DARESBURY LABORATORY

INFORMATION QUARTERLY for PROTEIN CRYSTALLOGRAPHY

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

Number 7 September 1981

Contents

Editor: Pella Machin — Science & Engineering 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

SOME THOUGHTS ON A SABBATICAL IN CANADA

Lindsay Sawyer (Edinburgh)

The amazing thing about the brain is its capacity to keep working from morning till night, day after day and it continues to do so; until it is required to produce an article for a journal such as this when, mine at least, stops completely. My study leave in Edmonton is fast coming to an end and Pella wondered if I would put down on paper something of my experiences over the past months.

I consider myself extremely fortunate to have been able to spend time in the Biochemistry Department, University of Alberta and in particular in the laboratory of Mike James. My period in Bristol introduced me formally to the serine proteases and I had been wanting to re-open my work on elastase for a number of years. Unfortunately, being in a 'poly' has meant that my research effort has necessarily been limited and so the chance of a period of full time research was eagerly taken up when it presented itself.

The choice of Edmonton was based partly on the interest in the serine proteases and on the expertise available in protein refinement. Despite being about the latitude of Birmingham, an altitude of only half way up Ben Nevis and further from the sea that it is possible to get in Britain, the climate has been quite enjoyable! Mind you temperatures did drop below -30^{o}C but life is geared to cope and mostly all that we were aware of was the central heating working harder than usual. However, they say that this last winter was rather mild!

Shortly after arriving, we discovered that others too had wanted to work on elastase. In particular, interest in the structure at -70^{o}C seemed to be the main concern of two groups: Edgar Meyer at Texas A. & M. who was collecting data in Germany with Huber's group, and Xuong at La Jolla who is collecting typically 160,000 reflections with a count time of some 6 mins per reflection, albeit not sequentially! This was rather a blow but we decided to press ahead with room temperature studies of some peptide substrates in order to answer the question of the apparently anomalous binding to elastase of product inhibitors like Ac-ProAlaProAla-OH.

Collecting data on a CAD4, we discovered intensity losses due to radiation damage greater than had been previously encountered. At first, we thought that this was the fact that the crystals were not freshly prepared. However, both penicillopepsin and α-lytic protease suffered the same alarming decay and this could be compared directly with the decay obtained from the same crystals under the same laboratory conditions but using a Picker FACS-1. The conclusion, then, was that the much shorter focus to crystal distance was responsible and, after finding that none of the several monochromator crystals tried was able to produce a satisfactory beam, we went back to the original Ni-filtered radiation but this time with an extra thickness of Ni-foil glued to the collimator. This worked perfectly!

Data collected for the Ac-ProAlaProAla-H derivative gave an excellent difference map showing clear binding in approximately the same manner as Ac-ProAlaProPhe-H in SGPA or the chloromethyl ketone derivative in γ-chymotrypsin. However, data collected for Ac-ProAlaProAla-NH_2 at two separate concentrations have given only a suggestion of binding, but in a different manner. This is rather frustrating since photographs of soaked crystals showed clear intensity changes but neither $|F_o| - |F_c|$, native α_{calc} nor $|F_o| - |F_c|\ \alpha_{calc}$ showed convincing binding. An even higher concentration is being tried at present.

While the problems with data collection were being ironed out, it was possible to begin learning about high resolution protein refinement in a practical way. The Hendrickson-Konnert restrained least squares program, which was so successful in refining SCPA was, and is, the method in full swing here particularly since it has been modified to use the array processor on the Amdahl (the AP costs nothing to use, yet!) by Bill Furey in Pittsburg. This, coupled with the MMS-X graphics system operating under Colin Broughton's M3 software has made refinement an efficient and relatively inexpensive process. Data for the protease B from Streptomyces griseus had been collected to 1.7 Å and refinement had ground to a halt at R ∿ 28%. Starting from this point examination of the $2|F_o| - |F_c|$, α_{calc} map revealed a series of misinterpretations of the MIR map and one ortwo slight variations in the sequence. These corrections started the R-factor on its way down again and after several rounds of 'water-divining', side chain amendment and the like, the R-factor has reached 15.1% for the 17,500 data with $I \geq 2\sigma(I)$. Fairly tight restraints on the geometry were used throughout and, in retrospect, this probably served to slow down convergence somewhat

although it inhibited some of the poorer regions becoming grossly distorted and did away with the need for idealisation. We are currently examining the final(?!) maps; the problem is how to convince oneself that one should stop!

I have found working in Edmonton a most stimulating and rewarding experience not least because the personnel in the lab, to whom I would like to say a heartfelt thank-you, have been cheerful and helpful and have made my stay so enjoyable. I must also express my gratitude to the MRC of Canada and to Lothian Regional Council for making this sojourn possible.

RIBONUCLEASE-A : LEAST-SQUARES REFINEMENT OF THE STRUCTURE AT 1.45Å RESOLUTION

This work was presented at the IUCr Congress in Ottawa in August 1981 by Nivedita Borkakoti, David S Moss and Rex A Palmer (Birkbeck College)

The crystal structure of bovine pancreatic ribonuclease-A has been refined by restrained least squares analysis employing X-ray diffractometer data to 1.45Å resolution. The current R-factor for 19,238 reflections is 0.26 and 0.24 for 17,427 reflections with $I(hk\ell) > 0$. The rms deviation from ideality of bond lengths is 0.01Å. Corections, mostly of minor character, to previous models of the secondary structure have been made and a quantitative analysis of intra-molecular hydrogen bonds is given. A total of 130 solvent molecules have been clearly identified around the enzyme molecule and included in the least squares analysis. A sulphate anion occurs in the active site and has also been refined as part of the structure. Further new features of the structure to emerge are: alternative sites for the His-119 side group with occupancies, refined in the analysis of 0.80 and 0.20 respectively; a solvent molecule hydrogen bonded to the N-terminal amino group; and extensive disorder of the side chains in the region of residues 35-39. The rms deviation in atomic position between the current model and the starting model is 1.1Å including some shifts of 7 - 8Å where major rebuilding of side groups was necessary. Recent energy calculations on the active site residues have shown the presence of two energy minima corresponding closely to the alternative sites of His-119. A detailed analysis of the hydrogen bonds with certain hydrogens in calculated positions is now being carried out.

A COMPUTER PROCEDURE FOR EXTENSION OF MODEL ELECTRON-DENSITY FOR DENSITY MODIFICATION OR FOR REFINEMENT PURPOSES

A poster presented at the XIIth IUCr Congress in August 1981 by

T N Bhat and D M Blow (Imperial College)

SUMMARY

An efficient computer procedure has been developed for the extraction of regions of continuous well-connected high density from a three-dimensional electron-density map. This may be used to produce a more realistic electron-density distribution from a poorly phased protein map at moderate resolution, by a cyclic process. The procedure may be used to generate an extended model volume from a smaller volume, based, for example, on a starting atomic model of the partial structure; or it can identify possible structural features in the uninterpreted regions of the electron-density by picking up large continuous islands of positive electron density. The procedure can also be used to trace molecular boundaries or solvent regions in an automatic way. Once the density within the extended model volume has been scaled to the starting model, they can be used together as an improved electron-density distribution of the molecule. Phases calculated from this density may be used in a phase combination procedure to improve the starting phases for the next cycle of a convergent process. Where portions of the electron-density cannot be interpreted at the atomic level, they can be incorporated in the calculation of phase angles during least-squares refinement of structural parameters, giving improved convergence to the refinement. The procedure has been applied to the structure determination and refinement of tyrosyl-tRNA synthetase.

STEPS INVOLVED IN DENSITY MODIFICATION

The procedure includes seven steps which form an iterative cycle:

(1) Determination of 'occupancy' for each residue in the tentative model of the molecule as observed in the current electron-density map.

(2) Calculation of electron-density for the tentative model, using the occupancies determined in step (1).

(3) Extraction from the current electron-density map of (a) features corresponding to the tentative model used in steps (1) and (2), or linked to it through regions of high electron-density (and (b), optionally, other electron-density features which probably correspond to real features of the structure), to form the extended model volume.

(4) Generation of the extended model density on a finer grid, appropriate scaling of the model density to the electron-density map.

(5) Calculation of structure factors based on the extended model density.

(6) Modification of the phase information for each reflection, based on the calculated phase, pre-existing phase information derived from isomorphous replacement or any other source, and on the global agreement between observed and calculated structure factors.

(7) Calculation of a revised electron-density distribution using these modified phases.

This new electron-density map is used for re-assignment of occupancies (step 1), and the whole procedure is iterated, using the same tentative model of the molecule. The process is found to converge within a few cycles, and then the latest revised electron-density distribution forms the basis for building a new tentative model. If this model contains a significant number of new features, a further application of the whole iterative procedure may be made.

Although our application of the procedure has always used a starting model, one can envisage starting from a map in which no interpretation can be given in terms of atomic co-ordinates. In this case, strong positive features on the electron density map can be picked up as structural features in step 3(b) and steps (5), (6) and (7) can be pursued to generate a new electron density map.

COMPUTER ALGORITHM FOR PICKING LINKED ELECTRON DENSITY

Three logical arrays S(y), E(y), O(y) are used. S and O can be considered as 'searching' and 'output' arrays. E(y) is a starting electron density array.

Throughout the procedure, any point in the electron-density map may be marked as S, E or O, or none of these. No point is ever more than one of S, E or O. The whole procedure may be envisaged as the placing of one of three kinds of marker, S, E or O, at points of the electron-density map.

In each step of the procedure, S represents points in the map which are the growing surface of the extended model volume. One by one, each point in S is deleted from S and entered into O. We then consider every neighbouring point which is not yet part of O or S. Such a neighbouring point becomes part of S if the starting density is large (i.e. it is part of E), and if a sufficient number of its neighbours are part of S, E or O (i.e. if it is sufficiently well connected to points where the starting density is large).

To start with, all points in the model density, M (y), which exceed a chosen value c_2 are accepted as part of S. Any of these points which are already marked E (where the starting density ρ (y) exceeds c_1) are removed from E.

This algorithm may be expressed formally as follows:

Initially, for all y:

$|\underline{O}(y) = \text{false}$
$\underline{S}(y) = \text{true}$ $\underline{\text{if}}$ $\underline{M}(y) \geq \underline{c}_2$, $\underline{\text{else}}$ $\underline{S}(y) = \text{false}$
$\underline{E}(y) = \text{true}$ $\underline{\text{if}}$ $(\rho(y) \geq \underline{c}_1$ $\underline{\text{and}}$ $\underline{S}(y) = \text{false})$,
$\underline{\text{else}}$ $\underline{E}(y) = \text{false}.|$

Repetitive procedure for each y where $\underline{S}(y)$ is true:

$|\underline{O}(y)$ becomes true $\underline{\text{and}}$ $\underline{S}(y)$ becomes false.
For each y' = $\underline{\text{adjacent}}$ y, where $\underline{O}(y')$ is false:
($\underline{\text{If}}$ $\underline{E}(y')$ is true $\underline{\text{and}}$ more than $\underline{n}$ trues are found
(amongst ($\underline{E}(y'')$ $\underline{\text{or}}$ $\underline{S}(y'')$ $\underline{\text{or}}$ $\underline{O}(y'')$,
where y'' = $\underline{\text{adjacent}}$ y' then $\underline{S}(y')$ becomes true
$\underline{\text{and}}$ $\underline{E}(y')$ becomes false)).|

The repetitive procedure is continued over each remaining true point in $\underline{S}$, until all members of $\underline{S}$ are false. The array $\underline{O}$ then defines the extended model volume. A point y' is defined as $\underline{\text{adjacent}}$ y if $|y' - y| \leq$ some pre-determined neighbour distance $\underline{d}$. In our work $\underline{d}$ has been taken as the spacing of the grid on which the density values $\rho(y)$ are defined.

APPLICATION OF THE PROCEDURE TO IMPROVE ELECTRON DENSITY

The method was applied to tyrosyl-tRNA synthetase, a crystalline protein whose asymmetric unit contains one sub-unit (molecular weight 45000) of the dimeric molecule. Three isomorphous derivatives had been used to obtain phase information to a resolution of 2.7A though beyond 3A measured structure amplitudes were weak. Two substantial fragments of the amino acid sequence had been determined but these had not been recognised in the electron density. Application of two rounds of density modification lead to substantial retracing of the chain connectivity and to the alignment of chemical sequence for 321 residues.

APPLICATION OF THE PROCEDURE IN REFINEMENT

Electron density map of tyrosyl-tRNA synthetase has an ordered domain corresponding to about 320 residues and a disordered domain corresponding to approximately 100 residues. Application of the density modification procedure lead to an interpretable map for the ordered region and some improvement in the disordered region. Attempts to improve the model of ordered region alone by the conventional least-squares procedure (Konnert, J H & Hendrickson, E A, Acta Cryst, 1980, A36, 344-356) with fixed occupancies was not successful as R-factor remained above 45% even after 12 cycles of refinement. The structure was then refined to an R of 31% (3A) with the starting occupancies (determined by step of density modification) and with a fixed structure factor contribution coming from a fairly large peak (picked up by step 3(b) of density modification) from the disordered region. Structure factors coming from this peak corresponds to approximately 10% of the total calculated structure factors.

CONCLUSION

The density modification procedure is a good friend to approach when your war with heavy atom derivatives is hard to win - doesn't matter even if your computing grant application has been rejected and you can barely afford a few FFTs.

ROTATION CAMERA FILM PROCESSING AT LEEDS

R. Cooper and N.P.E. Gammage
Astbury Department of Biophysics
University of Leeds

The hardware in Leeds for processing rotation camera photographs consists of a PDP 11/45, 3 disc drives, magnetic tape deck and a Joyce-Loebl Scandig 3 rotating drum film scanner which utilises a standard D.E.C. DRS 11E interface.

Films are digitised in an offline mode and stacked up on magnetic tape; each film taking up approximately 200 ft of tape. A modification to the Joyce-Loebl supplied software enabled this to be done more efficiently. The program now writes a block of data on tape equal in length to the number of pixcels read in a film stripe (this is currently 1100). Digitisation of one film takes approximately $6\frac{1}{2}$ minutes.

Spots for orientation refinement are measured manually from still photographs and input to the program IDXREF. This program was modified by Dr A.J. Wonacott (Imperial College) to enable us to index and refine cells of triclinic geometry. This program is run on a VAX 11/780 for speed and convenience, taking up to 200 spots from 7 stills.

The spot generation program GENVEE, next in our suite has caused us (and others since!) many problems. On some film packs, depending upon missetting angles, setting matrix, and resolution limit, the program loses itself and so generates 3 non-existant spots ad infinitum. Modifications by Alan Wonacott have included the addition of a parameter to swap the scan and step directions earlier so that the program never reaches this "point of no return". This is now under hardware switch control. Frills added in Leeds have been a hardware abort switch to terminate a "lost" generate cleanly (some generates hang up at or near the end and are therefore usable), and some graphics routines to display the spots (and flash the partials) during generation.

At this stage two Bristol written routines PLODCR (plot to screen) and PLODPL (plot to plotter) allow plotting of the generate file for comparison with the photographs. These routines have been extremely useful as the missetting angles always have to be altered manually to correctly predict the partials.

Integration of spot intensities is done by the program MOSCO (written in Cambridge, modified in Bristol). A films worth of data is read from the magnetic tape onto an empty RK05 disc. Spot positions from the generate file and spot intensities from the disc are matched up, refined and summed. Bristol modifications allowed for choice of background position and for this reason the Bristol raster parameters were used in the program GENVEE. Modifications in Leeds have been to allow a manual search for the fiducials marks on the film using the graphics and input of fog level when a fiducial is too large. MOSCO can be run in automatic mode, integrating a full tapes worth of data once the parameters have been set for the first film.

continued ...

The binary direct access generate file is now converted back into ASCII using a program written in Leeds called PREPAR. The data is then transferred via a HASP link to the University's AMDAHL 470 V7A, there being no magnetic tape link between the Biophysics PDP and the AMDAHL.

The data is scaled, sorted and merged using 4 Cambridge programs and 2 Leeds ones. The Cambridge programs are PROTIN, PAIROT, ROTAVATA and AGROVATA all of which are well documented. The Leeds programs are used for sorting 3A1 and 9A2 binary files because the Leeds system does not have the standard IBM sort package programs.

In order to use this system more efficiently a device driver for the Joyce-Loebl scanner is presently being written and it is hoped to use this with the RSX11M multi user operating system.

The scanning of the films is at present carried out using software supplied by Joyce-Loebl under the single user operating system RT11, density data from the scanner being transferred directly onto magnetic tape, and as such undergoes little processing. The processing is carried out under DOS/BATCH (another single user system) but can be left running over night.

However, the data collection is a lengthy process which ties up the computer for long periods of time during the day - due to constant operator intervention required to change the films. The same data collection system could be run far more efficiently under the multi-user system RSX11M which is now a fully functional operating system at Leeds. Work is well in hand on the writing of a device driver to interface the Scandig scanner in real-time with the RSX operating system. This will allow processing to be carried out in parallel with data collection and program development increasing the throughput of films.

The driver makes use of non processor requested data transfer and so requires very little processor intervention, and as the scanner operates under interrupt control the processor is free to continue with another task whilst it awaits input from the scanner. The driver is written in assembly language but so as to be easily accessible it is called by a set of FORTRAN routines that can be incorporated into a user's program.

Eventually the driver will be incorporated into a FORTRAN program which will emulate the Joyce-Loebl assembly language software and also to provide the basis for on-line scanning of films.

SRS PROTEIN CRYSTALLOGRAPHY WORKSTATION ON THE AIR

John Helliwell (Daresbury/Keele)

After the necessary period of commissioning the instrument, the workstation is now available to users. An intensive period of work this summer enabled on-line control of alignment to be made both of the monochromator and Arndt-Wonacott camera. With the SRS commencing at the lowly figures of 1.5 GeV and 30 mA current we were able to accurately set the wavelength using calibrated metal foils (e.g. nickel, cobalt). Also by using a narrow slit at the focus and the monochromator set at the Guinier condition, a narrow $\frac{\delta\lambda}{\lambda} < 10^{-3}$ could be obtained adequate to resolve the near edge absorption structure (white line) - see figure 1 (Dysprosium LIII edge). Then as the SRS increased power (1.8 GeV and higher currents) a detailed investigation of the measured flux versus calculated flux revealed a 1500% discrepancy (wrong way!) which was resolved because of unnecessary absorption effects - mainly a spurious layer of vacuum grease on a beryllium window. Presently we are within 30% of the calculated value.

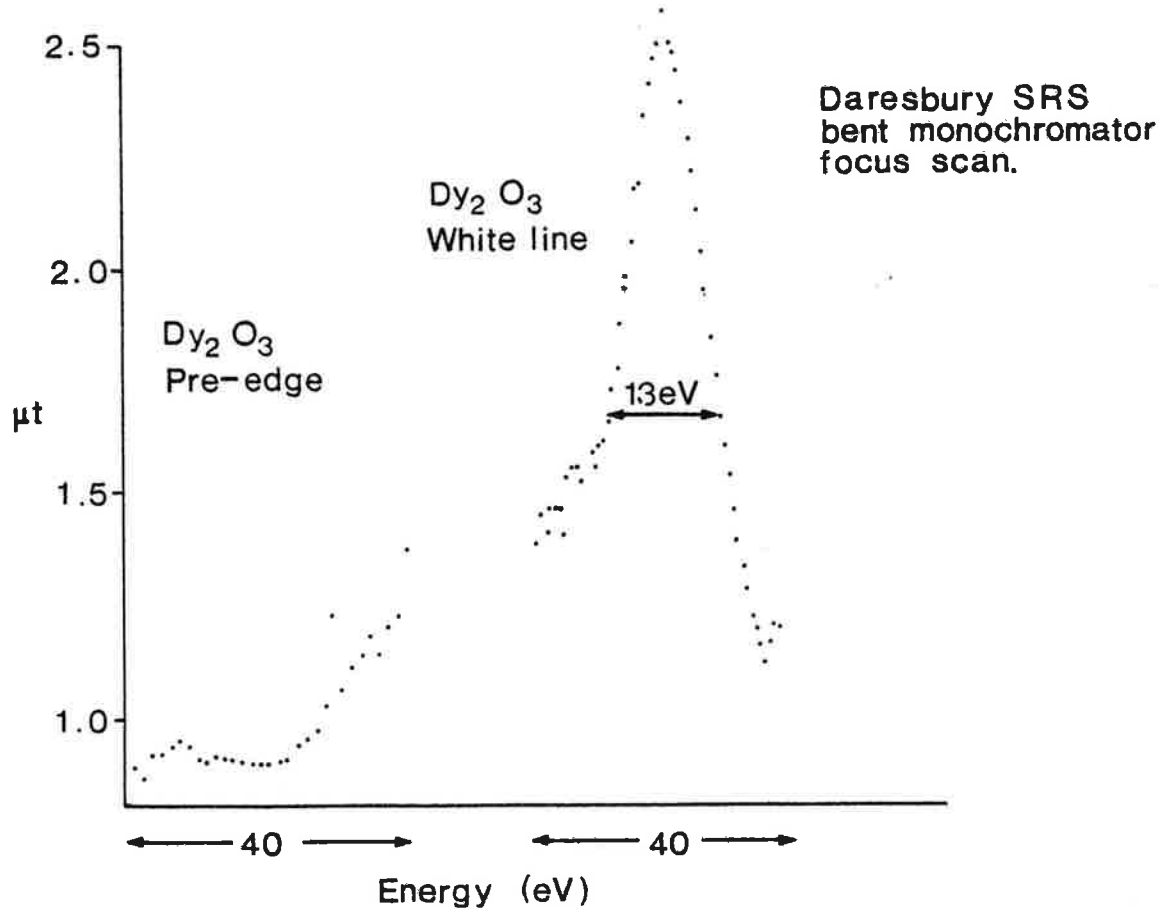
It was at this stage, with the SRS performance becoming acceptable for routine use, that outside users could come and collect worthwhile data. The first (decent!) photograph was obtained by the Oxford group (M Adams, E Brigg, S Gover and R Pickersgill) - see figure 2 showing a v-shaped cassette photograph from a single crystal of 6-phosphogluconate dehydrogenase, spots out to 1.87 Å, 20 minutes exposure time. Subsequently, data has so far been collected at room temperature from glyceraldehyde-3 phosphate dehydrogenase (A J Wonacott) and Rabbit F_c (B Sutton). Work by the Birkbeck Group (G Taylor) was unfortunately impeded by sample cooling problems, which are being sorted out.

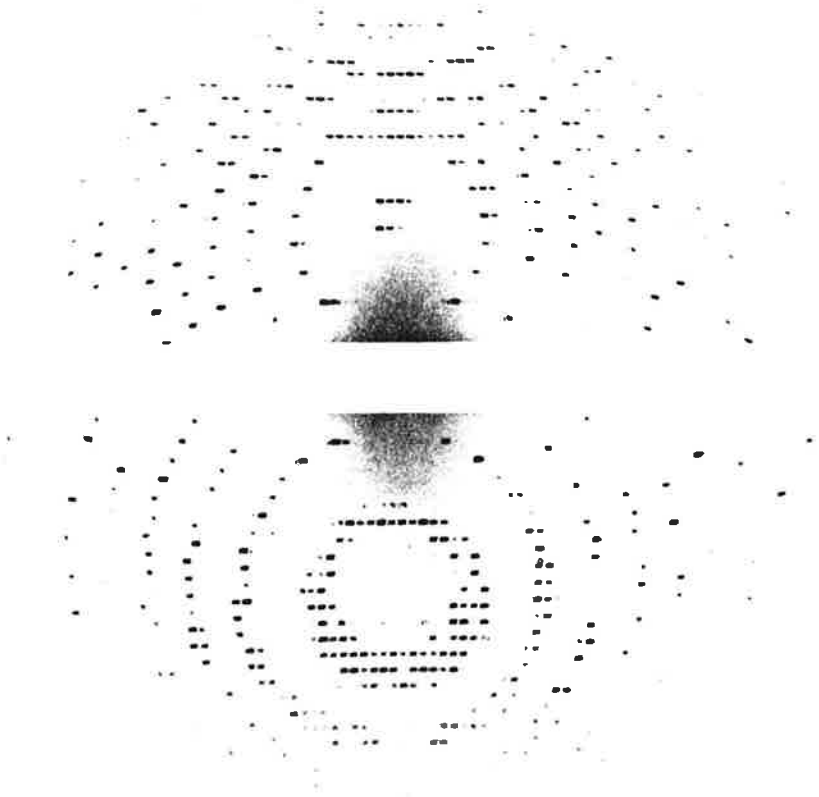
The SRS is presently (September 1981) operating at $\sim$ 1.87 GeV and currents up to 170 mA at injection giving about 40 x the GX6 rotating anode with a graphite monochromator (or $\sim$ 15 - 20 Ni filtered GX6).

We now look forward to a productive period of data being collected on the workstation at the SRS.

Figure 1

Absorption edge scan across the monochromator focus at the Guinier position revealing the white line near edge structure in Dy_2O_3 (L III edge, 1.59 Å)





VEE SHAPED CASSETTE PHOTOGRAPH FROM A SINGLE CRYSTAL OF 6-PHOSPHOGLUCONATE DEHYDROGENASE FROM SHEEP LIVER. SRS OPERATING AT 1.6 GEV, 80 MA, EXPOSURE TIME 20 MINUTES, RESOLUTION LIMIT 1.87 Å, ROTATION ANGLE INCREMENT 1.25°.

REVISION OF THE FFT PROGRAMS

Following discussions at a CCP Working Group Meeting in May 1981, Phil Evans agreed to "clean-up" the existing FFT programs while implementing them on the MRC (Cambridge) VAX computer. This work is now almost complete and as well as thanking Phil, I though it might be of interest to include some details of the new versions here.

Pella Machin

Fourier calculation routines : an implementors guide

This version of the FFT programs was put together by Phil Evans form the original FFT programs by Lynn Ten Eyck, with extra space-group routines by Gerard Bricogne and Eleanor Dodson, and front-end routines by Ian Tickle and John Campbell.

The function of the front-end routines is to allow input column selection, taking of differences, weighting etc, and symmetry expansion and permutation to increase the flexibility of the programs.

The subroutines are divided into four files:-

1. FFTMAIN a short main program to call the the other routines
2. FFTSGP Space-group specific master subroutines, and routines for reading and writing the direct-access scratch file.
3. FFTFSUBS General routines, front end routines, and map output routines.
4. FFTTRNFM One-dimensional Fourier transform package (Ten Eyck). This is also used for the reverse transform.

In addition, the following general routine packages are used.

(a) LCF routines : reflection input

(b) Map output : routines from MSUBS and DISKIO

Organisation of the calculation

1. The main program (FFTMAIN) calls subroutine CARDIN to read all the input control data. All input is read using Fortran free format routines GTNUM etc (in FFTFSUBS), which interpret a line read as 80A1. The input specifies the parameters needed for the front-end routines,to set up the required Fourier coefficients (selection of input columns, symmetry expansion, permutation etc) and parameters for the Fourier transform (index limits, sampling intervals, map limits, scaling etc).
The structure factor input is taken from a binary file in the Daresbury LCF format (although it may easily be changed to other data formats). Subroutine CARDIN sets up the column assignments, calling subroutine GTASGI to set up the columns required, and subroutine RHLCF1 (from the LCF package) to assign the columns to the file.

2. The main program then calls the subroutine FFTn, where n is the space-group number. This routine FFTn controls the Fourier calculation. The present main program uses the space-group number read by CARDIN to branch to the appropriate subroutine call, but for other computers it may be better to link the program with a different main program for each space-group, to avoid loading code for all space-groups on every run.

3. The routine FFTn calculates how many levels of constant h can be taken together in the first pass (P1), and how many output sections

C can be generated together in the second pass (P2). These numbers P1 &
C P2 depend on the array size, the maximum indices, the map sampling
C and the space-group. The record length and the number of records in
C the direct access scratch file are calculated.
C The assignment of te direct access file is liable to be machine-
C dependant. In particular, on the IBM 370 the record size etc in a
C DEFINE FILE statement have to be constants, ie set at compile time,
C which is awkward. These programs call a Rutherford library subroutine
C DEFINU (called from FILEO) which is a run-time DEFINE FILE, and an
C equivalent routine DEFINU is provided for the VAX in file FFTFSUBS
C (using an OPEN statement).
C
C 4. Three steps are then repeated as required to calculate the first
C Fourier transform.
C
C (a) Read a block of structure factors for some h (P1 levels), all k
C and all l. Subroutines SFIPYC, SFIPYR, SFIPZC, SFIPZR, and SFIIYC
C are called depending on the section axis and whether the space-group
C is centric or not (SFIIYC is special for body-centred I222), and
C these routines in turn call GETHKL and GGETF to read the file and
C prepare the Fourier coefficients.
C
C (b) Calculate the Fourier transforms along k (or along l if
C z-sections are being calculated).
C
C (c) Write the intermediate results to the direct-access scratch
C file (subroutines YOUT... , ZOUT... etc). For space-groups up to
C orthorhombic symmetry, these files are written in blocks which
C contain all of the available values of h, P2 values of k (or l),
C and all values of l (or k). If the space-group is one of the
C orthorhombic space-groups for which the intermediate results for
C -h k l are derived from +h k l by symmetry relations, the symmetry
C relations are applied and the intermediate results for -h k l are
C also written. For tetragonal and trigonal space-groups, the scratch
C is organized differently: each record corresponds to one plane of
C the intermediate transform, and data may be patched into these
C records in different passes.
C
C 5. The scratch file is positioned at the beginning, and three steps
C are repeated as necessary to calculate the other two transforms.
C
C (a) Read through the scratch file picking up all of the values of h
C and l (or k) for the range of k (or l) values currently being
C processed. The data are placed in memory in such a way that after
C the last transform the resulting real numbers are stored by planes
C of electron density.
C
C (b) Calculate the remaining two transforms
C
C (c) Write out the electron density sections (subroutines PRINTY,
C PRINTZ). If this is the first block of sections, calculate the scale
C factor if required.
C See write-up MAPFORMAT.DOC for map format. Routines from MSUBS
C are called as follows: MWRHDR sets up the map header; MSYPUT copies
C symmetry data for the appropriate true space-group from the library
C file assigned to SYMOP (stream 4) to the map file; MSPEW or MWRSEC
C write out one section to the file; and MCLOSE writes the header
C block and closes the file.

Density generation & structure factor calculation

This collection of routines was put together by Phil Evans from Eleanor Dodson's density generation program (GENDEN) & Lynn Ten Eyck's reverse FFT programs, with additional space-groups by Gerard Bricogne and Eleanor Dodson. It is set up so that the routines may be linked together to do density generation, or structure-factor calculation (SFC), or both (GENSFC), depending on the availability (or desirability) of virtual memory etc..

The subroutines are grouped into 6 files.

1. GENMAIN A small main program to set the working array size, and the maximum record length allowed for the scratch file.
2. GENDEN Density generation routines
3. FFTSFSGP Space-group specific routines for structure-factor calculation, including routines for writing and reading the direct-access scratch file.
4. FFTSFSUBS Other subroutines, some of them space-group dependent, including routines for reading the map and for writing the structure-factor file.
5. FFTTRNFM One-dimensional FFT routines as for Fourier synthesis.
6. SFC Main program and routines for structure-factor calculation without density generation (to be linked with 3,4, & 5).

In addition standard routines are needed

(a) from the LCF package for reflection input and output.

(b) from MSUBS and DISKIO for map input and output.

Flow of calculation for GENSFC

1. GENMAIN calls the master routine DATRED.
2. DATIN is called to read all the control cards (stream 5) and to set up the form-factors.
3. Subroutine SFSET (FFTSFSUBS) is called to set up the parameters for the structure-factor calculation. The [input and] output LCF files are opened.
4. Subroutine SGTEST sets up the space-group specific parameters: axis order, map limits, the number of groups of sections and layers of h for each pass of the transform (P1 and P2), and the record length and number of records for the direct-access scratch file. If the record length is longer than the maximum allowed (which may happen,particularly on the VAX, with very small problems or very large arrays), the array size to be used is reduced (function routine LRCSIZ in FFTSFSUBS), and the parameters are recalculated.
5. Subroutine SFCTL (in Genden) is called. This calls ATSORT (Genden) to read the coordinates file, generate any symmetry related atoms required for the map, sort them into order on the section axis, and write them

to a direct-access scratch file. This file is written and read by subroutine BUFIO (in Genden) which buffers the short atom records (5 words) into 2000 word blocks for efficiency.

6. SFCTL either calls GENDEN to generate the map and PRINTZ to output it, or it passes control to SFCALC (FFTSFSUBS) to calculate structure-factors.

7. For structure-factor calculation, SFCALC branches to call a space-group specific routine SFCn, where n is the space-group number. It may be necessary to change this way of calling the routines for other machines, in order to link routines for just one space-group at a time: eg all the SFCn routines could be given the same name but stored in different files, and picked up selectively at link time.

8. The rotines SFCn (in FFTSFSGP) control te reverse FFT. For the first two transforms, they call either GENDEN to generate a slab of calculated map, or (for SFC when no density generation is being done) REDMAP to read a slab of map from a file (MAPIN). The intermediate transform is written to the direct-access scratch file.

9. The final structure-factors are written to an LCF output file by the routines OUTSFx (OUTSFY for y-sections, OUTSFZ for z-sections, OUTSFI for body-centred y-sections (I222)). These routines either output the whole array of structure-factors within the resolution limits, or read an LCF file of 'Fobs' and output only those reflections for which there is an entry in the Fo file, and which lie within the resolution limits.

Note that there are two space-group specific parts of the reverse FFT routines. The subroutine SGTEST has a piece of code for each space-group, and the routine SFCALC calls the specific routines SFCn.

For structure-factor calculation only (SFC), the routines in file SFC consist of a cut-down DATRED and DATIN (cf GENDEN) which call SFSET etc and then call SFCALC to control the FFT

GROUP MEETING OF PROTEIN CRYSTALLOGRAPHERS OF NORTHERN ENGLAND

Pella Machin (Daresbury)

Another one-day meeting for Protein Crystallographers (PCONE) from Daresbury, Keele, Leeds, Sheffield and York was held in the Biophysics Department of Leeds University on 29 September 1981 organised by Maxine McCall. This was the third meeting in the series of six monthly meetings, aimed at promoting communication and collaboration between the groups.

During the morning various reports were given by people who had attended International meetings over the summer. Maxine McCall (Leeds) began by describing talks given at the Buffalo meeting on small molecule interactions with DNA. Pauline Harrison (Sheffield) reported on some of the structures presented at the Ottawa meeting and Colin Reynolds (York) described the Ottawa session on protein refinement. To finish off this section Tony North (Leeds) summarised some interesting developments in Biophysics reported at the International Biophysics meeting in Mexico City.

Two specific topics were discussed in the afternoon. Firstly John Helliwell (Keele/Daresbury) gave details about the current state of the Synchrotron Radiation Source (SRS) and reported that data had been collected for five film data sets, all reasonably successful except for one which suffered from problems with the cooling apparatus. A discussion followed concerning scheduling of time on the SRS. The second topic concerned graphics systems and Sandy Geddes (Leeds) described the various systems he had seen in the USA while visiting laboratories after the Ottawa congress. Tony North reported on some graphics developments in the UK and a discussion followed on the type of system required here and the possibilities of setting up communal equipment.