CHAPTER 3

DATA COLLECTION AND PROCESSING

3.1 Choice of Data Collection Method.

The number of reflections from a crystal which has to be measured to solve a structure to a given resolution is proportional to its unit cell volume, while the intensity of each reflection is inversely proportional. Thus a highly efficient strategy must be used for the virus crystals where the unit cell is so large that each reflection must be measured for a significant fraction of the crystal lifetime in order to count enough photons for an accurate determination. The most efficient method available for this work is the oscillation method (Arndt and Wonacott, 1977). Since high resolution electronic area detectors are not presently available, photography has been used to record the data, and automated densitometry of the resulting films has been employed to read the intensities of the recorded reflections, using the well-established SCAN12 system (Crawford, 1977).

The geometry of the oscillation method is shown in figure 3.1. The crystal is rotated between angles ϕ_1 and ϕ_2 on a spindle perpendicular to the direction of the x-ray beam. The region of reciprocal space which is the volume of revolution of the Ewald sphere between angles ϕ_1 and ϕ_2

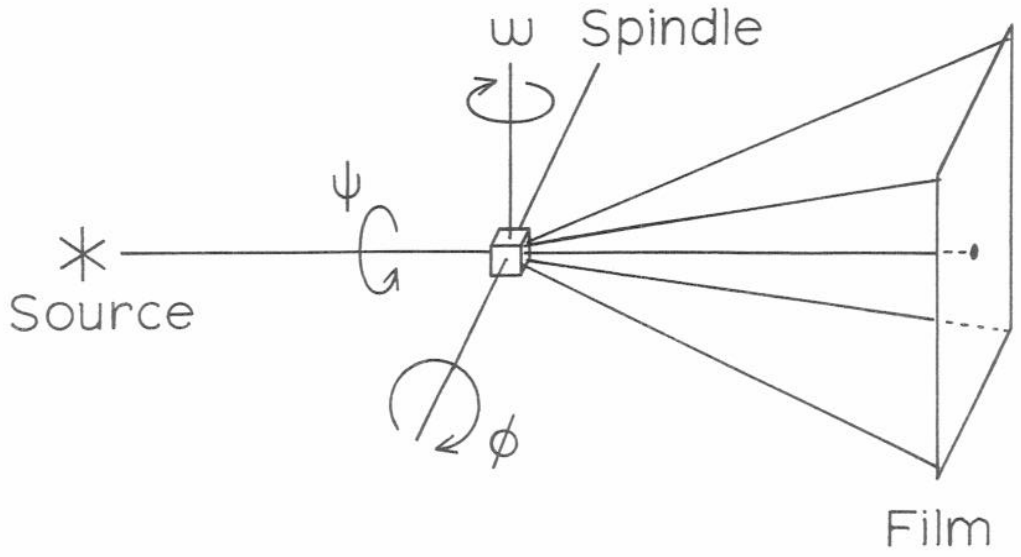


Figure 3.1. Oscillation geometry for data collection. ω and Ψ are the crystal orientational angles corresponding to the goniometer arcs. ϕ is the angle of the crystal on the spindle axis that is made to oscillate during the exposure.

defines the set of reflections which will be recorded. Because of the mosaic spread of the crystals, the non-monochromaticity and imperfectly parallel nature of the x-ray beam, each reflection has an effective angular width, Y. Each reflection is therefore contributing Bragg intensity to its recording for a range of ϕ angles and will only be accurately determined if that range is completely contained between ϕ_1 and ϕ_2; the crystal rotates at constant angular velocity so the integrated intensity over time is the integral over ϕ, but only when this condition is true. Thus the flagging of partially recorded reflections is necessary and is an important feature of the SCAN12 system.

3.1.1 Oscillation Range.

The choice of the optimal oscillation range depends upon many factors. A large range means that each reflection is only contributing intensity for a small fraction of the time, so a longer exposure is required to obtain the same photon count in the spot, which leads to larger background level and correspondingly greater error in the scanned intensity measurement; the crystal lifetime may also limit the maximum exposure time. A small oscillation range means that only a small fraction of the spots is fully recorded; if partial spot weighting is used (see below) the corresponding errors are larger, and if the partials are excluded altogether the overall data collection efficiency is reduced.

In the case of high resolution data collection for compact TBSV (Harrison, et. al., 1978) the oscillation range is restricted by the requirement that the spots in adjacent 'lunes' (corresponding to neighbouring layers of the reciprocal lattice) do not superimpose. $\Delta\phi = 0.5^{o}$ was chosen in this case, but the consequence was that 70% of the reflections recorded on a typical film were partial and it was therefore essential to use a 'postrefinement' scheme (Winkler, et. al., 1979) to recover them. Postrefinement is the determination of the precise crystal orientational parameters, the exact angular width of each reflection ('rocking curve width', Y), and unit cell parameters individually for every film by comparison of the measured partial spot intensities with a reference file of spots known to be wholly determined (i.e. located well away from the edges of their lune on the film). Knowing these parameters, the recorded fraction can be estimated accurately for every spot and used to scale the measured value accordingly. The method works well for the compact TBSV case, but the crystals of the expanded virus have several serious disadvantages:

i) The number of spots obtainable per film is an order of magnitude smaller because of the absence of high resolution data.

ii) The number of unit cell parameters to refine is larger (C2 has four parameters instead of one for I23).

iii) Because of the lower symmetry, the redundancy of reflections in reciprocal space is much lower (C2 has 4 equivalent reflections on the full sphere, whereas I23 has 24).

All of these factors have the effect of reducing the observations-to-parameters ratio for the postrefinement to such an extent that it is doubtful that it would have worked.

Instead, since there were no high resolution data present to cause overlap problems, it was decided to collect 1^{o} oscillation films whence the proportion of partially recorded spots was less than 50%. A 12 hour exposure was found to be satisfactory. Two films were used in the oscillation camera cassette; the second, about threefold attenuated in intensity due to absorption inside the emulsion of the first film, had very few overexposed spots (i.e. that fell in the non-linear region of the response characteristic of the Kodak 'noscreen' medical x-ray film used). The background level on the first film was also satisfactorily low.

3.1.2 Oscillation Strategy.

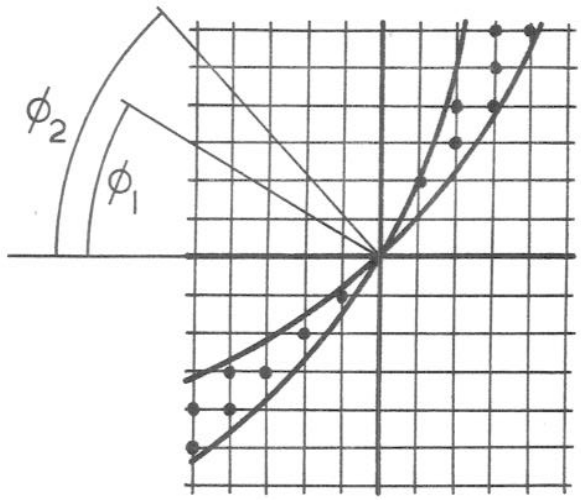
The space group C2 of the crystals of the expanded virus implies that a full quarter sphere of data in reciprocal space should be collected. A single oscillation film samples a 1^{o} spherical sector of reciprocal space that is

slightly curved because of the curvature of the Ewald sphere. Ninety consecutive 1^{o} pictures about an oscillation axis perpendicular to the 2-fold would cover the entire volume as shown in figure 3.2. However this scheme is inadequate in two ways:

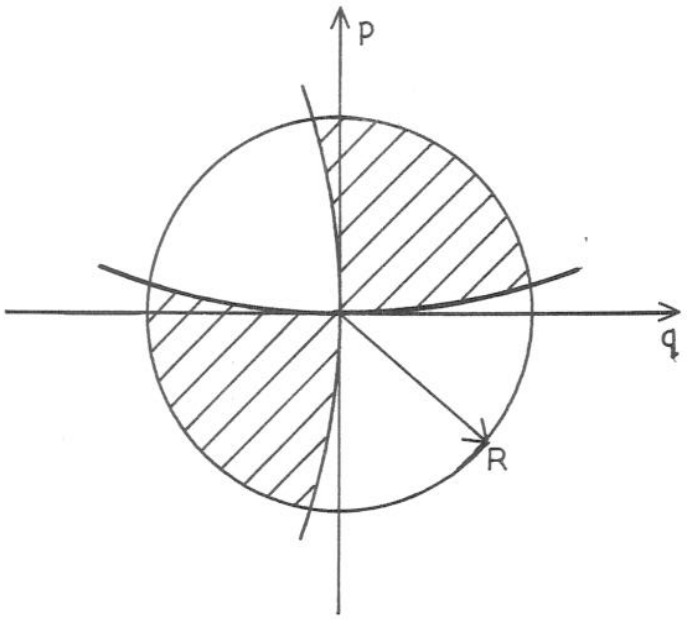
i) The two polar regions of reciprocal space lying close to the axis of oscillation are never observed. There are thus two systematic holes in the data.

ii) The scheme does not provide adequate cross-checking of the data. The procedure for finally merging the films uses reflections that are recorded on more than one film to calculate the agreement of each film with the rest of the dataset (see section 3.4). This is the only point at which erroneous data can be detected and filtered out of the final dataset. Most of the overlap among the films occurs between those with adjacent ø angles (the slight curvature of the Ewald sphere causes the Friedel mate of a reflection to lie on a plane only a few degrees away in ø) and those are frequently collected from the same crystal and scanned in the same way, so that an error in the determination of the polarity of the crystal (see section 3.3.1), for instance, would not be observed as inconsistencies in the data.

A safer oscillation data collection strategy is to include a number of 'cross-films' taken about a perpendicular oscillation axis as well as the basic 90^{o} about the first axis.



(a) The data collected on a film is determined by the Ewald construction. The two positions of the Ewald sphere relative to the reciprocal lattice at the extremes of the oscillation range determine the boundaries of the volume of reciprocal space that is sampled on that film.



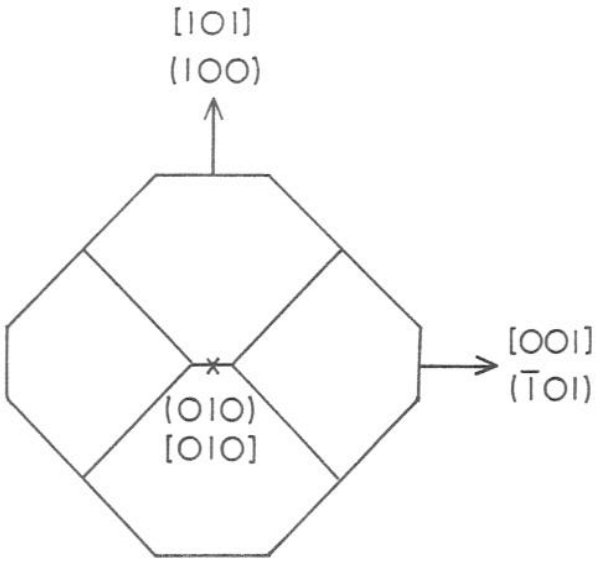
(b) Volume of reciprocal space collected in 90^{o} of oscillation data to resolution limit R. The crystal 2-fold axis must be along p or q, so that the unique volume is one of the four quadrants shown.

Figure 3.2. Collection of the complete unique volume of reciprocal space by means of 90^{o} of adjacent oscillation films.

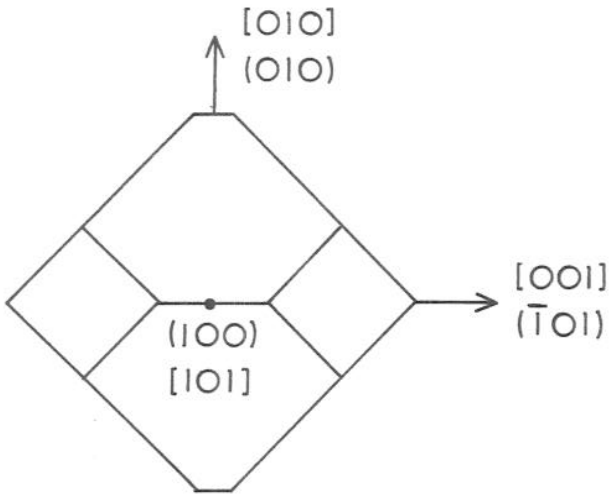
he choice of axis to use for the majority of the data collection was based on the consideration of the crystal morphology. Figure 3.3 shows the most commonly observed shape of the crystals. It was not always possible to pick out the b* axis by inspection; the figure demonstrates the ambiguity. However, use of the $(\bar{1}01)$ axis along the oscillation spindle does not require the ambiguity to be resolved at the time of mounting the crystal, but can be done instead by measuring the radii of the circular patterns in the alignment photographs. This choice has the additional advantage that the crystals, which tend to grow a little longer in the $(\bar{1}01)$ direction, prefer to lie down upon mounting with this axis along the capillary tube, which facilitates the alignment on the goniometer head and reduces the contribution of capillary and solvent scattering to the data photographs. Crystals which did not lie down this way could be used for oscillation about the (100) axis, which is mutually perpendicular to b and c, as can be seen in figure 2.2.

3.2 Data Collection Procedure.

The crystals were harvested into concentrated mother liquor (see chapter 2) in the same temperature-controlled room that the data collection was to take place; they were then never allowed to leave that room and always harvested less than a couple of weeks before their use.



(a) First setting.



(b) Second setting.

Figure 3.3. Crystal orientations for the first and second settings seen from above. The spindle axis runs across the page, and the beam runs up and down. Except for the [010] and (010) case, the zone axis and reciprocal lattice direction are only approximately coincident (see figure 2.2). It is not necessary to distinguish between these settings at the time of mounting the crystal.

Quartz 1.0mm and 1.5mm capillaries (Charles Supper Inc., Natick, Mass.) were flattened by application of a gas-oxygen blowtorch flame (method of D. L. D. Caspar). A crystal was selected and drawn into the capillary using a hypodermic syringe. The excess liquor was drawn off by means of a fine pipette and finally with filter paper, until the crystal was held to the side of the tube by only a slight meniscus. A drop of mother liquor was reintroduced nearby to prevent evaporation from the crystal by saturating the vapour and the tube was sealed at both ends with wax. The crystal was then aligned on a 2-arc goniometer by eye with the aid of a microscope and a model of the ideal crystal morphology; a crystal was usually brought to within 5^{o} of the correct orientation by this method.

The x-ray source used throughout this work was an Elliot model GX6 rotating anode with a 100 micron primary focus. The accelerating voltage was normally 40kV and the beam current was 20mA. 50mm Francks mirrors were used to focus the beam in the vertical and horizontal directions onto the plane of the film (Harrison, 1968). The final spot size on the film was always less than 250 by 300 microns. The camera used was a Supper precession/oscillation camera with a crystal to film spacing of 100mm, so the spots were separated by at least 700 microns. Kodak 'noscreen' x-ray film was used throughout. The developer was Kodak KLX and the developement time was 5 minutes at 68^{o}F. Films were fixed for 5 minutes in Kodak rapid fix, washed for one hour

in running water and air dried.

For the data collection in the 'first' and 'second' settings with ($\bar{1}01$) on the spindle, alignment 'still' photographs were taken alternately down the [010] and [101] zones, each time correcting two of the three orientational angles (spindle setting ϕ, and two goniometer arcs) until all were correct to within 10' of arc; this usually took four phtographs of 30 minutes exposure. From the radii of the circles corresponding to upper layers of the reciprocal lattice, the lattice spacings were calculated and the b axis identified (or confirmed). A 1.5° precession photograph was taken along b to provide the polarity information needed to index the film correctly (see below). A series of data photographs was then commenced.

By convention $\phi = 0^{\circ}$ corresponds to the position with the beam along the [101] zone; $\phi = 90^{\circ}$ corresponds to [010]. Photographs in the first setting were taken with the direction of the beam near to the [101] zone, with $0^{\circ} < \phi < 30^{\circ}$. Those in the second setting had the beam near to [010] and $72^{\circ} < \phi < 90^{\circ}$. No data were collected in the regions $30^{\circ} < \phi < 72^{\circ}$ or $90^{\circ} < \phi < 180^{\circ}$. The choice between the first and second settings was made on the basis of which presented the less solvent scattering path length to the beam. The third setting, referred to above with ($\bar{1}00$) along the spindle, was aligned on [010], conventionally assigned $\phi = 90^{\circ}$, and [100], assigned $\phi = 180^{\circ}$. These are both major

zones of the crystal (see chapter 2). Data were collected in the range $90^{o} < \phi < 112^{o}$.

At the end of every picture, the crystal was translated horizontally to expose a new area to the beam. On a large crystal, 6 sequential photographs could be taken in this way without realignment. Slight drift of the setting parameters by as much as 30 minutes of arc during such a procedure was considered acceptable; any films showing greater changes were considered to demonstrate crystal slippage and were rejected. A second sweep across the crystal could sometimes be made by use of a vertical translation, but realignment was needed between sweeps. In the early attempts at data collection, an additional translation was made after 6 hours of the exposure, because degradation of the higher resolution reflections was known to take place within this time. However, data collected this way proved to be no more accurate when merged together (see below) than when this step was omitted. Data on different films from the same crystal were found to merge better than average, so there was an advantage in having a large number of films per crystal. The 6 hour translation was omitted for the rest of the data collection.

Examples of data films are shown in plate 3.1, and the course of the data collection is summarised in table 3.1. The total time spent was about 9 months for the native data and 2 months for the heavy atom derivatives.

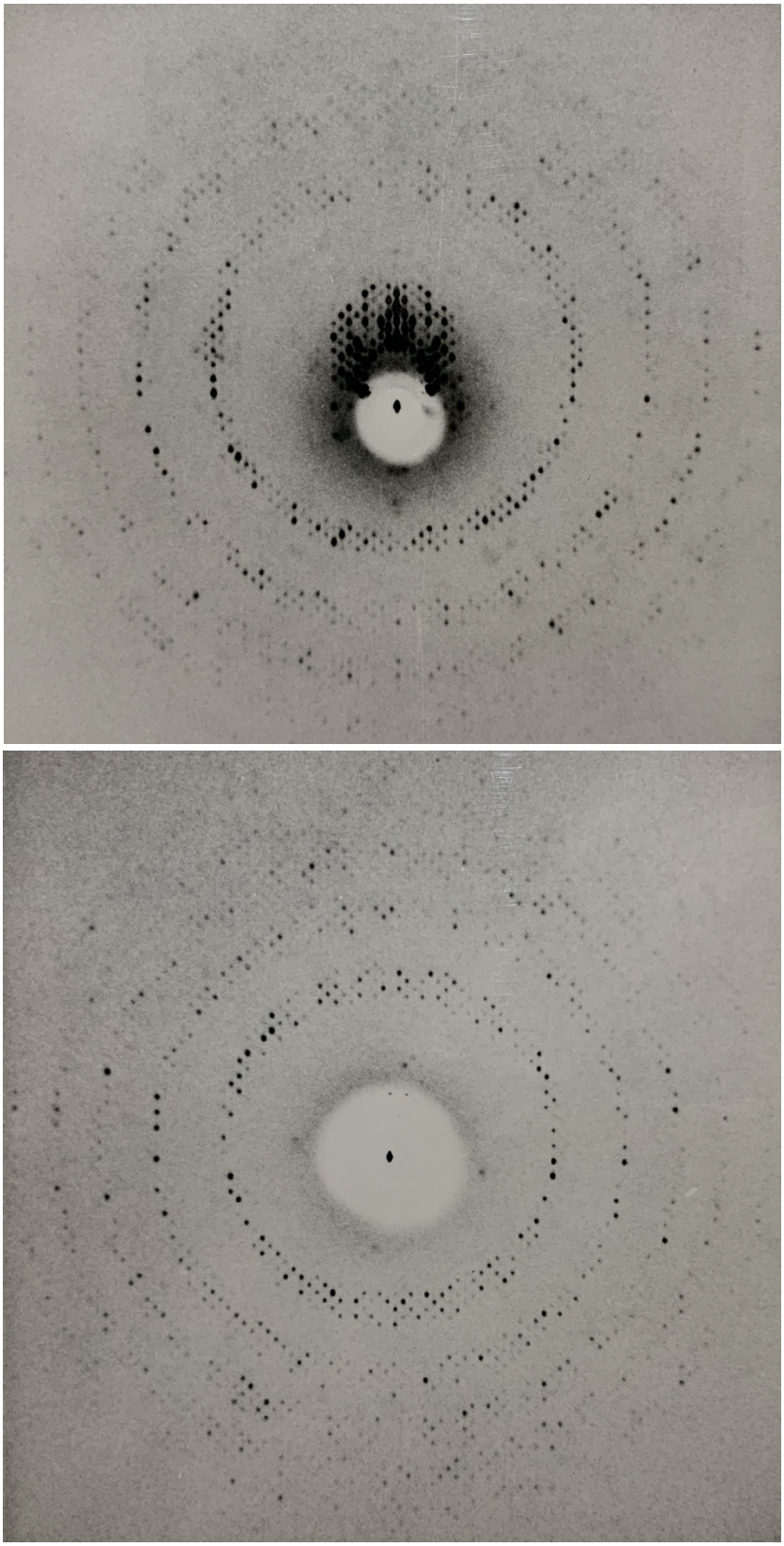


Plate 3.1 Examples of oscillation photographs used in data collection

3.3 Film Scanning.

The Harvard SCAN12 system for automatic densitometry of oscillation films (Crawford, 1977) had been developed and tested for several space groups, but never for a monoclinic one. Space group specific subroutines were required to define the transformation from the reciprocal lattice to an orthogonal local frame used by the program, and also to describe the systematic absences. These were written and easily incorporated into the modular design of the SCAN12 system.

3.3.1 Crystal Polarity.

Somewhat more difficulty was encountered with the identification of the polarity of each film as recognised by the scan program. The space group C2 is polar, meaning that the axes of the crystal have directionality. A photograph taken at a particular ø setting is the mirror image (about a vertical mirror, perpendicular to the oscillation axis) of one taken at a setting of ø+180°. No information is provided about the polarity of a crystal by the alignment photographs alone; the 180° photograph appears the same as the 0° one. For this reason, a small-angle precession picture was taken for every crystal around the b zone axis and, by comparison with a standard photograph, all films for that crystal were labelled either 'o' (ordinary) or 'f' (flip). Care was taken not to introduce a 180° rotation between the

alignment ø setting and the one used for data collection. The 'flip' of each film then determined whether it was turned face up or face down during the indexing and scanning operations.

The orientation of the crystal is entered into the scan program as three angles and an orientation code 'IOR' which specifies which of six allowed permutations (IOR = 1,...6) of the axes is to be taken (Crawford, 1977). Therefore, in principle, there are six equivalent ways of defining the same crystal orientation. Because of certain peculiarities in the algorithm for predicting the set of spots that will be present on a given film (program 'OSCGEN' in the SCAN12 system) when used with non-orthogonal space groups, not all of these ways were found to produce the full set of spots; several equivalent definitions of the orientation had to be tried for some films until the expected number of spots was generated. To add to the confusion, orientation codes IOR = 2, 4 and 6 are left-handed axis permutations, which meant that one additional film flip was required for scanning films using these codes.

Apart from one or two other minor amendments to the code of the SCAN12 programs, no further difficulties were discovered in the scanning of the expanded TBSV data.

3.3.2 Masks and Wedge.

Before scanning was commenced, a "mask" was created to define the shape and size of the spot (Winkler, et. al., 1979). Several spots from different areas of a trial film were densitometered and printed out. The central region of the mask over which the optical density values are added to give an integrated intensity was chosen to be one 50 micron raster unit larger in all directions than the largest spot image encountered. The outlying regions that are used to calculate the background optical density and standard deviation were placed in the four directions of greatest distance to neighbouring spots and were trimmed in size so that in no case could they ever include part of another spot, even if the mask was miscentred. The final mask contained 68 raster points in the spot and 50 in the background, and is shown in figure 3.4.

The "wedge" function which describes the photon count to optical density transfer function of the Kodak 'noscreen' film and the local development and fixing conditions had already been measured by J. Hogle; this was used without modification. The function is very close to linear up to an optical density of 1.7 then steepens to account for the saturation of the film, becoming more than twice as steep by O.D. 2.0. Spots containing O.D.s greater than 2.0 were rejected by the scanner program; in almost all cases, the exposure was such that these were not saturated on the

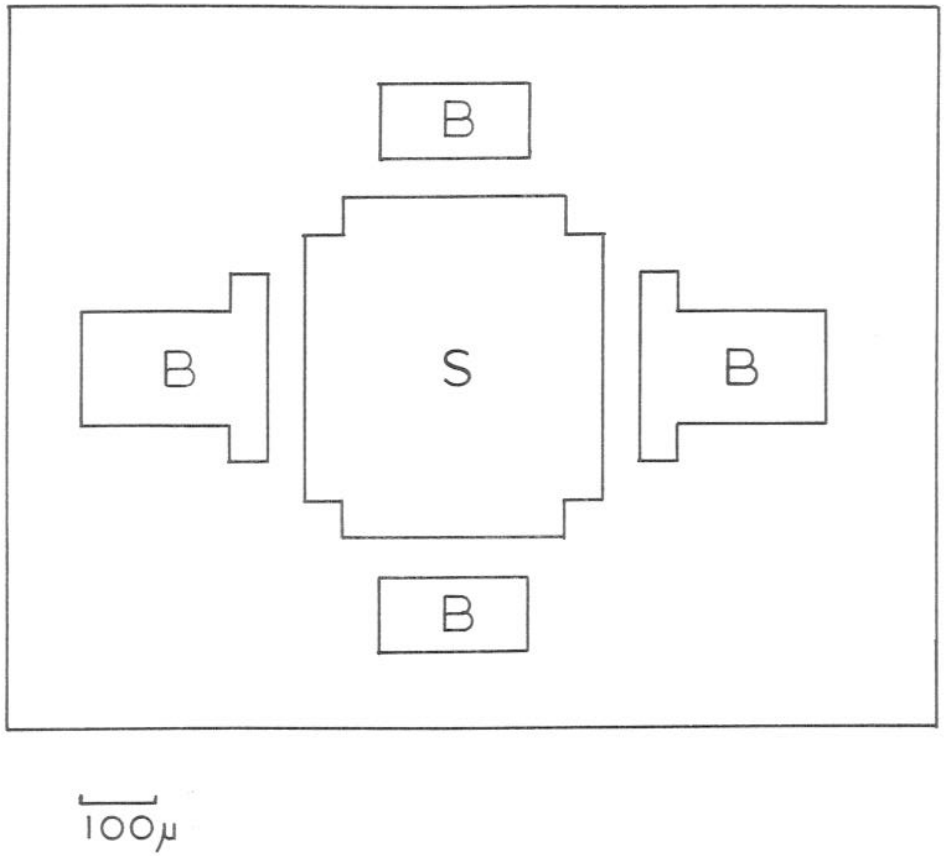


Figure 3.4. Mask used for scanning data films for expanded TBSV. This is defined on a 50 micron grid corresponding to the scanning raster. The optical density of the spot is integrated over the region marked 'S', and the background over the regions marked 'B'.

second film in the cassette and so were not lost altogether.

3.3.3 Scanning Procedure.

The first step in scanning a film was the manual indexing of a set of 12 partial spots used to define the exact orientation of the crystal during the exposure. Since all the data collected were within 30^{o} of one or other of the alignment directions (which are the major zones of the crystal), it was always possible to count the lunes corresponding to the layers of the reciprocal lattice perpendicular to these two major zone axes. A spot is thus assigned a 'lune number', n, by counting layers from the origin. For spots close to the centre of the film, the pincushion distortion was also small, so the position of a spot could be described as discrete (i,j) coordinates on an orthogonal grid (the unit cell angle β is very close to 135^{o} and $a \simeq 2^{1/2}c$, so the b* projection is very close to being a square lattice). Simple formulae were derived to convert (i,j,n) into (h,k,l) for each of the three crystal settings used.

Partial spots were selected on the grounds of their having a clipped appearance or a slight displacement from the line connecting their horizontal or vertical neighbours. Twelve spots were selected, distributed in all four quadrants of the film, six at each end of the oscillation range. Their recorded fraction was estimated very crudely.

The indices were fed into the 'halfspot' program of SCAN12 which produced the setting angles and an error estimate for each spot. Typically, one or two spots would show a large error, sometimes due to an arithmetic, counting or typing mistake; otherwise, the spot was replaced with a nearby one. The halfspot refinement was considered acceptable when the standard deviations of the best estimates of the setting angles were below 2 minutes of arc.

When the refined setting parameters were known, the film was mounted on the scanner. The scan procedure was automatically controlled by a command file, which was successful about 70% of the time; otherwise manual intervention was used. First the centre spot of the film marking the direct beam direction was located, then the angle of twist about the centre was determined by sequential refinement of the positions of strong spots lying at increasing radii. Finally, the film was scanned and the measured intensities and standard deviations were stored on magnetic tape. The front and back films were mounted on the scanner side by side and scanned together.

The times taken to index a film and to scan both films, were about 20 minutes and one hour respectively. A typical film yielded 2,700 spots to 7.0Å resolution, of which about 1,200 were partially recorded.

3.4 Data Processing.

The scanner output is in the form $(h, k, l, I_{meas}, \sigma_{meas})$ records in files on magnetic tape, one file per film scanned. To obtain a useable dataset, these files had to be scaled and merged together. The data processing method used was that devised by C. Schutt, F. Winkler and S. Harrison in this laboratory (Schutt, 1976) and made operational on the department VAX 11/780 by P. Kuttner and J. Hogle; the scheme is summarised in figure 3.5.

i) PRUNE and PREPRO. Each film was processed by these programs to remove unwanted reflections and provide statistical information. Firstly, measurements assigned a negative standard deviation by SCAN12 (because their centres were too far from their expected positions or because they were too dark) were discarded. Secondly, only positive measured intensities were retained on all the back films for purposes of front/back scaling; negative front film measurements were however retained. Thirdly, a resolution cutoff was made to each film, using individual limits for each film scanned (see section 3.4.1).

ii) SCALE. Every reflection that was present on both front and back films was saved for calculation of an overall scale factor by the method of Fox and Holmes (1966), correcting for the variation of this with the angle of incidence on the film. Back film reflections were then

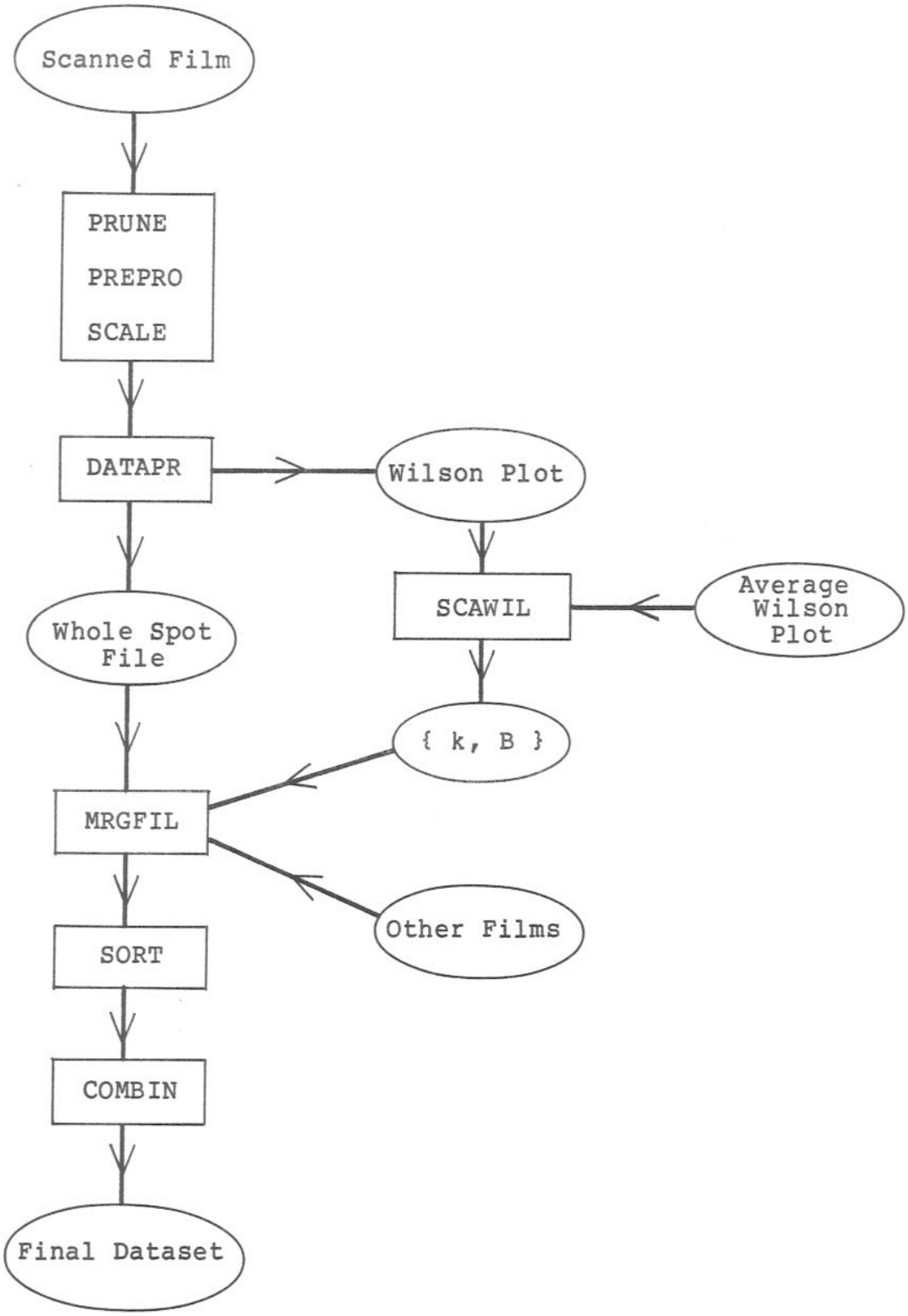


Figure 3.5. Flow diagram for the processing of the expanded TBSV data.

multiplied by this factor and merged with those from the front film, and an R-factor was calculated for the agreement between them. All the scale factors were found to be in the range 3.0 to 3.3, and the R-factors below 0.10.

iii) DATAPR. From the known setting angles of the film, the rocking curve width, and unit cell constants, the 'degree of partiallity' of every reflection was calculated. Only those reflections with a recorded fraction greater than 105% (allowing a safety margin for error in the input parameters) were retained as whole spots.

iv) SCAWIL. A 'Wilson plot' consisting of a list of the average intensities in each of six resolution ranges between 35 and 7.4 Ångstroms was calculated for every film by DATAPR. The first 30 of these plots were optimally scaled together and their average taken. Each film was then fitted individually to this average Wilson plot by choice of a scale factor, k, and a temperature factor, B, with the relation,

$$I_{obs} = kI_{meas}\, e^{-2B\sin^2\theta/\lambda^2} \qquad (3.1)$$

where θ is the Bragg scattering angle for the reflection, and λ the x-ray wavelength (Wilson, 1949).

v) MRGFIL and COMBIN. The I_{obs} were calculated from the I_{meas} by application of k and B as in equation 3.1. These data were tagged with a label indicating from which film they were obtained, merged together and

sorted into ascending order of (h, k, l). Statistics were calculated on the quality of the I_{obs} by comparing repeated measurements of the same reflection that appeared on more than one film. A reproducibility coefficient was derived in the form of a conventional R-factor on intensities and was broken down in resolution ranges as well as film by film. The bad films could therefore be isolated as being those with anomalously large R-factors.

This data processing scheme was used with minor modifications in the form of space group specific subroutines for DATAPR. The expanded TBSV as well as all derivative data were processed identically.

3.4.1 Individual Resolution Cutoffs.

Although all the films were scanned from 35 to 7.0 Ångstroms resolution, not all of them contained this full range of data. Some had a beam-stop shadow extending beyond 35 Ångstroms, while others contained no visible spots beyond some limit lower in resolution than 7 Ångstroms, even a few with nothing beyond 9 Ångstroms. The eye was considered to be a reliable way of determining the diffraction limit for a film and so these limits were recorded individually by inspection. The cutoffs were given as input parameters to PRUNE to reject the unwanted reflections. The extra effort was found to pay off in terms of the final R-factor.

3.4.2 Estimation of Rocking Curve Width.

Without postrefinement, there was no method for determining the rocking curve width, γ, for each crystal. To obtain a crude estimate for its value, the following experiment was performed. Eleven adjacent films from a single crystal were processed using several different trial values of γ and the final R-factors for merging were compared. By using only one crystal, good agreement between films is obtained and, since γ is expected to be the same for each film, a knee in the response should be visible. The results are shown in table 3.2. No strong variation is observed but a value of γ = .001 ensures that the reproducibility is as good as for the largest γ value tried, so this was settled upon for subsequent processing.

3.4.3 Data Processing Results.

Of 2,700 measured reflections typically on a front film, 2,400 would remain after the PRUNE and PREPRO steps, depending on the severity of the resolution cutoff used. This number reduced to 1,400 after rejection of partial spots. 1,400 is only 1.5% of all data to 7.0 Ångstroms resolution.

A total of 128 1^{o} oscillation photographs from 22 crystals were taken of which 92 were suitable for scanning. These films were merged after processing and all those with an R-factor greater than 0.30 were rejected. The final

Gamma	# Refls. Output	# Redundant Reflections	R-factor
.0025	8,223	460	0.139
.0015	10,649	739	0.135
.0010	12,089	942	0.140
.0007	12,970	1,106	0.142

Table 3.2. Variation of data processing performance of the data for crystal X116 (11 films) as a function of rocking curve width, γ. The advantage of the increased number of reflections output is offset by their lower accuracy (higher R-factor) below γ = .001.

dataset contained 58 films and had an R-factor of 0.196 with individual values distributed uniformly from 0.1 to 0.3. The Wilson plot is shown in figure 3.6, where it is compared with that of compact TBSV. The superexponential falloff, steeper than that of the native, is clearly apparent, suggesting that this is fairly well characterised as thermal disorder.

Heavy atom data collected from different crystals were <u>not</u> merged because of the expected variation in soak conditions from one crystal to the next. Fortunately, large crystals had become available by the time the heavy atom data were needed and sizeable datasets could be collected from each. Two $PtCl_4$ and two EMTS crystals were used and the four merged files were kept separate. The procedure was exactly the same as for the native dataset, using the same average Wilson plot and again rejecting films with an R-factor greater than 0.30. The largest overall R was 0.20. When a merge of the two $PtCl_4$ datasets was attempted, the R-factor increased from 0.16 to 0.28, justifying the decision to treat the crystals separately. The data collection statistics are summarised in table 3.1.

3.4.4 Selection of Films for Final Merge.

The rejection of films on the basis of large R-factor deserves some justification. It is believed that the greatest source of error in the whole data collection pro-

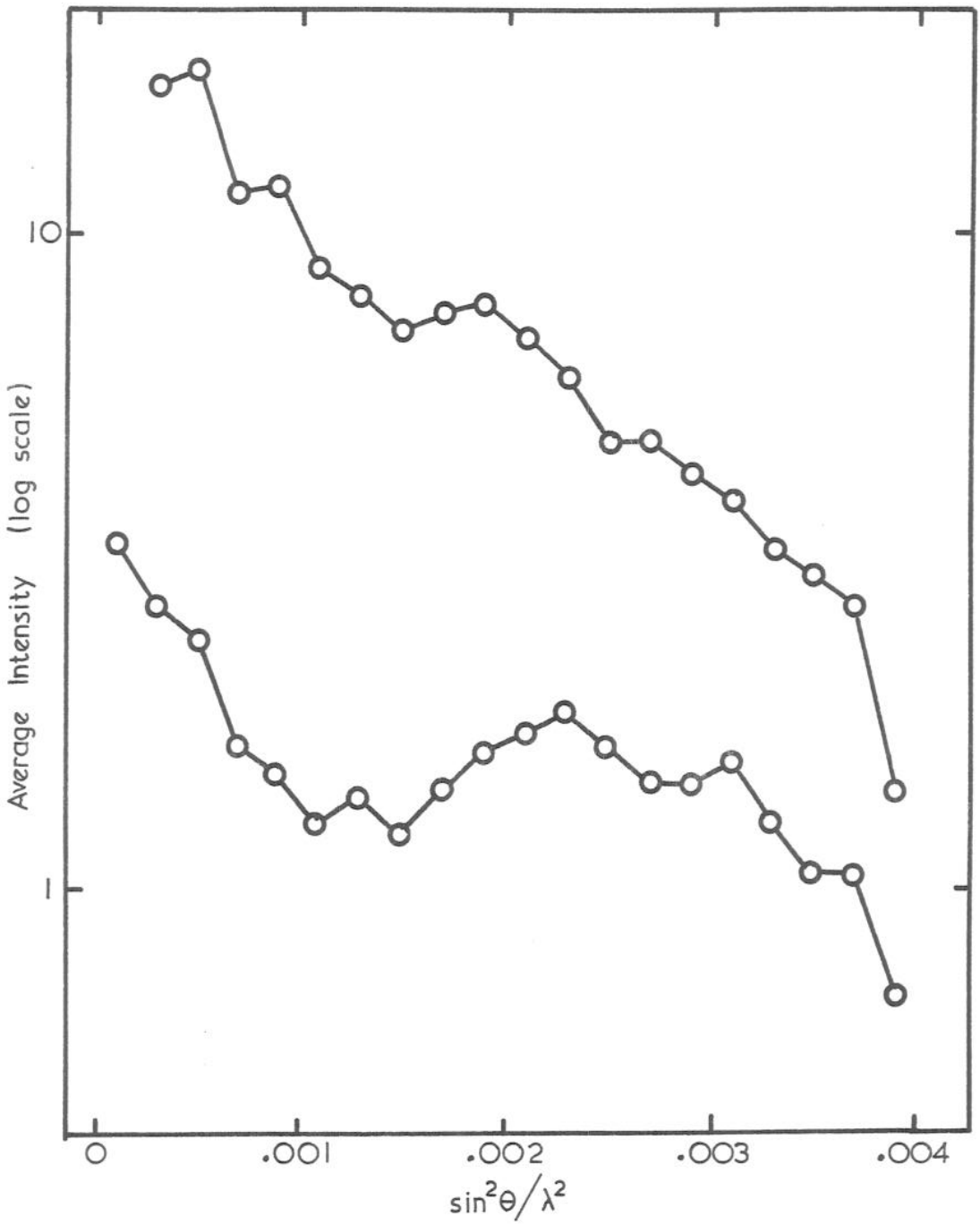


Figure 3.6. Wilson plot of the combined native dataset for expanded TBSV (top), compared with that of the compact virus (bottom; Winkler, et. al., 1977). The temperature factor relating the two slopes is 340Å^2, indicating substantial thermal disorder in the former.

(a) Number of films used at each stage.

Dataset	# Crystals	# Films Collected	# Films Scanned	# in Final Merge
Native	22	128	92	58
X131 (Pt)	1	12	12	8
X132 (Pt)	1	11	11	6
X140 (Hg)	1	16	16	14
X144 (Hg)	1	15	14	13

(b) Number of reflections and overall quality.

Dataset	# Reflections (final)	Fraction of 7Å Sphere	R-factor
Native	40,825	48%	0.196
X131 (Pt)	5,350	6%	0.158
X132 (Pt)	4,912	5%	0.156
X140 (Hg)	14,220	16%	0.195
X144 (Hg)	11,346	13%	0.200

Table 3.1. Data collection statistics. The heavy atom datasets are named after the crystals they were collected from.

cedure is crystal slippage during the exposure; this is partly due to the mechanical effects of radiation damage. Significant trends were observed in the refined setting angles when successive exposures were taken without realignment; angles were seen to drift by as much as 30 minutes of arc over several exposures. Any drift during an exposure would produce a number of spots near the lune edge that would be systematically underestimated in intensity. With some films the effect was so great that considerable difficulty was encountered in finding suitable partial spots for input to the scan program; spots originally estimated as being 60% present were then artificially reduced to 20% to give a satisfactory fit. Another source of error is the possible gross misscanning of the film by wrong centering or accidental wrong flip; these would both be expected to produce R-factors much greater than 0.30. Thus the sources of error fall into two distinct categories: inaccurate intensity measurement and false indexing. The second of these must be excluded altogether while the first is a part of the unavoidable experimental error. The selection of 0.30 as the rejection criterion in R-factor was based on the fact that it was the natural edge of the distribution: nearly all of the rejected films had R-factors considerably greater than 0.30 (and as large as 0.9) so are probably in the second category. In a few cases, rescanning a film led to a dramatic improvement in its agreement with the rest of the data.

The final quality of the native dataset is described in table 3.3(a). The R-factor for symmetry equivalent reflections is 0.12 on intensities for the lowest resolution range, which is as good as expected for this data collection and processing scheme (Harrison, et. al., 1978), but the quality falls at higher resolution. This is partly due to greater relative errors in the weaker higher resolution reflections, but also due to the inconsistency of the diffraction limit from crystal to crystal, it is believed. This trend is only partly accounted for by an individual temperature factor for each film. Table 3.3(b) shows the fraction of the full quadrant of reciprocal space that is covered by the final native dataset to different resolutions.

The five output files were preserved in the form of (h, k, l, I_{obs}, σ_{obs}) records. The negative observed intensities were retained as these are also valid measurements; although the best estimate of the corresponding F_{obs} is zero, this is still a constraint for the testing of model structures.

(a) Breakdown of native data by resolution.

Resolution Range	Number of Redundancies	Average Intensity	R_{sym}
35.0 - 12.5	7,847	255	0.118
12.5 - 10.0	7,516	170	0.197
10.0 - 8.8	6,476	109	0.332
8.8 - 8.0	3,321	67	0.485
8.0 - 7.4	688	40	0.575

(b) Number of native reflections to different resolutions.

Resolution Limit	# in Full Quadrant	# Observed Reflections	Fraction
7Å	90,209	40,819	45%
8Å	60,193	36,113	60%
12Å	17,323	10,356	60%
16Å	6,888	3,880	56%

Table 3.3. Breakdown of native data after processing. The reproducibility of the data is indicated for the redundant measurements by R_{sym}.