DARESBURY LABORATORY

INFORMATION QUARTERLY for PROTEIN CRYSTALLOGRAPHY

An Informal Newsletter associated with Collaborative Computational Project No. 4 on Protein Crystallography

Number 6 May 1981

Contents

Editor: Pella Machin
Science Research Council, Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England

Deputy Editor for Imperial College: Dr. Alan Wonacott
Imperial College of Science & Technology, The Blackett Laboratory, Prince Consort Road, London SW7 2BZ

Deputy Editor for Birkbeck College: Dr. David Moss
Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX

ACTIVITIES AT YORK

Guy and Eleanor Dodson

The trouble about asking for a report from York for the CCP is that we are too busy collaborating on computing projects. But Pella's pressure has persuaded us and we have decided on an account of our work which describes what programs and methods we have been using and which might be generally interesting and helpful.

Collaborations - these have happened in quite unplanned ways and have made our lives much more fun and our studies very productive. There is the collaboration with Zygmunt Derewenda (from Lodz in Poland) on the haemoglobin crystallised from polyethylene glycol (PEG). The crystals are similar to other deoxy haemoglobin crystals prepared in Werner Love's laboratory, but are grown in air. Zygmunt demonstrated the crystals could be liganded without damage and apparently oxygenated without oxidation. In this study 2.1Å data has been collected at LURE, the film developed in York, and processed at Imperial College. The film data was reduced, scaled and merged and all the subsequent crystallographic calculation (rotation and R value minimization, refinement of the Hb model atomic parameters by FFT procedures and Fourier maps) done on the DEC 10 at York. And all in 6 months. This was possible because of the generous help of the laboratories mentioned already and because of Zygmunt's acceptance of the non-Polish work ethic. He found the facilities he needed (however much used - including the saturated DEC 10 at York) were entirely his at hours like 4 a.m.!

Although the PEG Hb maps are most interesting and Zygmunt returned to Poland ecstatic about the effectiveness of the protein crystallographic community there were several worrying episodes - leaving out the appalling behaviour of the Home Office and Immigration Department, and Zygmunt's immersion in the icy lake on the way to a late night calculation. There was the appearance of marked diffuse scattering on the films exposed at LURE, which we had not detected before. To our inexperienced eyes the films looked unlikely to give accurate data. Here Alan Wonacott's expertise and flexible film processing procedures came to the rescue. There was anxiety too when the disparity between partial and fully recorded reflections was found to range between 10 - 25%

amongst the different crystals; possibly owing to an overly simple treatment of the peak shape. We simply gritted our teeth and applied fudge-factors - with Zygmunt's time running out it wasn't possible to reprocess the films more rigorously. The DEC 10 gagged more than once while the 195,000 observations were being handled but we were saved by the Christmas lull.

We were glad that in the data processing we held on to all the observations even though those beyond 2.3Å seemed very weak. The crystallographic refinement clearly benefitted from these high angle terms which were found to be usefully accurate. The refinement itself was a pretty large problem for the DEC 10 - there were 5,000 atoms and 37,000 terms. It took 55 minutes per cycle usually calculated between 3 and 6.30 a.m. For us, used to puny (E.J.D.) insulin these calculations had a new dimension, but, once a fundamental error had been spotted (well done John Campbell) convergence was smooth. The starting positions (Fermi's deoxy haemoglobin coordinates properly oriented in the PEG Hb cell) had an agreement factor R of 42%; after some 6 cycles this fell to 28% for a geometrically correct structure.

A second collaborative project is the study on bacterial ribonuclease which has an interesting history. The enzyme and crystals were prepared by Bob Hartley at the NIH. Bob brought them to Cambridge in the early 1970's (we understand) where David Blow, Cyrus Chothia and Gerard Bricogne and most of all, Tony Jack worked on them. The crystals, space group $P3_1$ ($P3_2$) had 3 molecules in the asymmetric unit. Analysis of the one excellent heavy atom derivative (a gold salt) indicated the molecules were not related by any simple symmetry. A map calculated from the isomorphous phases which had been refined by symmetry averaging, was uninterpretable. A conversation, we gather between Cyrus Chothia (still interested in the structure) and Phil Evans, led to the old paper tapes and some freshly prepared crystals coming to York. Here a series of individuals helped: Federico Gordiano from Naples, Peter Moody (now at Imperial College) reprocessed the original data from the grimy (EJD) tapes - taking great care with the anomalous scattering data and with the indices!

Bob Hartley sent some crystals in which the enzyme had been specifically iodinated; 2.4Å data from these were collected at LURE (by Yves Mauguan who has played a very large part in the analysis).

The heavy atom derivative was refined by the familiar (to us) FHLE procedure and the iodine atoms were placed from different Fourier maps phased by the isomorphous and anomalous differences of the gold derivative - a procedure that also defined the space group as $P3_2$.

The electron density map calculated from the two derivatives contained encouraging features - reasonable boundaries for the 3 molecules, reasonably similar density for the 3 molecules, well defined aromatic sidechains and pieces of peptide chain but for all that it was not interpretable. At this stage two of the originals (Gerard Bricogne and Cyrus Chothia) and ourselves with Yves Mauguan regrouped and we spent an incredibly intense week at Cambridge discussing how to use Gerard's symmetry averaging procedures. When we left the MRC, computing access was completely saturated and we were exhausted. Back in Paris Yves completed the symmetry refinement calculations, came to York and computed a beautiful map, readily interpreted. The coordinates are now being tidied up and we plan to extend the native data to high resolution so they can be refined. Bob Hartley is bringing fresh crystals in June for this operation.

Insulin just goes on and on, keeping us thinking crystallographically and structurally. Colin Reynolds has just completed the refinement of hagfish insulin with 1.9Å data and has refined (judiciously) the A_1 α-amine and B_{29} ε-amino crosslinked insulin whose data extends only to 3.0Å spacing.

We have worked particularly hard on crystals of an insulin modified by removing its 5 B chain C terminal residues. This insulin no longer aggregates and hence gives us a chance to look at the monomeric structure of the molecule. Four related crystal forms have been grown in York and one in Peking. So far Federico Gordiano, John and Sue Cutfield from Dunedin in New Zealand and Bi Ru Chang (who coming from Peking brought some of their X-ray data with him) have worked on this problem. The crystals contain unusually little solvent (about 20%), are small and appear a little unpredictably - but they diffract well. Sue Cutfield's expertise in crystal growing has been much appreciated in this particular project. We have failed to get suitable heavy atom substitutions (so far - S.C.). Our approach has been to use the rotation function and R factor searches based on a suitably tailored insulin molecule - although we were aware these crystals were not good candidates for such studies. And we were right to worry; the correct peaks turn out to be about 10% above the background. The contrast with the PEG Hb where we did similar calculations

could not be more marked. Nonetheless after a nightmare of symmetry operations we were able to select a consistent set of peaks. The difference Fourier maps based on the appropriately tailored and positioned insulin molecule showed well defined and sensible electron density at the positions of certain selected side-chains when omitted from the phasing, which encourages us a lot.

It must be obvious that we have been greatly helped over the last two years by being able to use facilities and skills at other centres. To these groups we are grateful - and would remind those who might be interested that we can offer hospitality and some experience in crystallographic refinement if anybody needs it.

John and Sue Cutfield have been in York on Sabbatical leave for the last year. Their presence has added to the momentum on the insulin studies - especially the writing up of the 2 Zn insulin refinement ! and the growing of some very nice recombinant DNA human insulin crystals from material sent to us by Eli Lilly. We look forward to another sabbatical visitor next August, Ed Hough now at Trømsø, originally from Imperial College; we are assuming York University will still be in existence by the time he comes.

GROUP MEETING OF THE NORTHERN PROTEIN CRYSTALLOGRAPHERS

Maxine McCall (Leeds)

A one-day meeting for protein crystallographers from Daresbury, Keele, Leeds, Sheffield and York was held in the Biophysics Department of Leeds University on 15 April 1981. It was the second in a series of meetings which has been planned to promote informal contact, communication and collaboration between the represented groups.

Mike Elder (Daresbury) reported on the Film Scanning meeting at Oxford and John Helliwell (Keele) gave a progress report on the Enraf-Nonius Fast TV Area Detector for Daresbury. Phil Bourne (Sheffield) commented on the experiences of a small group computing with the CCP; Elias Eliopoulos (Leeds) talked on secondary structure prediction methods, and Guy Dodson on mutant human insulin. The benefits from the collaboration which already exists between various combinations of the groups were evident from these talks.

The next meeting will be held at Leeds University on Tuesday, September 29.

OXFORD FILM PROCESSING MEETING: MARCH 19 - 20 1981

Mike Elder

Introduction

Keith Wilson's parting gift to CCP4, shortly before his departure for Berlin, was the organization of a successful 2-day meeting on the processing of rotation photographs. This meeting was held in Oxford on March 19 and 20. Participation was limited as far as possible to those who were actively involved in protein film processing: there were 21 participants including at least one representative from each of the UK protein groups together with representatives from the Munich, Groningen and EMBL Hamburg groups. Keith's efforts in organizing this meeting on behalf of the Collaborative Project are gratefully acknowledged.

The principal aim of the meeting was to bring the participants up-to-date with developments in film processing and in the various solutions to problems which have been incorporated in the different program systems that are available in the UK. The meeting also served a secondary purpose of bringing together all those who are involved in the comparative study on film processing using the ferritin data set which will be presented at the Ottawa IUCr meeting.

This necessarily brief summary of the meeting follows the major headings: film and microdensitometers, data collection strategy, crystal orientation, post refinement, film orientation and integration, merging and data reduction, systematic errors.

X-ray Film

Although some groups wisely stock-piled quantities of now unavailable film it is clear that the number of alternative sources of suitably fast X-ray film have been greatly reduced. Kodak No-screen (USA), Agfa M3 (W. Germany) and CEA Reflex 25 (Sweden), the latter only available from Sweden and at an excessive price, are the main alternatives. Some preliminary results from the survey of film characteristics, obtained using Ni-filtered fluorescent Cu $K\alpha$ radiation, were quoted (M Elder) and are summarized in Table 1. The Groningen apparatus for producing step wedges using graphite monochromatized, Ni-filtered, Cu $K\alpha$ radiation was described (R Kampuis). Results from this work are quoted in the table in parentheses for comparative purposes, for films which are currently available.

Table 1 - Film Characteristics

Film Type	Manufacturer	Speed (Ilford G=1)	Ag (mg cm^{-2})	Fog (O.D)	Film Factor	Price (100 shts)
No Screen	Kodak (USA)	1.50	4.0 (4.0)	.54	3.4 (3.7)	£51
Osray M3	Agfa (W. Germany)	0.77	1.9 (1.8)	.29	1.8 (1.8)	£43
Reflex 25	CEA Werken (Sweden)	1.24	2.9 (3.0)	.22	2.4 (2.5)	£160
Singul X	CEA Werken (Sweden)	0.73	1.2	.33	1.5	£40

Kodak No-screen is fast but has high fog. Agfa M3 has lower fog but is slower and has a low film factor. CEA Reflex 25 has perhaps the most satisfactory characteristics but is extremely expensive. In subsequent discussion it emerged that the variation in the fog level of No-Screen with time and storage conditions was well-known, and that lower fog levels than those quoted are possible from fresh batches of film.

Microdensitometers

In essence the choice of microdensitometer reduces to a choice between the updated Optronics Photoscan instrument and the newer Joyce-Loebl Scandig-3. Apart from the hybrid machine at Cambridge there were no other machines known to the participants to be in use by protein crystallographers. The principal differences between the two machines were summarized (A Wonacott).

The Scandig is the faster machine with its data transmission rate dependent upon a drum speed of 12 revs/sec compared with the Optronics at 8 revs/sec (both figures should be halved for 50 μm operation) but this advantage of the Scandig is often thrown away by software which wastes alternate revolutions which are spent transferring data to magnetic tape. The Scandig uses a photodiode detector system which should give better linearity and long-term stability than the Optronics photomultiplier. The Scandig light source is cheap but very short-lived whilst the Optronics light source lasts much longer but costs more.

A number of differences arise from the Scandig's use of a plastic sleeve for film mounting, rather than the air gap of the Optronics. The sleeve is to be preferred for ease of film mounting, and for ensuring that films are mounted

with a minimum of distortion. It is not liked for the problems of data accuracy which arise with the need to keep the sleeve clean and unscratched, and for the fact that a poor autozero reading (arising from a mark on the sleeve) will affect all the data from one drum revolution whereas a similar problem should not arise with the air gap used for autozero with the Optronics.

The Imperial College (Scandig) experience with the effects of vibration (due to movement of the optical carriage along the drum axis between scans) upon instrumental noise indicate that noise can be reduced significantly if the scanner is modified to make this movement immediately after the transfer of data ends during each revolution, rather than immediately before the autozero reading is taken.

It was observed that both Optronics and Scandig machines have warm-up problems and that drift can be avoided by allowing an adequate period after switching on before data are measured.

Data Collection Strategy

The principal considerations in deciding upon a data collection strategy were summarized (A Bloomer). It was often necessary to resolve a conflict between opposing factors in terms of the particular problem: there are no generally applicable recipes. The obvious points to consider are:

Crystal life-time, affecting the number of films which can be collected with one crystal.

Crystal habit and orientation in the solvent tube will often limit the possible rotation axes.

Unit cell dimensions are important - rotation about the long axis may allow longer rotation ranges and hence fewer films, but longer ranges give more background.

The measurement of Bijvoet pairs is more accurate if they occur on the same film. This is important for derivatives but not for native or substrate crystals.

Choice of resolution: straight to high resolution or staged? If only low resolution data are required then exposure times may be reduced to avoid saturation.

Different film types can be mixed in a pack with a sensitive film at the front to pick up weak reflections, then less sensitive films to increase the dynamic range.

For an orthorhombic crystal a total of 90° rotation about an axis (c) will give Bijvoet pairs on the one film, 4 measurements of most reflections, but lose some reflections in the cusp-shaped blind region. The same total rotation about a diagonal ([110]) will also give Bijvoet pairs on the one film, 4 measurements of some relfections, 2 of the rest, and no loss owing to the cusp.

An alternative strategy allows almost all reflections to be measured with 60° total rotation (say $10 - 40^{\circ}$, $50 - 80^{\circ}$) and then, if the derivatives seem OK and structure solution is possible, the complete dataset with full redundancy may be obtained by filling in the gaps. A similar method (W Bennett) employs alternate 10° ranges.

It was observed (K Bartels) that many of the considerations about data collection strategy are based upon reducing the number of films to save money and processing, but that these should not necessarily be paramount when using a synchrotron source. The great cost of a shift of beam time should mean that considerations of saving beam time should be paramount.

Crystal Orientation

The use of the screen-less rotation method for large unit cells requires that a series of photographs be taken using small oscillation ranges to avoid reflection overlap. These small ranges and the finite period that each spot spends in a diffracting position due to beam divergence and crystal mosaicity mean that a high percentage of reflections will be partially recorded. It will be necessary to calculate accurately the expected positions of reflections on a film for integration purposes, and which reflections are partially recorded and must therefore be combined with the other partially recorded part from an adjacent oscillation range or corrected for the degree of partiality, or rejected.

The following parameters are relevant to the problem:

Unit cell parameters - Up to 6 parameters determining the relative positions of spots on a film. Depending upon the lattice and the crystal orientation not all parameters will necessarily be derivable from one film.

Orientation angles (ϕ_x, ϕ_y, ϕ_z)	-	It is convenient to describe the alignment of the crystal by defining a relationship between the beam (x), spindle (z) and vertical (y) axes of the camera and the cell axes, and then to measure the mis-orientation in terms of rotations about these three axes.
Crystal-film distance	-	Determining the relative spacing of spots on the
Wavelength	-	film. Wavelength is likely to be important for synchrotron data especially if it varies during an exposure !
Film centre and rotation x_c y_c ω	-	Parameters relating film coordinates to those determined by the scanner. Usually determined from the fiducial mark positions (or the direct beam mark), but ω is effectively ϕ_x - rotation about the beam axis.
Camera corrections (twist, tilt and bulge)	-	Parameters defining deviations from normal beam geometry important for predicting accurate spot positions.
Beam divergence	-	Which determine the length of time a reflection
Crystal mosaic spread (Υ)	-	is in the diffracting position and hence are
Wavelength spread ($\delta\lambda/\lambda$)	-	crucially important to distinguishing fully and partially recorded reflections.

With a blithe disregard for technical difficulties it is possible to state that the number of reflections recorded on a single protein oscillation photograph is so great that the determination of these parameters is a hugely over-determined problem and should be straightforward.

The program systems described were MOSCO (as used at Oxford - D Stuart), Munich (W Bennett) and OSCAR (E Stura, K Wilson) and the program manuals themselves or Arndt and Wonacott's "The Rotation Method in Crystallography" should be consulted for details.

In MOSCO crystal orientation is determined by an analysis of still photographs. At least two stills are taken 90^o apart on the rotation axis, with further stills at the ends of the rotation range. The coordinates of about 20 strong

spots are measured from each film using a data tablet connected to an Evans and Sutherland graphics device. It is assumed that strong spots are about half-way through the Ewald sphere at the point where the still is taken, rather than just on the edge, but no attempt is made to estimate where each spot is with respect to the sphere. The fiducials determine x_c, y_c and ω, which allows spot indices to be predicted, positions on the film calculated and the camera constants (x_c, y_c, ω, crystal-film distance) to be refined against the difference between observed and calculated positions. Then follows refinement of the orientation angles, minimizing the calculated reflection distances from the Ewald sphere. Finally, it is possible to refine the cell dimensions against spot position, which converges well if there are no indexing errors. Accuracy is judged from the rms residuals for spot positions and distances from the Ewald sphere, and from the standard deviations of refined parameters. MOSCO uses a fully predictive technique: spot integration box positions are determined from the refined cell and orientation parameters and not adjusted by reference to the observed spot pattern which makes it important to confirm the accuracy of this refinement process by plotting a prediction of the film pattern for comparison purposes.

The Munich system is broadly similar, with the major difference being that the orientation angles are defined with respect to the crystal position at the start of each oscillation range and thus differ for each film. Still photographs are taken at the beginning and end of each oscillation range. This method has the disadvantage that consistent treatment of partially recorded reflections is not possible since the orientation angles may refine to effectively different values for adjacent oscillation ranges. The Munich system is semi-predictive in that an evaluation box is placed around each predicted reflection position and is allowed to move to accommodate differences between observed and predicted spot positions. An analysis of these movements provides evidence for cell parameter or crystal-film distance errors, but these parameters are not refined against the positions of reflections on the stills.

OSCAR is an off-line system, but uses, once again, a broadly similar method. Differences arise from the use of the direct beam to mark the film centre, thus avoiding the x_c, y_c refinement of MOSCO, whilst the rotation of the still about the beam axis is included in the orientation angle about this axis. Two stills are taken at 90° to each other, and the inclusion of a third still at 45° often

improves the results. The coordinates of spots on the stills are prepared by program, scanning the digitized copy of the still to find the spots with maximum intensity. OSCAR is mid-way between the fully-predictive MOSCO and the semi-predictive Munich system: one set of orientation angles is used for each crystal data set but when the predicted pattern of spots is compared with each film it is possible to adjust the angles manually in a series of random steps in order to get better agreement.

There seems little doubt that existing methods solve satisfactorily the problem of calculating accurate positions for spots which are likely to occur on a particular film. What seems less certain is whether or not they can accurately predict which spots will pass through the Ewald sphere during the recording of a particular oscillation range and hence will appear on the film. The crux of the problem seems to be in the determination of Δ, the combined effect of beam divergence and crystal mosaic spread. Underestimate this and reflections will be classified as fully recorded which are in fact partials. Overestimate it in order to avoid this problem, and unnecessary spot overlap will result and the scope for the introduction of partial bias (see later) will be increased. A number of methods for measuring crystal mosaic spread were described.

Crystal Mosaic Spread Considerations

It is possible to check spots measured from stills in the regions of film where reflections move most rapidly through the Ewald sphere (E Stura). Assuming such (strong) spots are close to 0.5 recorded then γ can be adjusted manually to achieve a good fit. Post refinement of γ (see next section) has been attempted (A Wonacott) but the refinement leads to a broad minimum and depends upon the rocking curve. It should be possible (K Wilson) to take a series of stills at, say, 0.05° apart to find the angular spread of selected reflections, but is this possible for every crystal which may be needed for data collection and in view of the fact that mosaic spread can increase with crystal ageing in the X-ray beam ? It may be possible to compare integrated intensities of reflections from a 1° oscillation photograph with the intensities of the same reflections on a still recorded exactly in the centre of this 1° range (A Wonacott). If the distance of the still spots from the Ewald sphere is known then γ may be calculated from the ratio of the intensities, but this condition requires an accurate knowledge of the orientation angles and hence of γ.

Perhaps the best hope is held out by the suggestion that it is not so much the calculation of γ which is at fault but its application in calculating which spots are in fact partially recorded (T Greenhough). With a careful derivation of the predicted rotation range during which a reflection passes through the Ewald sphere for given values of beam divergence, mosaic spread and wavelength spread, he demonstrated that the conventional formulae ignore certain factors. When, in a test, the corrected formulae were applied to the apoferritin test photographs some 6% of fully recorded reflections were reclassified as partials. This work is about to be published.

Post Refinement

This technique of using the output from densitometering the films in order to improve the orientation matrices etc was briefly surveyed (A Bloomer). Post refinement can be used after some films have been measured in order to improve the parameters which will be used for predicting spot positions on the remaining films, or even to indicate that the first films should be reprocessed. Alternatively it may be used when a complete set of data has been obtained in order to improve the orientation parameters and hence the assignment of fulls and partials. Partials which become fulls after refinement can have one component discarded as background; fulls which become partials can either be thrown out or scaled up in proportion to the estimated fraction recorded.

In both cases the usefulness of post refinement rests upon the fact that after a film has been densitometered there will be a greatly increased amount of data available to determine cell parameters (spot coordinates) and orientation angles (which spots are observed). It is of particular importance for virus crystals where only one photograph may be possible for each crystal. Where successive oscillation photographs have been taken from one crystal then the refinement of a set of cell parameters and a set of orientation angles for each photograph, although requiring a lot of tedious book-keeping, can accommodate the effects of crystal slippage.

Film Orientation and Spot Integration

Once the crystal orientation has been determined it is possible to produce a list of film coordinates for reflections which are going to be measured. The relationship between film coordinates and drum scanner coordinates must then be established before spot integration can begin. The steps involved were illustrated by a description of the Oxford system derived from MOSCO, run as an off-line minicomputer version (D Stuart).

1. Transfer film density data to disk. A number of film scans will be stored on magnetic tape and must be transferred one at a time to disk.

2. Apply criteria for overlapping reflections.

3. Locate the 3 fiducials, at the same time estimating the fog level and the variance of fogged background for later use. If the search for fiducials fails then the film density is displayed on the Evans and Sutherland and the fiducials are selected manually.

4. Choose about 18 low resolution spots (by a centre-of-gravity calculation weighted by density) for a least-squares refinement of x_c, y_c ω and a scale factor which corrects for the scanner drum radius.

5. Then divide the full film into 18 areas and choose 2 spots from each for the refinement of twist, tilt and bulge corrections. Spots are rejected during refinement if their observed and calculated positions vary by too much ($\sim$ 60 μm).

6. Choose integration and background boxes by displaying a few spots on the vdu. The program has a feature for expanding the integration box for high angle spots (increased area because of oblique incidence) but this is often suppressed since boxes are usually over-estimated in order to accommodate the intense low angle spots and do not therefore need increasing at high angles. The 4 background boxes may be placed N, S, E, W or NE, NW, SE, SW relative to the integration box.

7. In the first integration pass the spot intensity is measured, data accumulated to allow a plane to be fitted to the background values, and the deviation of the spot centre-of-gravity from the integration box centre is measured.

8. In a second integration pass the integration box may be moved (up to some pre-set maximum) towards the observed c-of-g. Individual background values are checked against the values estimated from the calculated plane (using variance estimates previously determined from around the fiducials) and poor values are rejected. There are criteria for rejecting spots because of high background gradients or large c-of-g deviations.

A recent modification to this program is the inclusion of a profile fitting option using Rossman's method, which was found to reduce Rmerge by a factor of about 20% in a comparative study.

The other programs use the same general method. In the Munich system (W Bennett) the integration box size is specified for low angle spots and expanded to allow for oblique incidence at high angles. A picture frame background of width one raster is used, with the possibility of rejecting values which deviate too much from the background mean. The integration box is moved about within a small area in order to maximize spot intensity. A recent program improvement is first to estimate the background, then to determine the extent of the spot and hence its centre, and then to use the original integration box placed upon this centre.

Systematic Errors

Four sources of systematic errors in intensity data were discussed.

1. Partial Bias

Since partial reflections differ from fully recorded reflections only by the random choice of the boundaries of oscillation ranges then the expected value of the differences between the mean intensity values of full and partial reflections should be zero for all classes of reflection if there has been no systematic difference in their treatment. In the past this has been far from true, with the intensities of partially recorded reflections always being apparently over-estimated, often badly so.

More recently it has been found that this effect is still observed, but is not usually so large. The improvement must surely be attributable to program improvements but no very clear ideas were expressed as to the source of the bias or indeed, as to precisely which program changes were reducing it.

For 2.1Å tyrosin synthetase data (A Wonacott) 32,000 reflections gave Rsym = 12% with a small partial bias of around 2%. No variations were apparent when the bias was analyzed by intensity or $\sin\theta/\lambda$. At Oxford the experience with OSCAR was that a partial bias of 5 - 6% was common, and could be worse than that. At MRC Cambridge partial bias was found to vary with the crystals. One improvement was found to be the use of intensities rather than F-values to avoid the need

to set negative intensities to zero. As would be expected this reduced the partial bias for low intensity reflections.

It was generally felt that better crystal orientation matrices, and better estimates of γ, lead to a reduction in partial bias. It would follow that deliberate over-estimates of γ might make the partial bias worse, but there is no evidence for this.

2. Thermal Diffuse Scattering

It was suggested (E Stura) that TDS effects could explain partial bias. In essence, the argument goes that each of the components of a partially recorded reflection could be over-estimated by the inclusion of TDS peaks, whereas a fully recorded reflection will only contain one TDS contribution. Whether or not the TDS contribution to a partially recorded reflection should itself be partially recorded implying that the two TDS components of a partially recorded reflection would add to the TDS component of an equivalent fully recorded reflection seemed to be a matter for some debate.

If the TDS argument holds then Oxford crystals would seem to exhibit more TDS than other crystals. It would also follow that partial bias is going to be a problem with synchrotron data for a number of photographs were displayed that showed considerable TDS effects for synchrotron photographs.

3. Absorption

The method used by the Groningen group (R Kampuis) of measuring crystal transmission as a function of the rotation angle by recording the attenuated primary beam using the same collimator conditions, etc., was described. The results for papain have shown significant variation with transmission factors varying by up to 15%.

4. Radiation Damage

The usual treatment is to ignore it, and hope that the use of individual B values for each film during scaling will accommodate any variations due to radiation damage. Cases have been observed when the resolution of diffraction patterns decreases with time, an effect attributed to radiation damage. Where this damage is likely to be severe (virus crystals where only one photograph is possible from each) then fast oscillation rather than slow rotation would have a time-averaging effect. Crystal cooling may help. The localized heating of flat-plate crystals in a synchrotron beam caused crystal collapse and was found to be alleviated by blowing cold air over the crystals.

EFFICIENT SCALING OF LARGE DATASETS

A.G.W. Leslie (Imperial College)

A new version of the Cambridge program ROTAVATA which uses the Fox and Holmes algorithm (Acta Cryst. (1966) 20, p.886-891) to perform inter-film scaling, has been implemented on the 370 at Daresbury. The original version of this program was restricted to refinement of the scale-factors K_i and temperature factors B_i in alternate cycles of refinement. Because of the high correlation between these parameters, this results in a rather poor rate of convergence. The new version allows the K_i and B_i to be refined simultaneously, leading to a considerable increase in the rate of convergence and consequent saving of cpu time. By using the option for determining starting values for the K_i based on the average intensity of each film, experience with a variety of datasets has shown that convergence should be achieved within four cycles of refinement. (Previously a total of 30 cycles of alternate K, B refinement had been required to achieve convergence for a dataset comprising 130 films.)

A number of additional options have been included which are primarily for use with very large datasets (e.g. more than 100 films). In these cases the program will require in excess of 440K of core to run if all the parameters are refined simultaneously. In order to reduce the core requirement, a facility has been added to divide the films up into two (or more) overlapping blocks of films, and then refine the parameters of each block in turn. This "blocking" procedure was tested on a 2.5Å GPD dataset, which was made up of a total of 132 films, including 20 "cusp" films. The dataset was divided into two blocks each comprising 76 films, so that 20 films were common to both blocks.

- CONTINUED -

The behaviour of the refinement depended crucially on the way in which the films were divided into the two blocks, in particular which films were chosen to be common to both blocks. An inappropriate choice resulted in the parameters for these overlapping films oscillating from one cycle to the next, as each block of films was refined. In this case it was found that when the cusp films were used as the common set, the refinement was very well behaved and converged to give identical parameters to those obtained by refining all the films in a single block. The core requirement and cpu time per cycle were 304K and 172 seconds for the two-block refinement compared to 672K and 720 seconds for the single block refinement. (Note however that one cycle of single-block refinement is equivalent to two cycles of two-block refinement.)

A further modification has been made to allow for the situation that arises when a small number of films is to be added to a large dataset whose film parameters have already been refined. (For example, to replace films which scale poorly with the rest of the dataset.) It is now possible to refine only the parameters of the additional films, all other parameters being held constant, which results in a very significant saving of cpu time associated with the inversion of the normal matrix.

The program currently uses the Cambridge 9A2 data format, although a version for use with LCF data files should be available in the near future. Program documentation is also available on request.

MY TWO PLUS YEARS WITH CCP-4 AND WHAT IT HAS MEANT TO THE SHEFFIELD GROUP

As my time in Sheffield draws to a close it might be worthwhile to reflect on what an association with CCP-4 has meant to the Sheffield group. This is in fact committing to paper a few thoughts which I briefly expounded upon at a recent Northern Protein Crystallography meeting at Leeds University (15.4.81).

I have been associated with CCP-4 more or less from the pleasures of its conception to a point where we have a 2½ year old tottering towards a set of rationalized programs which are able to communicate with each other through flexible binary reflection data files, atomic co-ordinate files and Fourier map files. For Sheffield, it has meant the evolution of some 30 plus programs all documented in an agreed way each with a flexible CLIST. This achievement which would never has been possible for such a small group (average size two postdocs. only) without tapping the large resources offered by CCP-4. The Sheffield suite must continue to expand to keep pace with the computations needed in the improvement of the structure of horse spleen apoferritin and the subsequent analysis of isoferritins. Whether this expansion occurs, for example, through being BOSSed around, or through some other alternative depends on how well the CCP-4 evolves and also sells itself.

Sheffield, I think, provides a useful yardstick by which to measure the success of a collaborative effort. We are a small group who had little previous protein crystallography software and who had access to the SRC network. Out collaboration has primarily been through the use of programs and reporting back on their faults and deficiencies. In this regard the spectrum of program experience could be highlighted by the following 3 examples (the programs will remain nameless):

a) The Good - a program which had a user supplied CLIST which proved to be adequate, no apparent program bugs (perhaps because symmetry, F432 for apoferritin, played no important part), and for which the write-up was adequate, although was later expanded to include user experience.

b) The Bad - a totally inadequate write-up, no CLIST and gave wrong results for F432.

c) And The Essence - a well written and documented program, having been used extensively by one of the larger groups, taken over and developed for Daresbury by the Daresbury people using, in part, experience gained by my use of the program.

At some pointperhaps Working Group II should consider each program to define its catagory, perhaps by asking a question. Who uses the x.LOAD version of a program with x.CLIST? I can't help feeling the answer might be a little embarrasing!

Unlike every Working Group II meeting I will not spend any time discussing data formats.We can do little more than acknowledge their usefulness and flexibility until they are implemented and well tested. Until then we must confess to keeping our own LOAD libraries bulging.

There seems little point in discussing hardware either, as sometimes, assuming machine and network are functioning we don't seem to have access to it anyway! Daresbury should be included as a resource in the next edition of 'Limits to Growth'.

As CCP-4 begins to grow older I have every confidence that it will overcome its teething problems, some of which I have touched upon, and provide a very useful tool to protein crystallographers. My best wishes go to everybody associated with the project, particularly to Pella Machin, John Campbell and Mike Elder who put in much effort. I thinkwe all agree it would be a great loss if a disease similar to that which led to the death of the NRCC were to prevent a blossom into maturity.

Phil Bourne

DATA FORMATS

Pella Machin (Daresbury)

The working groups of the Collaborative Computing Project (CCP4) have been most concerned to establish common data formats. Currently, much time and effort is spent converting from one format to another, and valuable disk storage space is used up by the resulting duplicate copies of data.

Agreement has been reached on format specifications for:

1. Coordinate data
2. Reflection data
3. Map data.

Details of these formats are given below.

1. Coordinate data

Atom coordinate data will be formatted according to the definition of the Brookhaven protein data bank. The data bank allows a variety of data cards to be present in a file, the type of data card being determined by the characters in the first 6 columns of each card. Atom coordinate records are formatted as follows:

Cols.

1 - 4	ATOM
7 - 11	Serial number (residues are given in order beginning with the amino-terminus)
13 - 16	Atom name
18 - 20	Residue name
21 - 27	Sequence identifier
31 - 38	X)
39 - 46	Y) Orthogonal Å coordinates
47 - 54	Z)
55 - 60	Occupancy
61 - 66	Temperature factor
68 - 70	Footnote number

FORMAT (6A1,I5,1X,A4,1X,A3,1X,A1,I4,A1,3X,3F8.3,2F6.2,1X,I3)

Note that:

(a) The coordinates are in orthogonal Å units with the orthogonal axes being such that

X // a
Y // c^* x a
Z // c^*

(These are different from the frequently used Rollett orthogonal axes X // a^*, Y // c x a^*, Z // c).

(b) The atom name consists of 4 characters:

1 - 2 Chemical symbol - right justified
3 Remoteness indicator (alphabetic)
4 Branch designator (numeric)

For example:

main chain atoms are ^CA^ ^C^^ ^N^^ ^O^^

side chain atoms are ^CB^ ^NH1 ^CG^ ^CD1 etc.

(c) Since the file may contain other types of data the characters ATOM should be put in colums 1 - 4 of each atom coordinate card to indicate the type of data present.

John Campbell (Daresbury Laboratory) has written some FORTRAN routines to handle input of Brookhaven data. These routines are available for use and are described in this newsletter in a separate article.

2. Reflection Data

The reflection data files will use the 'Labelled Column Format' (LCF for short and previously named NA2) proposed and developed by Dr T N Bhat and Dr A J Wonacott (Imperial College). In general terms the LCF file format uses fixed length records, the first records of the file contain header information (title, cell dimensions and column labels) and the remainder contain reflection data. For the reflection data, the columns are identified by alphanumeric labels - hence the name LCF.

Sets of subroutines are available for reading and writing standard LCF reflection data files.

Further information is available in the paper "Standard LCF Reflection Data Files and their use" which can be obtained from the Daresbury Laboratory.

3. Map Information

Most programs in the UK uniformly use one particular format for map information. It has been agreed that this format could be improved by including a few extra items. Provided these items are added by extending current records of the file, existing programs (not reading such information) can still function without error while reading new 'map files'.

The format agreed upon is as follows:

(a) A binary file

(b) Record 1 - title information

TITLE(1-80), NSEC, IU, IV, IW, NX, NY, NZ where

TITLE(1-80)	80 character title	
NSEC	Total number of sections in the file	
IU)	Fast)	axis order definition representing x as 1,
IV)	Medium)	y as 2, z as 3, e.g. IU = 3, IV = 1, IW = 2
IW)	Slow)	means z is fast, x is medium, y is slow.
NX)	Sampling intervals along x, y, z given by the total	
NY)	number of points along the whole cell edge.	
NZ)		

Note that the above 7 integers and all that follow are full integers (4 bytes on an IBM computer).

(c) Paired records for each section, as follows:

ISEC, MINF, MAXF, MINM, MAXM where

ISEC	Section number
MINF)	minimum value of fast index
MAXF)	maximum value of fast index
MINM)	minimum value of medium index
MAXM)	maximum value of medium index

Density values for this section (ISEC)

$\rho(ij)$ where $MINF \leq i \leq MAXF$

and $MINM \leq j \leq MAXM$

(d) No terminator record.

Reading of the file can terminate on a FORTRAN "END=" option or can be accomplished with the knowledge of the number of sections from the title record (NSEC).

4. Implementation of Formats

Work is underway (by groups at Birkbeck, Imperial, Cambridge and Daresbury) to modify existing protein crystallography programs to handle these new data formats. We aim to have a working set of initial processing programs with documentation in the near future.

These programs run on the IBM 370/165 at Daresbury. The identifier PCZ is available for storing this new program suite of modified programs and the associated CLISTS for running them.

It is hoped that these programs will be used soon and that protein crystallographers will routinely choose to run these versions when they have new data to process. The availability of documentation may be an encouragement to use such, and future program enhancements will only be made to these versions.

Groups computing at other sites (York, Cambridge, Leeds) are also planning to try these formats. Program exchange will be greatly facilitated if these data formats are accepted and used by all groups.

SUBROUTINES FOR HANDLING BROOKHAVEN FORMAT COORDINATE FILES

JOHN W CAMPBELL (DARESBURY LABORATORY)

1. INTRODUCTION

Some subroutines have been written at the Daresbury Laboratory for reading and writing coordinate records for Brookhaven format coordinate files. The Brookhaven format defines a standard setting of orthogonal axes with respect to the crystallographic axes and it was agreed by the CCP that this standard should be used within the programs developed for the CCP. The standard set of orthogonal axes XO, YO and ZO is defined as follows:

$$\begin{aligned} XO &\parallel \underline{a} \\ YO &\parallel \underline{c}^* \times \underline{a} \\ ZO &\parallel \underline{c}^* \end{aligned}$$

Within a Brookhaven format file, however, coordinates may be held with respect to other sets of axes. After consultation with David Moss at Birkbeck College, it was felt that the best solution to the problem of different orthogonalisations was that the subroutine, reading coordinates from a Brookhaven format file, should always return the coordinates with respect to the standard axes. Thus, if a program requires fractional coordinates then, provided that the cell dimensions are available, these can be readily calculated knowing that the coordinates returned from the subroutine are in a standard setting. The subroutine RBROOK in the set of subroutines described below will re-orthogonalise the coordinates to the standard setting provided that the coordinate file contains the CRYST1 card (holding the cell parameters) and the SCALE cards (holding the transformation matrix from stored to fractional coordinates). Coordinates in a file not containing the CRYST1 and SCALE cards are assumed to be in the standard Brookhaven setting.

2. THE SUBROUTINES

Two pairs of subroutines have been written for handling Brookhaven format coordinate files. These are:

(a) RBROOK, RBINIT for reading coordinates

(b) WBROOK, RWBFIN for writing coordinates

The subroutine RBROOK is used to read coordinates from a Brookhaven format coordinate file (and if necessary transform these coordinates to the standard setting) in a single pass through the file. The subroutine RBINIT is called for initialisation prior to reading coordinates via RBROOK.

The subroutine RBROOK also has an option to write the non-ATOM/HETATM records to an output file as they are read from the input file. The subroutine WBROOK is used to write the updated ATOM/HETATM records. The subroutine RWBFIN may be used to copy all remaining records from the input file to the output file.

3. THE SUBROUTINE 'RBINIT'

The subroutine RBINIT is used to rewind an input coordinate file and to perform initialisations prior to calling the subroutine RBROOK.

Subroutine call: CALL RBINIT (IUN)

Subroutine parameters:

IUN	Unit number of input coordinate file

4. THE SUBROUTINE 'RBROOK'

The subroutine RBROOK is used to read coordinates from a Brookhaven format coordinate file. The subroutine RBINIT must be called prior to reading or re-reading a file via RBROOK. If required the non-ATOM/HETAM records may be written to an output file as they are read.

Subroutine call:

```
CALL RBROOK (IUN, ISER, IZ, IRESN, X, Y, Z, B, OCC,
* ATNAM, RESNO, RESNAM, IOUT, MSG1, MSG2, ITER, &1, &2)
```

Subroutine parameters:

IUN	Unit number of the input coordinate file
ISER	Atom serial number
IZ	Atomic number (returned as 7 for ambiguous atoms)
IRESN	Residue number as an integer
X, Y, Z	Orthogonal Angstrom coordinate in the standard Brookhaven setting
B	Temperature factor
OCC	Occupancy
ATNAM	Atom name (packed characters, 4/word left justified)
RESNO	Residue number (packed characters, 4/word unjustified)
RESNAM	Residue name (packed characters, 4/word unjustified)

IOUT	Unit number to which non-ATOM/HETATM records are to be written (may be Ø if reading only)
MSG1	Unit number for flagging ambiguous/unknown atom types (may be Ø if messages are not required)
MSG2	Unit number for listing matrix applied to input coordinates (if calculated by the sub-routine) (may be Ø if listing of matrix is not required)
ITER	Flag = 1, Return via Return 1 if TER card found
	Flag = 0, Do not return when TER card found.
&1	Return on TER card found (only if ITER = 1)
&2	Return on end of file found.

If the user wishes to access details from the ATOM/HETATM card which are not returned via the parameter list, then the complete record read is held in character form in the common block /RBRKXX/IFCRYS, IFSCAL, IFEND, R(3,3), S(3,3), T(3), IRTYPE, IE, IBROOK (64).

IRTYPE	holds columns 1 - 4 packed
IE	holds columns 5 - 6 packed
IBROOK	holds columns 7 - 70 one character/word.

5. THE SUBROUTINE 'WBROOK'

The subroutine WBROOK is used to write an ATOM/HETATM record to an output file. It is used in conjunction with the subroutines RBINIT and RBROOK which set up any transformation matrices required and which will copy non-ATOM/HETATM records to the output file. All calls to WBROOK must be paired with corresponding calls to RBROOK so that information from the input record, not passed via the parameter list, e.g. the footnote number, etc. can be written to the output record (see the example in section 8 below).

Subroutine call:

```
CALL WBROOK (IOUT, ISER, IZ, IRESN, X, Y, Z, B, OCC,
* ATNAM, RESNAM)
```

Subroutine parameters:

IOUT	Unit number of the output coordinate file
ISER	Atom serial number
IZ	Atomic number (may be zero if the atomic symbol is a single character, e.g. C, N, O, H, S)

IRESN	Residue number as an integer (max. of 3 digits)
X, Y, Z	Coordinates (standard orthogonal)
B	Temperature factor
OCC	Occupancy
ATNAM	Atom name (packed characters, 4/word left justified)
RESNAM	Residue name (packed characters, 4/word - left three characters are written).

6. THE SUBROUTINE 'RWBFIN'

The subroutine RWBFIN is used to copy the remaining records of a Brookhaven coordinate file from an input to an output file. It will normally be used after part of the file has been copied and updated using the subroutines RBROOK and WBROOK but may be used to copy a complete file after calling RBINIT.

Subroutine call: CALL RWBFIN (IUN, IOUT)

Subroutine parameters:

IUN	Unit number of input coordinate file
IOUT	Unit number of output coordinate file

7. EXAMPLE OF READING A FILE

The example below outlines the way in which the subroutines RBINIT and RBROOK are used to read coordinates from a Brookhaven format coordinate file (on unit 1)

```
          CALL  RBINIT(1)
10        CALL RBROOK(1,IS,IZ,IR,X,Y,Z,B,OC,AT,RNO,RNM,Ø,Ø,6,Ø,&10,&100)
          :
          process data
          :
          GO TO 10

100       STOP
          END
```

8. EXAMPLE OF READING A FILE AND WRITING A MODIFIED FILE

The following example reads in a set of coordinates and processes them (e.g. in a refinement). The input file is then re-read and an output file is created with the updated coordinate data.

```
C
C  READ  IN COORDINATE DATA (UNIT 1)
C
          DIMENSION X(1000), Y(1000), Z(1000), R(1000), A(1000)
          N = 0
          CALL RBINIT(1)
10        N = N+1
          CALL RBROOK(1,IS,IZ,IRS,X(N),Y(N),Z(N),B,OC,A(N),RNO,R(N),Ø,Ø,6,Ø,&10,&100)
          GO TO 10
C
C REFINE COORDINATES
C
100       NMAX = N-1
          :
          refine coordinate values
          :
C
C WRITE UPDATED COORDINATE FILE (UNIT 2)
C
          CALL RBINIT(1)
          DO 150 N = 1, NMAX
          CALL RBROOK(1,IS,IZ,IRS,XX,YY,ZZ,B,OC,ANM,RNO,RNM,2,Ø,6,Ø,&200,&200)
          CALL WBROOK(2,IS,IZ,IRS,X(N),Y(N),Z(N),B,OC,ANM,RNM)
150       CONTINUE
200       CALL RWBFIN(1,2)
          STOP
          END
```

9. AVAILABILITY OF THE SUBROUTINES

Two load modules are available at the Daresbury Laboratory as follows:

PCZ.LOAD (RBROOK)	contains the subroutines RBROOK and RBINIT (also FVAL and PACKB)
PCZ.LOAD (WBROOK)	contains the subroutines WBROOK and RWBFIN

The source code for the subroutines may be obtained on request from the Daresbury Laboratory.

CCP4 PROGRAM LIST

Pella Machin (Daresbury)

A list of the Protein crystallography programs which are available at Daresbury in association with CCP4 is now held in a file on the IBM 370/165 called PCB.CCP4PROG.TEXT.

An entry for a particular program consists of a program name, brief program description, program status (F = uses new formats, I = running on IBM, C = running on Cray, U = under development), distribution status, documentation status and a person to contact for information.

Anyone with access to the IBM may list the file to get up to date information on the programs. I will update the file with new information - other users may do so too if they know of any necessary changes.

The current list of programs follows.

CCP4PROG.TEXT

PROTEIN CRYSTALLOGRAPHY PROGRAMS ON THE DARESBURY IBM 370/165 AND CRAY COMPUTERS

NAME	DESCRIPTION	PROG STAT	DIS	DOC	CON
1.Data Processing					
OSCAR	Film data processing	I		B	JH
VALDIF	Validation of diffractometer data (input is specific for Birkbeck diffractometer data)	I		B	IJT
DIFCOR	Applies LP, absorption, radiation scale to output of VALDIF program for diffractometer data.	I FIU		B	IJT IJT
2.Scaling and Merging of 3D Data					
INTOLCF	Converts Imperial College film data to LCF files	FI		C	AJW
ROTOVATA	Determines relative scale and temperature factors using the method of Fox and Holmes. Jan 81 new formats being put in by Alan Wonacott.	FIU		B	AJW
AGROVATA	Merges datasets and calculates statistics. As above re new data formats.	FIU		B	AJW
HRS4	Scaling of datasets - methods of Hamilton, Rollett, and Sparks. Used mainly on diffractometer data.	I FIU		B	IJT IJT
DSCALE	Scaling of heavy atom derivatives. (6 methods are available eg weighted Kraut).	I FIU		B	DSM DSM
CAD	Collect assorted data, apply symmetry transformation and format changes.	FI		A	JWC
NORMAL	Calculation of overall temperature and scale factors - from MULTAN-78. (Original author is Peter Main at York University).	I		B	?
TRUNCATE	?	IU		B	?
ANSC	Analyse and scale derivative data	FI		B	JWC
ANISISC	Scale derivative data using anisotropic temperature factor scaling function.	FI		B	JWC

3.MIR Phasing and refinement

PHARE	MIR phasing and refinement. (Bricogne's method)	I	B	TNB
REFINE	Refinement of heavy atom positions. (Based on the Oxford BOSS programs, using HFLE)	I FI	B	IJT JWC
PHASE	Calculation of isomorphous phases , (produces Hendrickson-Latman coefficients)	I FI	B	IJT JWC

4.Rotation function

ALMN	York (Eleanor Dodson) version of Rotation function, using method of Crowther. All 3 stages (DLM ALMN FRFSUM) are combined in one program . All symmetry related peaks are output. Allows up to 30 bessel functions.	FI	A	PAM
ALMN60	As ALMN30 , but allows 60 Bessel functions.	FI	A	PAM

5.Translation Function

SEARCH	Search and Rfactor calculation, by Eleanor Dodson.	FI	B	PAM

6.Fourier

FFT	Fast fourier transforms, method of Ten Eyck. Available for subgroups of all space groups (Acta Cryst (1973) A29, 183)	FI I	Y	A B	JWC IJT

7.Map Contouring

PLUTO	Fourier map plotting and/or atom plotting	FI	A	PAM
EXTEND	Cambridge version of map extension for PLUTO	FI	B	JWC
CONTOUR	Map contouring , output on the FR80. Necessary unless PLUTO is enhanced to optionally output to the FR80.	I	B	IJT

8.Graphics input

CONPROG	Contours FFT output into a form suitable for Builder, Diamond's interactive graphics display programs	I	B	TNB
MAPMIX	Combines several outputs from CONPROG	I	B	TNB
AJCON	contours FFT output for FRODO on the graphics (ie for A.Jones program)	I	B	IJT

9.Model Drawing

RIBBONS	Ribbon program, method of McLachan (output on FR80)	IU	C	DSM
PLUTO	Highly modified version of Sam Motherwell's original program, main changes by Phil Evans (as in 7. above).	FI	A	PAM

10.Model Stereochemistry

MODFT	Regularization of peptide geometry (Acta Cryst (1976) A32 311)	I FI	B	PAM JWC
EREFN	Levitt energy refinement (Acta Cryst (1978), A34 ,931)	IU	B	DSM
TORSION	Torsion angle calculations	I	B	PEB
DISTAN	Protein geometry	I FI	B	PAM JWC
FISIPL	Ramachandran plot - from Brookhaven data bank	I	B	PEB
MAIN	Generation of a polyalanine structure (Starting from C alpha positions ?)	I	B	PEB
BALUPLOT	Balusubramanian plots of phi,psi	I	C	DSM

11.Fourier Map and generation manipulation

GENED	Map generation from model co-ordinates	I	B	TNB
ISOLATEM	Isolates protein molecules.	I	C	DSM
KEBAB	Skew planes using Bricogne double sort method Beware there are some space group specific parts	I	B	TNB
ROTMAP	Rotates map (applies the angles obtained from the rotation function and applies them to the electron density map)	I	B	IJT

12.Structure factor calculation

SF	Fast fourier structure factor calculation Available for space groups P1, P21, P212121, P21212, P41212, P3121, R3.	I	A	JWC
	For P1 only a new version.	FIU	B	JWC
FCALC	Structure factor calculation using overall temperature factor. Subgroups available for all space groups.	I	B	DSM
FCALB	As FCALC only with individual structure factors	I	B	DSM

13.Weighting schemes

SIMW	Application of Sim's weighting scheme to partial structures.	I	B	IJT

14.Phase Recombination

HLSIMW	Hendrickson Lattman phase combination with Sim weighting. Input is from the PHASE output, plus Fc's calculated by RESTRAIN. Its function appears to be very similar to COMBIN	I	B	IJT
COMBIN	Bricogne's method of combining Hendrickson-Lattman coefficients.	I	B	TNB
		FI	B	JWC

15.Least-squares

SF	Isaacs-Agarwal fast Fourier refinement (Acta Cryst (1978) A34 791). Available for space groups : P1, P21, P212121, P21212, P41212, P3121, R3	I	A	JWC
		FIU		JWC
RESTRAIN	Restrained Least-squares (either geometric or phase) refinement for any space group.	I	B	DSM
		CU	B	DSM
KONNERT	Hendrickson-Konnert refinement (Acta Cryst (1976), A32, 614). Two versions available, both on the Cray - consult contacts for details. Available space groups are P21, P21212, P212121, P43212, C2, I222.	C	B	WP
		C	B	TNB

16.Program packages

ROCKS	Crystallographic system from George Reeke at the Rockefeller University, New York, encompassing many aspects of protein crystallographic computing	IU	N	B	JWC

17.Ancillary programs available from Brookhaven

TAPDIR	Print directory of Brookhaven Data base contents F.Bernstein, Brookhaven is the main contact at Brookhaven for all these programs	IU		B	PAM
BENDER	Parameters for best wire models	IU		B	PAM
CONECT	Generate peptide connectivity	IU		C	PAM
DGPLOT	Diagonal plots on the printer	IU		C	PAM
DSTANCE	Distance calculations	IU		C	PAM
FISIPL	Phi/Psi plots on the printer	I		B	PEB
NAMOD	Ball and stick model display	IU		C	PAM
PHIPSI	Main chain torsion angles	I		B	PEB
STEREO	Extract X, Y, Z from stereo diagrams	IU		C	PAM
TORSRU	Complete torsion angles	IU		C	PAM
TOTALS	Validation of master record	IU		C	PAM

18.Small molecule programs

SHELX	Structure solving package written by George Sheldrick, Gottingen.	I	N	B	ME
MULTAN78	Direct methods for structure determination written by Peter Main, York University.	I	N	B	ME
PLUTO	Molecular plotting program. Original version for small structure drawings, written by Sam Motherwell, Cambridge.	I	N	B	ME
DIRDIF	Direct methods using difference structure factors Acta Cryst (1979) A35, 765. Written by P. Beurskens at Nijmegen.	I	N	B	?

19.Utilities

MESSAGE	Message and program logging program (PCLOG).	I	A	JWC
DODLCF	Convert from Dodson Binary Format to LCF	FI	A	JWC
BIRKLCF	Convert from Birkbeck binary format to LCF	FI	A	JWC
CARDLCF	Convert from Card image free format to LCF	FI	A	JWC
LCFDUMP	Dump/summarise contents of an LCF file	FI	C	JWC
DISKLIST	List, in summary the contents of a private disk	I	C	JWC
COMPARE	Compare program to compare two source files	I	A	JWC
SETPAR	CLIST for handling ROUTE/CLASS/filename parameters. (general description is available)	I	B	JWC
LCFUTIL	LCF utility program to create a new or rearranged or edited LCF file from one or two existing LCF files.	FIU	B	TNB

NOTES ON THE ABOVE TABLE

Introduction

The programs listed above are being developed in association with the Collaborative Computing Project in Protein Crystallography (CCP4). Many programs are third or fourth generation i.e.after initial coding they have been modified by several different laboratories.

The table is divided horizontally into sections on particular areas of computing. The meaning of the abbreviations and the content of the various columns of the table are given below:

1.Program name

This is an 8 letter code given as the program name for shorthand reference.

2.Program description

A few words to outline the purpose of the program, perhaps the program method, historical development and current limitations.

3.Program status (PROG STAT)

Three columns are used to describe the implementation status.

F	denotes a program version using the new formats
I or C	a code to describe which machine the program is available on I = IBM and C = Cray
U	if present denotes the program is still under development

4.Program distribution (DIS)

One column is available for one of three symbols..

Y	Program is available for distribution, i.e. a version with new formats is available and documentation is on the word processor.
N	The program is not available for distribution and will not be in future. This may apply to program packages obtained from elsewhere which are always distributed from the original site.
blank	A program which is not yet ready for distribution but which may be sometime in the future.

5.Documentation status (DOC)

One of three symbols may be used to state this.

A	Documentation in standard form is available on the word processor.
B	Written documentation
C	'Back of envelope' documentation

6.Contact (CON)

The initials of the person responsible for the implementation at Daresbury are given. Usually initial questions should be directed to this person, who will know about program versions, methods of running the program, etc.
A key to the initials is given below :

PEB	Phil Bourne	SHEFFIELD	0742 - 78555 ext 4242
TNB	Bhat	IMPERIAL	01 - 589 - 5111
JWC	John Campbell	DARESBURY	0925 - 65000 ext 528
ME	Mike Elder	DARESBURY	0925 - 65000 ext 350
JH	John Helliwell	KEELE/DARESBURY	0782 - 62111 ext 307
DSM	David Moss	BIRKBECK	01 - 580 - 6622
PAM	Pella Machin	DARESBURY	0925 - 65000 ext 528
WP	Bill Pulford	OXFORD	0865 - 56789 ext 394
IJT	Ian Tickle	BIRKBECK	01 - 580 - 6622
AJW	Alan Wonacott	IMPERIAL	01 - 589 - 5111 ext 1853

READY