi_{0} \), the mirror image of the starting angle for the oscillation range, to generate the P sphere. It is also rotated by \( -\phi_{0} \) (the end of the range) to generate the Q sphere. The reflections \( b, k, l \) will be included in the data if the unrotated reciprocal lattice point \((x^{*}, y^{*}, z^{*})\) falls inside only one of the two spheres P and Q. Since the square of the distance from the center of a sphere is a quadratic form, it is possible to test through the reciprocal lattice, testing inclusion between limiting spheres by keeping track of the squares of the distances from their centers. If \( r \) is the distance of the point \( b, k, l \) (at \( x^{*}, y^{*}, z^{*} \)), the distance of the “next point” (for example, \( b+\Delta b, k, l \)) can be found from the equation

\[
(r + \Delta r)^{2} = r^{2} + \Delta r^{2} + 2r \Delta r = r^{2} + (\Delta b)^{2} + 2(r) \Delta b
\]

Since most of the terms in this equation are constants, calculations are very efficient. The list of reflections is sorted on frame number, and then the frames are processed sequentially, keeping an active window of ten frames, and integrating the reflection as their centers fall in the middle of the active window. A three-dimensional integration mask, whose shape can be dependent on spot location, is used for integrating each reflection. The mask can be shifted a few pixels from the predicted position of the spot if the spot’s centroid is not exactly at the predicted position.

On VSG crystals \( (a = b = 96.3 \text{ Å}, c = 111.3 \text{ Å}) \), generating the list of predicted reflections and positions required 4.2 seconds of VAX 11/780 central processor time.

---

Table 1. Data from weakly diffracting crystals. The numbers in parentheses represent the number of reflections; \( d \) is maximum resolution.

<table>
<thead>
<tr>
<th>Crystal composition</th>
<th>Space group</th>
<th>Cell constants (Å)</th>
<th>Daltons per asymmetric unit</th>
<th>( d ) (Å)</th>
<th>*R_{sym} ) detector</th>
<th>( R_{sym} ) film</th>
<th>( R_{sym} ) conventional diffractometer</th>
<th>( R ) detector to film</th>
<th>( R ) detector to diffractometer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influenza HA†</td>
<td>P4₁</td>
<td>163.2, 177.4</td>
<td>200,000</td>
<td>3.5</td>
<td>0.09 (3000)</td>
<td>0.11 (3000)</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>434 Repressor-DNA</td>
<td>1422</td>
<td>166.4, 189.4</td>
<td>35,000</td>
<td>7.0</td>
<td>0.087 (1171)</td>
<td>0.086 (1134)</td>
<td>0.078 (1036)</td>
<td>0.020 (1331)</td>
<td>0.061</td>
</tr>
<tr>
<td>434 Repressor-Br DNA</td>
<td>1422</td>
<td>166.4, 139.4</td>
<td>35,000</td>
<td>7.0</td>
<td>0.087 (1171)</td>
<td>0.086 (1134)</td>
<td>0.078 (1036)</td>
<td>0.020 (1331)</td>
<td>0.061</td>
</tr>
<tr>
<td>Variable surface glycoprotein of trypanosome brucei (VSG)†</td>
<td>P4¹,2</td>
<td>96.3, 111.3</td>
<td>43,000</td>
<td>6.0</td>
<td>0.069 (1094)</td>
<td>0.068 (7619)</td>
<td>0.068 (7619)</td>
<td>0.080 (1365)</td>
<td>0.065</td>
</tr>
<tr>
<td>VSG IgG†</td>
<td>P4¹,2</td>
<td>96.3, 111.3</td>
<td>43,000</td>
<td>6.0</td>
<td>0.069 (1094)</td>
<td>0.068 (7619)</td>
<td>0.068 (7619)</td>
<td>0.080 (1365)</td>
<td>0.065</td>
</tr>
</tbody>
</table>

\*\( R_{sym} = \frac{1}{N} \sum_{i} \sum_{j} (I_{ij} - \bar{I}) / \sqrt{\sum_{i} \sum_{j} I_{ij}} \) where \( N_{s} \) = number of symmetry-related reflections, \( N \) = total number of sets, \( b \).

†These factors are from data processed with no variation in integration mask shape.

An empirical absorption correction (17) was applied to the 3 Å VSG data sets with a program written by C. Schutt and modified for use here (18).

---

Fig. 3. (A) Isomorphous difference Pattersons at 6 Å resolution. The sections shown are the Harker sections \( z = 1/4 \) and \( z = 1/2 \) (space group P4₁,2). The origin is at the upper left hand corner; there is mirror symmetry across the diagonal. At 6 Å resolution, three unresolved peaks occur on \( z = 1/4 \) and two unresolved peaks occur on \( z = 1/2 \). The origin peak in both maps was set to 999. The standard deviation for the diffractometer map is 19; the map is contoured in increments of one standard deviation, starting at 2 standard deviations above zero. The standard deviation for the Xentronics map is 12; the map is contoured in increments of one standard deviation starting at 2 standard deviations above zero. The Patterson maps were calculated with the coefficients \( (P - P_{0})^{2} \) for data between 16 to 6 Å. (B) Anomalous difference Patterson 6 to 4 Å Xentronics data. The sections shown are the same as in (A). The peaks numbered 1 to 5 can be interpreted in terms of a single site. Peaks 2 and 3 partially overlap. The standard deviation of the map is 12. The section is contoured at 22, 26, 30, 34, 38, 42, 46. The section is contoured at 26, 30, 34, 38, 42, 46. The section maps were calculated with the coefficients \( (K_{iso} / 4)(P - P_{0})^{2} \), where \( K_{iso} = 6.4 \).
unit (cpu) time per 1000 reflections. Integration of those spots required 101 cpu seconds per 1000 reflections. A major advantage of area detector data collection is the increased speed and ease of data collection. Data collection rates can be 100 times faster than conventional diffractometry (10) because many of the simultaneously occurring reflections can be collected at one time. Data rates are not as dramatically increased over oscillation photography, because oscillation photography already records the simultaneously occurring reflections. For identical beam geometries and crystal to detector (film) distances, there is only one essential difference between oscillation photography and area detector methods: background scatter adds to intensity within the area of a spot during the entire oscillation range of film, whereas it adds to spot intensity only on those detector frames included in the integration (that is, for only the φ range of the rocking curve). As pointed out above, the ratio of these two ranges is typically about 10:1, giving an increase in speed for comparable counting statistics of V10 = 3.2. Contributions from chemical fog and film scanning errors make the comparison even more favorable. For a particular choice of counting time, the detector also has the advantage of large dynamic range. Film to film scaling errors within a pack are avoided—an especially significant contribution for strong reflections.

Data collection from crystals unsuitable for diffractometry because of large unit cell sizes and weak diffraction is a particularly stringent test of an area detector. Such crystals require very long exposures. For example, a one-degree oscillation photograph of the influenza HA crystals requires a 12-hour exposure (Franks focusing optics, 100-μm focus, 40 kV, 20 mA, Elliot GX-6); since each reflection is 6 to 10 minutes of arc wide, individual reflections are exposed for two hours each, in contrast to the 1- to 2-minute "exposure" time for reflections from protein crystals suitable for conventional diffractometry. The current data collection system was tested on five weakly diffracting protein crystals for which high resolution data were uncollectible on a conventional diffractometer (Table 1). The unit cell dimensions range from 96.3 to 177.4 Å and the daltons per asymmetric unit from 35,000 to 200,000. R factors were calculated by comparing intensities of symmetry-related reflections. In the two cases where oscillation film data sets were available on the same proteins (influenza HA and 434 repressor-DNA complex), data measured on the detector system were of better quality than those measured by film. Low resolution data on the VSG had Friedel R factors comparable to 6 Å resolution data collected by automated diffractometry. In the high resolution (3 Å) data set of VSG (see Table 1), 3968 reflections (2106 unique) were collected from a typical crystal of the 13 used in the complete data set in the resolution range from 15 to 2.7 Å, with a symmetry R factor on intensities of 4.2% when 22 reflections were rejected.

Data from native VSG and from a mercurocic iodide isomorphous derivative were measured to 6 Å resolution on the detector system, allowing a difference Patterson to be calculated. Figure 3A shows a comparison of the two Harker sections at z = 1/4 and z = 1/2 (P41212) of this difference Patterson from data sets collected on the detector and on an automated diffractometer (Nicolel P3). The difference Patterson from the detector data sets is readily interpretable in terms of a single heavy atom site and is approximately comparable in signal-to-noise ratio to an identical difference Patterson calculated from diffractometer data. Anomalous differences at 6 Å resolution between Friedel pairs of reflections were also measured both by automated diffractometry and by the detector and found to correlate with a coefficient (11) of 0.54, indicating significant agreement (12). An anomalous difference Patterson (Fig. 3B) calculated from data collected on the detector in the resolution range of 6 to 4 Å showed major peaks corresponding to those in the isomorphous difference Patterson (Fig. 3A), indicating that high resolution anomalous scattering differences were detected accurately.

Data were collected to 7 Å from crystals of the complex of the DNA-binding domain from bacteriophage 434 repressor with a 14-bp DNA operator (13), as well as from isomorphous crystals of such a complex with DNA substituted at one position with 5-BrU (for T). The intensity differences were very small, but threefold noncrystallographic symmetry averaging produced an interpretable difference Patterson and led to a successful solution of the crystal structure (14). The noncrystallographic Harker section is shown in Fig. 4.

Data at 2 Å resolution have also been collected from more conventional, strongly diffracting protein crystals of a chemically cross linked ribonuclease, with the use of a similar detector at Genex (15).

One of us (M. B.) has recently developed an on-line integration system that allows automatic re-refinement of setting parameters during the data collection (16). The program runs on a second microprocessor.

Fig. 4. A difference Patterson section perpendicular to the body diagonal [which is parallel to a noncrystallographic axis of threefold skew symmetry (13)] at Z = 1/3 from the I434 repressor-DNA (see text for details) cocrystal difference Patterson. (A) Unaveraged; (B) averaged about the body diagonal of the Patterson unit cell, which is a noncrystallographic threefold rotation axis.

### REFERENCES AND NOTES

8. Optima optics and corresponding specimen-detector distances are not straightforward to calculate. With a nondivergent beam, such as we can achieve in principle with mirrors, there is an advantage to placing the detector at the same distance from the beam as is consistent with its aperture, the desired resolution, and so forth, since the background will fall off as 1/A, while the diffracted beam will hardly fall off at all. In practice, mirror aberrations give some effective divergence and loss in signal-to-noise than expected.
11. Correlation = $\left(\frac{\sum (I_{data} \times I_{model})}{\sqrt{\sum I_{data} \times \sum I_{model}}}\right)^2$

30 MAY 1986
where $\Delta_{\text{obs}}$ is the Bioket difference $(\text{I}^+ - 1^-)$.

19. We thank Kevin Ulmer and Wes Parker of Genex, Inc., for their help in providing the instrument for this work. We thank colleagues at Genex, R. Salomen, P. Weber, D. Ohtendolr, S. Sheriff, B. Finzel, T. Poulos, and A. Howard, and acknowledge support from NIH CA13202, NIH GM29109 (SCH), NIH AI15684, NIH AI21324 (DCW), and NSF PCM-79-22159 (computing hardware). J.A. is a Burroughs Wellcome Fund Fellow of the Life Sciences Research Foundation. R.M.D. held a Fulbright Maintenance and Travel Award while a special student in biophysics at Harvard.

12 April 1985; accepted 22 January 1986