DARESBURY LABORATORY

JOINT CCP4 and ESF-EACBM NEWSLETTER on PROTEIN CRYSTALLOGRAPHY

An Informal Newsletter associated with the SERC Collaborative Computational Project No.4 on Protein Crystallography and the ESF Network of the European Association of the Crystallography of Biological Macromolecules

Number 24 December 1989

Contents

Editors: Kim Henrick
Science and Engineering Research Council
Daresbury Laboratory, Daresbury
Warrington, WA4 4AD, England

Keith Wilson
EMBL c/o DESY
Notkestrasse 85, D-2000 Hamburg 52
Federal Republic of Germany .

Protein Crystallisation on the JUNO Mission

David Blow

Blackett Laboratory, Imperial College, London, SW7 2BZ

The Juno mission is a Russian–launched rocket mission to the Mir space–station, with a specific UK involvement. The flight, in April 1991, will carry a UK astronaut and a number of UK–designed experiments. The mission is expected to stay in microgravity for 5 days.

The Imperial College group have proposed a protein crystallisation experiment on Juno which has a number of special features:

— a large number of small individual samples;
— use of an automated microdispenser;
— an enclosed batch technique for crystallisation;
— thorough controls to evaluate effect of microgravity.

Based on the experience that the ideal conditions for crystallisation are very difficult to reproduce, we consider that a successful protein crystallisation experiment needs to use **multiple samples** so that a range of values can be covered for several parameters. To keep the experiment small, and to minimise consumption of protein, small samples are used down to about 2 microlitres.

These samples are dispensed with great accuracy and with a minimum of labour by using an **automatic microdispenser** system which has been developed in our laboratory.

Using the automatic microdispenser, we find that the techniques of vapour diffusion and microdialysis, commonly used in protein crystallisation, are not necessary; good crystallisation can be obtained in a single drop. This is known as the **batch** method of crystallisation. For microgravity crystallisation, two separate drops will be dispensed before the flight, which will be merged by centrifugation when the microgravity environment is entered. In microgravity, the sample will float as a spherical drop in the oil. Paraffin oil and silicone oils are both highly insoluble in water, and no ill–effects

on crystallisation have been found in their use. The technique as currently used would not be appropriate for intrinsic membrane proteins, and perhaps not for soluble but highly lipophilic proteins.

In order to evaluate the effects of microgravity on crystallisation, we propose two series of **control crystallisations** at the same time as the microgravity experiment. One control series will follow, as far as possible, exactly the same environmental history except that it does not travel in the rocket and is in a 1g environment at all times. A second control crystallisation series will be maintained under optimum conditions in our laboratory. In this series the samples will be mixed as they are dispensed, and it will not be necessary to mix two separate drops at a later time.

Crystals grown in the experiments will be evaluated by microscopic examination, and the best crystals of each protein will be compared by X-ray diffraction techniques. Microscopic examination will show the size and external perfection of the crystals. X-ray diffraction will be used to evaluate the internal perfection (twinning or other crystalline inclusions), and internal degree of order (B–factor and mosaicity) of crystals.

Timetable

In order to ensure that the experiment runs smoothly, earlier trials will be carried out before the main experiment.

January–December 1990: preliminary trials.

Every protein will be subjected to preliminary crystallisation trials in the microdispenser, to be sure that appropriate conditions of supersaturation are reached, and if possible that crystals of some kind can be grown. Any particular problem of the batch method and the microdispenser in relation to a specific protein will be resolved at this stage.

January 1991: advance trial.

In the advance trial a complete series of samples will be dispensed and will have the same environmental history as the microgravity sample is anticipated to have, except that it will be maintained at 1g throughout. This will allow a realistic evaluation of the time needed to prepare the samples, and a rigorous check of the technique for merging

the two drops under all the conditions to be used. Any problems due to the long period of incubation of the separate drops before they are exposed to crystallisation conditions, etc should emerge at this stage. Further adjustments to crystallisation conditions will be possible when the results of the advance trial are assessed.

April 1991: microgravity experiment; 1g control experiment; laboratory control experiment.

The microgravity and 1g samples will be dispensed into special sample–holders designed to allow the mixing of two drops, while the laboratory experiment will use standard sample holders.

May–July 1991: evaluation of results.

All samples will be rerturned to London and examined microscopically and by X-ray diffraction. For evaluation of crystal mosaicity special X-ray facilities will be required, probably at the Synchrotron Radiation Source, Daresbury, UK.

Sample requirements

Protein will be required for preliminary trials, advance trials and the three experimental series. Each sample will normally be evaluated by gel electrophoresis on receipt and the quantity of protein confirmed by A280 measurement. Protein samples will be concentrated to small volumes as required using the Centricon system. Each crystallisation sample will be dispensed by mixing up to four solutions which are typically protein, precipitant, and up to two additives.

Recommended sample size is 2 to 15 microlitres, giving drops of diameter 1.5mm to 3mm. The samples will be dispensed into trays of about 100 samples, using at least 5 trays in the microgravity experiment. We hope for at least 50 samples of each protein.

For example, 50 samples of 2 microlitres, each containing 10 mg/ml protein, would require 1mg protein. For the preliminary trials, advance trials and three experimental series, the total amount of protein would be the order of 5mg, of which 3mg would be required in April 1991.

Crystallisation conditions

The samples will be transported to microgravity at 4°C. Crystallisation at a controlled 4 deg environment will be the standard; another temperature, possibly 18°C, may be available as an alternative. We have stipulated that the temperature control must be maintained during descent and until the samples are returned to the laboratory for evaluation.

During re–entry and landing, the sample trays will be exposed to considerable accelerations and vibrations, but the sample drops are surrounded by a viscous oil, which will cushion them from these shocks.

The Juno mission itself is planned to last only 5 days, but there is some possibility that this crystallisation trial may be carried one way on the Juno flight and in the other direction on another flight manned entirely by USSR personnel. In this case the microgravity environment would be maintained for about a month. We have stated that the longer period for crystallisation would be advantageous.

Call for protein samples

We are inviting protein crystallographers, and interested protein chemists, to offer samples for trials of microgravity crystallisation. It is important that general conclusions about the value of microgravity techniques for obtaining large or perfect crystals be based on the widest possible range of proteins. A small number of the crystals of each protein grown during the flight will be used for tests of crystal perfection, but the others can be returned to the supplier after microscopic examination in our laboratory.

Proteins will be accepted subject to preliminary trials during 1990, as explained in the Timetable above. In the event that selection is necessary amongst the samples offered, a final decision will be made after the preliminary trial.

Offers of samples should be made to Professor David Blow, Blackett Laboratory, Imperial College, London SW7 2BZ

Some Experiences with Protein Data Collection using the Hamburg Image Plate System

Andrew G.W. Leslie and Penny E. Stein

MRC Laboratory of Molecular Biology, Cambridge UK

ANDREW@UK.AC.CAMBRIDGE.MRC-MOLECULAR-BIOLOGY

The image plate system at the EMBL outstation at DESY, Hamburg, is based on the Fuji image plate and the Hendrix-Lentfer rotating disc scanner, with control software by M. Boehm (see EACBM newsletter No 5, pp4-7, Feb 1989). It has been in operation for about one year, and has already established itself as a very attractive method for collecting high-quality protein diffraction data. The experiences described below relate to the collection of two datasets, both to 2 Åresolution, using the scanner on the X31 beamline. The first dataset was from a mutant of chloramphenicol acetyltransferase (CAT), (space group **R32**, cell parameters $a = 72.5$ Å, $\alpha = 94.5°$) and the second from native ovalbumin (space group **P1**, cell parameters $a = 62.9$, $b = 84.7$, $c = 71.5$ Å, α =87.5, β =104.0, γ =108.5°)

The system itself is very straightforward to use, thanks to features such as the automatic beam alignment and the user-friendliness of the controlling SCRAP software. The first step, as with film data, is to record a trial exposure which is displayed on the workstation monitor and this is used to judge the diffraction limit and the required exposure time. The quality of the display is such that it is very similar to looking at an exposed X-ray film, making decisions about the strength of diffraction relatively easy. Once the exposure time, resolution and rotation range have been determined, the user simply requests the number of images to be collected (each image corresponding to an oscillation "film") and the data collection proceeds automatically. For every image, the shutter is opened during the crystal rotation, then closed while the image plate is scanned and the digitised image written to disk, and finally the image plate is erased ready for the next exposure.

With the present scanner, the time required to scan and then erase the image is about 2.5 minutes, giving an overall cycle time of about 5 minutes for a "typical"

exposure time of 2–3 minutes. The beam utilisation is of course increased for weakly diffracting specimens which require a longer exposure, but is correspondingly reduced for strongly diffracting samples or when collecting a low-resolution dataset (see below). The scanner is therefore rather well matched to the beam intensity on the X31 beamline (which is approximately 20 times weaker than station 9.6 at Daresbury).

The obvious attraction of the system is that it runs essentially automatically. The only user intervention required is to open the main port shutter after an injection (which was typically every 2–3 hours when operating on X31). The elimination of film loading and developing saves a considerable amount of labour, and means that the station could be run single-handed, particularly if the crystal lifetime allows a large number of images to be collected from each crystal.

The principle disadvantage at present is the limited dynamic range. Essentially, each image ressembles the "A" film of a film pack, so that when collecting high resolution data (eg 2 Å) a large percentage of the lower resolution reflections are overloaded. The limitation is not due to the intrinsic dynamic range of the image plate, but the use of a 14 bit Analogue to Digital converter, limiting values to the range 0—16383. In practice, this means that a second data collection run is usually necessary, using a much shorter exposure time (a factor of 16 for the CAT data) and repositioning the detector to collect only the low resolution data that was overloaded in the first run (For CAT and ovalbumin, the second run collected data to 2.7 Å). This significantly increases the total time required for data collection, since although the exposure times may be less than 20 seconds, there is still the 2.5 minute overhead for scanning each image. Generally a larger rotation range can be used for the low resolution data, reducing the number of images required, but obviously it would be preferable to avoid this second run altogether. A possible solution would be to reduce the overall gain of the system (currently 1 to 2 X-ray photons per ADC unit at 8KeV). Reducing the gain by a factor of (say) 20 would extend the dynamic range by the same amount. In order to try to assess the effect of a reduction in gain on the quality of the data, the 2A CAT images were reprocessed performing an integer division by 20 on every pixel prior to processing (ie truncating to the nearest multiple of 20). The resulting merging R-factor (from AGROVATA) was

6.7%, identical to that obtained with the original images, suggesting that a reduction in gain by a factor of 20 would not significantly affect the data quality.

Apparently the "Mark II" scanner, which should become operational early in 1990, will avoid the requirement for the second data collection run.

A secondary disadvantage, though far less serious, is that it is not possible to examine an image without halting data collection. This was not a serious problem for CAT or the ovalbumin crystals, which have a relatively long lifetime in the beam, but in many cases it would be valuable to have some way of monitoring crystal decay continously, and the simplest procedure would be to allow the display of each image once it had been scanned but without interupting data collection.

Data Processing

As data collection on the image plate system mimics film data collection, it is possible to use existing film processing software to process the images with relatively few modifications. (In the case of MOSFLM, which was modified by Howard Terry and Sandra McLaughlin at EMBL, the biggest change involved working with an image in which the counts are stored as integers rather than bytes). This means that the user does not have to face an entirely new software package in order to process the images, which has obvious advantages, although it does mean that the system does not take advantage of the improved temporal resolution offered by other area detectors which work with much smaller (in phi rotation) image sizes. The availability of reliable auto-indexing programs has greatly simplified the problem of determining the crystal orientation (the REFIX program written by Wolfgang Kabsch, modified by Howard Terry to interface it to the MOSFLM suite, is used at Hamburg). In fact it allows the crystal orientation to be determined very rapidly from the first "trial" exposure, so that the PHI angle can be adjusted to collect the required data. The image plate also eliminates fiducials and variable camera constants (which are probably the greatest single cause of problems in film processing).

Clearly, any data collection system which simplifies the practicalities of collecting data at the expense of a reduction in data quality is only going to be of limited applicability. However, the results obtained so far strongly suggest that the Hamburg image plate system does not have this drawback, and data of a quality equal to that

possible with film methods is possible. (A comparison with other area detectors is not strictly valid as the image plate system works with relatively large rotations per image, which inevitably degrades the signal to noise relative to systems using 0.1 or 0.2 degree rotations.) In the case of the CAT data, the final merging R-factor (on intensities from AGROVATA) was 6.7% (9.9% for data between 2.00 and 2.10 Å). This compares with an R-factor for the best film data (to 2.2 Åresolution) from a different CAT mutant of 3.1% , collected on the Wiggler line by Michael Gibbs. Although the R-factor is higher for the image plate data, it is essential to note that the analysis of the standard deviations provided by AGROVATA suggests that in both cases the errors in the data are dominated by genuine counting statistics. This implies that the R-factor for the image plate data could be reduced by simply increasing the exposure time. The merging R-factor for the strongest reflections (5.4% of the total number of observations) in the 2.7 ÅCAT images is 2.4%, whereas the corresponding value for the best film data is 2.6%, supporting the view that the image plate data can equal film data in terms of quality. Unfortunatley no data have been collected on both film and the image plate for the same CAT mutant, so it has not been possible to do a direct comparison in order to detect any systematic differences in the intensities.

The statistics for the ovalbumin data are very similar: the overall merging R-factor for the 2 Åimages alone was 6.8% (11.8% at 2.0 Å), for the 2.7 Åimages alone 4.1%, and the R-factor for merging the 2 Åand 2.7 Ådata was 5.2% The 2 ÅOvalbumin data was compared with a 2.5 Åresolution dataset collected on a FAST diffractometer on a rotating anode source. The R-factor on F was 8.2% overall, and only varied slowly with resolution (7.7% at 10 Åand 10.8% at 2.5 Å). The average (delta F)/F was between 0.3% and 3.4% (greatest at very low and very high resolution). There was no suggestion of any significant systematic differences between the two datasets.

A number of practical points are perhaps worth mentioning in connection with processing the image plate data:

1) Because the image plate is physically larger than film, greater accuracy in cell parameters and crystal orientation are required to achieve the same positional accuracy in predicting the positions of spots on the image plate. To avoid systematic errors when

profile fitting it is essential to use the best possible values of the cell parameters. In the case of the CAT images, a positional residual of 25-30 microns between predicted and observed spot positions was achievable (ie only slightly worse than with an "A" film), but a change in cell parameters of less than 1 part in 1000 would noticeably increase the residuals. The optimum values were determined from the POSTCHK post-refinement program using abutting images. This is possible with CAT because there are only two independent cell parameters (space group **R32**). However, for ovalbumin (**P1**) this approach is not feasible, because not all the parameters will be well defined if data from a single pair of abutting images is used. In such cases, data from pairs of images separated in rotation angle by a minimum of 10 degrees (and preferably at least 20) must be used in POSTCHK, but this procedure will only work if there has been no crystal slippage. For this reason, if the true cell dimensions are not known very accurately, or if there is any likelihood of variation in the cell parameters between different crystals, it would seem worthwhile collecting two "still" images, so that the IDXREF program can be used to determine the cell parameters. This is particularly true when dealing with low symmetry spacegroups.

2) For reasons which are not entirely clear, the autoindexing program REFIX can give solutions in which the orientation is in error by 0.1 degrees or more, significantly larger than the errors we have found when testing the program on films from crystals belonging to a variety of spacegroups. The cell dimensions can also be in error by up to 1% because (unlike IDXREF or POSTCHK) they rely on an accurate knowledge of the crystal to detector distance. While a solution from REFIX is generally acceptable when processing the image from which the solution was obtained, it is frequently found that when processing a series of images the positional residuals increase steadily through the series. This is a reliable indicator that the cell parameters or orientation (or both) are in error. Post refinement is then necessary. In some cases, the convolution method of refining the orientaion within MOSFLM gives a better starting orientation. Unfortunately the radius of convergence of POSTCHK is rather small, and so it may be necessary to do several rounds of OSCGEN,MOSFLM and POSTCHK before a stable solution is obtained.

In conclusion, in our experience the Hamburg system has proved a very convenient and attractive method for collecting high resolution data of a quality comparable to that obtainable using film methods. There do not yet seem to be any grounds for claiming that the image plate actually gives higher quality data (at least in the case of CAT), in spite of its inherent advantages (no chemical fog or Wooster effect, and the ability to work at greater crystal to detector distances where the background level is lower relative to the diffracted intensity). However, it may well prove to be the case that future improvements in both hardware and software will eventually allow the full potential of the image plate to be achieved.

We would like to thank all the personnel at EMBL, Hamburg, for their friendly and invaluable support during the CAT and ovalbumin data collection. AGWL would also like to acknowledge fruitful discussions with Colin Nave (Daresbury Laboratory) and Phil Evans and Frank Mallet (LMB) on the problems of the dynamic range of the image plate scanner.

The Oxford XENTRONICS: two years as a multi-user facility.

Elspeth F. Garman and David I. Stuart.

Laboratory of Molecular Biophysics, University of Oxford
Oxford. OX1 3QU

Since October 1987 over 160 useable data sets from a large range of proteins have been collected using the Oxford XENTRONICS area detector and Rigaku RU200H X-ray generator (0.3x0.3 mm effective focus with graphite monochromator). To date 18 3-D structures of 11 different molecules (from 1.2 kD [for a peptide] to 104 kD [for Tumour Necrosis Factor] in the asymmetric unit) have been solved (e.g. Refs 1 to 4), and a large number of high quality difference maps (e.g. Ref.5) and general crystallographic data (e.g. Ref.6) on a variety of projects have been obtained.

The range of data sets obtained spans primitive cells with dimensions from 15.1Å to 293Å corresponding to crystal to detector distances of 7.5cm to 34cm. The highest resolution to which data have been measured, at a detector swing angle of 85°, is 0.9Å . The high resolution data were essential for the solution of the particular peptide structure under study, as it had resisted previous onslaughts using lower resolution diffractometer data.

The hardware has been modified in several ways since the installation of the detector system, in order to improve the data quality. As supplied, there was no adjustment to allow alignment of the system (a situation rectified on systems supplied after ours). We therefore built a simple, but very mechanically stable, alignment system, whereby the whole detector table can be rotated, and translated in x and in y, about a point beneath the monochromator. The table is also on adjustable legs to give the height degree of freedom. We optimise the settings by scattering the X-ray beam from 4 thicknesses of magnetic computer tape at the crystal position and monitoring the flux meter on the Xentronics Position Decoding Circuit, and rotating Bowden cables outside the radiation shelter to move the table.

The background air scatter falling on the detector has been reduced by 60% by shortening the distance from the end of the collimator to the crystal from 11mm to

4mm, and additionally changing the crystal to backstop distance from 27mm to 7mm. The backstop is mounted on a rod allowing adjustment of its distance from the crystal, and comes towards the crystal at an angle of 45 degrees. This enables free movement of the crystal tube (usually at chi=45 degrees) around the whole omega circle. The usual collimation consists of a 1mm diameter aperture 61mm from the crystal, and a 0.5mm aperture 17mm away. We have smaller collimators available, as well as a home made adjustable jaw (rotating tantalum jaws with a D shaped cross section) collimator, but we find that for most crystals our usual collimation is satisfactory.

Lastly, we have designed and built a simple device, which uses two Peltier cooling modules (one 75W and one 50W) to cool the laboratory compressed air supply (dew point -40°C) and provides stable temperatures down to 5 degrees Centigrade at the crystal. The advantages of the device as a whole has been its simplicity, cheapness, and temperature stability. This latter was a surprise, since we never put a feedback temperature control circuit on the device.

The Rigaku generator has given few problems. We normally run at about 70% full power (which is 5.4kW), and filaments and magnetic seals last on average 2,500 and 5,000 hours respectively. We operate the generator at about equal kV and mA, as our measurements show that for a given power, more X-rays are output for an increase in kV than in mA (e.g. 60kV 60mA gives higher output than 40kV 90mA).

As the area detector is used by 30 different researchers within the laboratory and several regular outside groups, it has been necessary to give a large amount of training and to establish strict experimental and processing protocols in order to maintain the integrity of the data and safeguard the equipment. For instance, we perform flood field correction and brass plate calibrations every time we change the crystal to detector distance. These protocols are available for general use, and will be sent over JANET on request (Elspeth@UK.AC.OX.BIOP). All experimenters fill in log sheets after their run giving the essential information on data collection and processing, and this has been very useful in monitoring the detector performance. Our present detector has been operating for 18 months, and the optimum detector bias has started to drift slowly upwards due to gradual poisoning of the detector gas. The bias is now checked fortnightly (N.B.

we use a strong diffraction spot from a crystal for this test, NOT the magnetic tape mentioned above, as the tape gives inelastically scattered X-rays of lower energy, for which the optimum bias is too high). However, the drift problem was first noticed when larger than normal r.m.s. errors in brass plate calibrations between observed and interpolated spots were reported by experimenters on the log sheets. The detector system is scheduled at the end of each month for the next month, so that experimenters know when they will be collecting data. On the whole this system of time allocation operates very well, as long as a degree of flexibility is retained.

We are currently using the XENGEN (Ref. 7) software processing package, and have found it very satisfactory over a large range of problems. In our experience the scaling is the weakest part of the programme suit, and in the case of very weak sparse data, we use an in house scaling package 3DSCALE (D.I.Stuart, unpublished). We do not align our crystals, relying on the autoindexing routines to determine orientation. Even with completely unknown crystal morphologies, we have found this a staisfactory way to proceed.

Several hardware developments made this year have enabled us to widen the scope of our in house data collection. We now have a very robust flow through helium path for use with crystal to detector distances greater than 18.5cm. This was designed and built in the laboratory, and the length is varied by inserting 12.5cm diameter sections of tube (plastic drainpipe!) into the central portion of the device. The front and back are sealed with mylar windows held by screw on flanges over O-rings. The path fits into the detector dovetail, and moves round with the detector if two-theta is changed. The backstop is situated in the air gap between the back window of the tube and the area detector, as near the mylar as possible. The backstop holder is attached to the top of the back of the tube, and allows the backstop to be adjusted vertically and horizontally (used for non zero two-theta). The helium path has reduced the background air scatter so that at a detector to crystal distance of 29cm, we have been able to cut collection times by a third while keeping the signal to background ratio constant.

We have constructed an extension tailpiece which enables the detector to be placed up to 50cm from the crystal for the study of very long primitive cells. The largest cell

length from which we have refined data so far is 293Å . These data were collected without changing our relatively simple X-ray optical arrangement of graphite monochromator and two collimating apertures. Obviously the resolution range for a single two-theta setting of the detector at such large distances is limited (e.g. 9Å at 32cm at two-theta=0), but we have still found this useful in characterising unknown crystal forms.

We have also installed an Oxford Cryosystems liquid nitrogen cooling system so that we can flash freeze our crystals. It is supported on a home built adjustable mount which is suspended across the roof of the radiation shelter. The Cryostream does not require a pressurised dewar, uses only 0.6 litres of liquid nitrogen per hour, has a flexible stainless steel transfer line, and is temperature stable to within 0.1K over our longest run with it (1 week). So far we have tried flash freezing crystals of three different proteins, and have successfully collected three data sets on one of these proteins (beta lactamase). In this case, the crystals lasted without apparent degradation for 48 hours in the beam, whereas their normal lifetime is on average 10 hours.

In conclusion, we have found our area detector data collection system very reliable, with 80% uptime (ratio of time actually collecting data / all hours in all days since installation * 100) to date. We have obtained consistently satisfactory R factors over two years, and the area detector has revolutionised our in house data collection.

References

1. E.Y. Jones, D.I. Stuart, and N.P.C. Walker. Structure of Tumour Necrosis Factor. *Nature* **338**, (1989) 225-228

2. J.B. Rafferty, W.S. Somers, I. Saint-Girons, S.E.V. Phillips. Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor. *Nature* **341**, (1989) 705-710

3. D.G. Brown, M.R. Sanderson, J.V. Skelly, T.C. Jenkins, T. Brown, E.F. Garman, D.I. Stuart, and S. Neidle. Crystal Structure of a Berenil-Dodecanucleotide Complex: the role of water in sequence-specific ligand binding. Submitted to EMBO November 1989.

4. P.D. Jeffrey. The Structure and Specificity of Immunoglobulins. D.Phil Thesis, University of Oxford 1989 (unpublished).

5. J. Martin. Molecular Interactions involving Glycogen Phosphorylase. D.Phil Thesis, University of Oxford 1989 (unpublished).

6. J.V. Skelly, D.A. Suter, R.J. Knox, E. Garman, D.I. Stuart, M.R. Sanderson, J.J. Roberts and S. Neidle. Preliminary Crystallographic Data for NAD(P)H Quinone Reductase Isolated from the Walker 256 Rat Carcinoma Cell Line. *J.Mol.Biol.* **205**, (1989) 623

7. A. Howard, XENGEN Version 1.3 1988. Unpublished Guide to XENGEN Area Detector Software Package.

The Implementation of CCP4 on the IBM 3090 at Glasgow

Neil Isaacs and Fran Young.

From: GACB24@UK.AC.GLASGOW.VMS1

With the support of IBM(UK) we are implementing CCP4 on the University's 3090 computer. THe primary aim of the project is to produce a working version of CCP4 n the 3090 and then to produce vectorised code for those programs which will benefit from this. THe software will be documented where appropriate for IBM users and made available to Daresbury for distribution.It is intended that this IBM version will maintain future compatability with other distributed versions of CCP4.

A secondary project is to implement CCP4 on an IBM workstation opr operating with windows under under UNIX. The aim here is to use the windows enviroment to link related programs and to develop a menu driven input to maximise the ease of use of the programs for the users

Both of these versions are using the UNIX version of CCP4 provided by Kim Henrick as the basis.

The Meiko Computing Surface at Birkbeck

Roman Laskowski, Doug Jones, Julia Goodfellow, David Moss and Ian Tickle*

Laboratory for Molecular Biology, Department of Crystallography, Birkbeck College, University of London.

(* to whom correspondence should be addressed.)

Introduction

This item describes the department's Meiko Computing Surface which is a parallel computer based on the INMOS transputer. It describes the machine's modes of operation, means of use, and limitations. It also outlines some of the applications that we have developed to run on it in parallel.

A transputer is a processor chip which has about 2 to 3 times the processing power of a microVAX II. You can think of each transputer as a fast stand-alone microVAX II - minus screens, keyboards, and disk-drives. Each is capable of independently running sequential computer programs.

A collection of transputers can be used either to run several different programs at the same time, or to run a single program in parallel. In the latter case each transputer either performs a different part of the algorithm, or the same algorithm but using different data. Parallel running requires that the transputers be able to transfer data and messages between one another. For this purpose, each has 4 high-speed serial communications links. You can use these links, to wire the tranputers together in a wide variety of configurations.

Tranputers come in essentially two varieties: the T414 which can only perform integer arithmetic and has 2 kbytes of on-chip memory; and the T800, which can perform floating-point calculations and has 4 kbytes of on-chip memory. The on-chip memory, although very small (about an A4 page of text), has very fast access, so is useful for storing frequently-used data. All other data is held off-chip in main memory.

Current Meiko System

Our Meiko Computing Surface currently has six transputers: one T414 which handles the communication between the Meiko and the BOOT microVAX II (acting as

a "front-end" to the Meiko), one T800 with 8 Mbytes of memory which controls the machine's 300 Mbyte internal hard disk, and four T800s with 4 Mbytes of memory each which can be used for running parallel programs.

The machine has two modes of operation: stand-alone and multi-user.

1. Stand-alone mode

In the stand-alone mode, only one person can use the machine at a time, and all file access, which is performed via the BOOT microVAX II, is relatively slow. The Computing Surface is booted from the user account and puts you into the "folding editor" of the Transputer Development System (TSD). Here you have the whole machine at your disposal, and can edit, compile, and run programs written in OCCAM - the native computer language of the transputer.

You can also use the machine to run sequential FORTRAN 77 or C programs on a single transputer. To do so requires compiling the programs with the appropriate Meiko compiler, having corrected for the minor differences between the Meiko and the VAX. Because the transputer is faster than the microVAX II, sequential programs will generally run faster on a single transputer (except where they are slowed down by a lot of disk I/O).

More interestingly, you can run parallel programs written in FORTRAN 77 or C on the machine, providing that they are embedded in an "OCCAM harness" capable of performing all the required communication between processors. Meiko have provided us with a standard OCCAM harness which makes it relatively straightforward - though not necessarily easy - to convert a sequential program into a parallel one.

2. Multi-user

The other mode of operation is the multi-user mode. This currently allows either one user to have 4 transputers or up to 3 users to have a transputer each. In this mode, file access can either be via the BOOT microVAX II or to and from the Meiko's internal 300 Mbyte hard disk. File access on the latter is much faster.

An advantage of this mode is that is allows jobs to be submitted to the Meiko from a batch queue on the BOOT microVAX II. These can either be parallel programs if the

machine is configured for a single-user, or sequential programs if it is configured for 3 single-transputer users. One can, for example, submit several sequential programs to the queue, each of which will be run on a transputer as soon as one becomes free.

Limitations

The main limitation of our current machine is the small number of transputers that is has. Other academic users typically have at least 16 transputers in their machines, and benefit accordingly.

Our machine is generally used for development work during the day and so does not have much free time for running other users' production jobs. With more transputers it would be possible to partition the machine and have a number of transputers for running sequential tasks, and a number for development work.

Overnight and at weekends, the machine is generally used for running large production jobs in parallel (eg LOCROT and SPRED - see below). Again, with more transputers, sequential tasks could continue to be processed throughout the night.

Parallel applications

We have adapted a number of applications to run in parallel on the machine, making the fullest use of its parallel architecture and gaining significant speedup over the microVAX II. These are described briefly below.

CARTE. This is an energy minimisation program used for producing energy contour maps of the interaction between solvent molecules and nucleotides or protein side-chains. It involves a minimisation calculation at a large number of planar grid points around the molecule of interest. The calculations at each point are independent of every other point. Thus the program can apportion the grid points across all the transputers and have the calculations performed concurrently. With four transputers, the speedup over the microVAX II is about 8 times, and that over the VAX 11/750 is about 25 times.

RESTRAIN. We have split up RESTRAIN, the least-squares restrained refinement program, to run as a parallel program on the Meiko. Its most CPU-intensive

routines involve the calculation of the structure factors and derivatives for each reflection read in. As in CARTE, because each of the calculations is independent, they can be performed in parallel on separate transputers. Using 3 slave transputers gives a speedup of about 6.5 over the microVAX II. However, because we have so few transputers in our machine, its performance is not currently comparable with the supercomputers on which RESTRAIN is generally run. To match the Amdahl 5890/300 at ULCC, for example, would require a total of 24 transputers, while to match the Convex C1 would require about 37.

LOCROT. This is a program that performs a search in 5 dimensions (3 translational, 2 rotational) for local rotation (ie non-crystallographic) axes of a given symmetry in a low resolution and/or poorly phased SIR or MIR electron density map, when the heavy atom constellation does not conform to the non-crystallographic symmetry. Its purpose is to provide an accurate fix on such axes for use when improving the map by the technique of molecular averaging. The program runs about 9 times faster in parallel on 4 transputers than on the microVAX II. It has been used to help locate the 5-fold axis of SAP (serum amyloid P-component) more accurately, and is currently being used to locate the three 2-fold axes in the bPP (bovine pancreatic polypeptide) structure. Typical run times for SAP were 13 hours for a coarse search and weekend runs of 40 hours for a fine search (corresponding to run times of about 117 and 360 hours, respectively, on the microVAX II).

BERNAL. This is a Monte Carlo simulation program for solute/solvent systems. It has been parallelised by having each transputer maintain a map of all the molecule positions and perform the calculations for a fraction of these. Each new random move can be generated simultaneously on each of the slaves by taking advantage of the behaviour of pseudo-random number generators. Results for each new configuration are collected by a master transputer which decides whether the configuration is to be accepted. Like RESTRAIN, the program is generally run on the Cray X-MP/28 at ULCC; there it achieves 216,000 configuration per hour. On our 4-transputer Meiko it currently manages 30,000, though the relative performance of the Meiko improves as the number of molecules in the system is increased.

SPRED. This is a program for predicting protein secondary structure. It uses a pattern matching algorithm to compile a database of amino acid patterns that are consistently associated with a particular secondary structure. The compilation uses 76 protein sequences of known structure. At the moment SPRED runs only as a sequential program on a single transputer, but will shortly be modified to have the pattern-generation and searching routines performed in parallel.

European Workshop on Laue Crystallography

Trevor Greenhough

University of Keele and Daresbury Laboratory.

A European Workshop on Laue Crystallography, sponsored jointly by the S.E.R.C. and the European Science Foundation Network on Crystallography of Biological Macromolecules, was held at Daresbury Laboratory between Friday October 27th and Monday October 30th. The meeting was intended not only to provide practical experience for the uninitiated, but also to provide a forum for recent results and a flavour of what might be achieved in the near future. The meeting was organised jointly by Trevor Greenhough (Keele/DL) and Keith Wilson (EMBL, Hamburg), while Marjorie Harding (Liverpool) and John Campbell (DL) organised the data analysis sessions and Janos Hajdu (Oxford) managed to keep control of a constant stream of samples on station 9.7.

By bringing together the majority of the available expertise in Laue Crystallography it was hoped that a more widespread involvement might be encouraged and that the ideas and contacts required for future developments would be stimulated. With many research groups now beginning to capitalise on the experimental and data analysis successes at Daresbury, this workshop fulfilled a growing need to disseminate both results and expertise and to promote Laue Crystallography as a technique worthy of serious consideration in the investigation of molecular structure.

The Workshop brought together students and staff from many of the European Laboratories where protein crystallography is a major research area. Countries represented included the U.K., Germany, France, Sweden, Greece, Portugal, Holland, Switzerland, Italy and the U.S.S.R. A full program included lectures on all aspects of Laue crystallography, research talks on the latest results from a variety of Laboratories, and extensive and exhausting sessions of data collection and analysis. The 44 participants divided evenly between "students" and instructors/lecturers, the vast majority of the latter being from the U.K. groups who have pioneered the technique.

The workshop began on Friday with a series of Lectures concerning all practical, theoretical and experimental techniques, given to an audience swelled to twice the expected number by Daresbury staff and visitors from several U.K. Laboratories. Janos Hajdu began with an overview of Laue Crystallography, addressing not only the ideas which first brought it to our attention as a potentially powerful tool for the time resolved study of macromolecular processes, but also the successes, failures and possibilities of the method. This was followed by a lecture from John Helliwell who gave a presentation of his work to establish the credentials of the Laue method through detailed studies of both proteins and small molecules using both monochromatic and Laue X-ray analyses. Durward Cruickshank then provided the definitive lecture on the theoretical aspects of the technique. His treatment of Gnomonic Projections and their application turned out to be something of a revelation in the hands of Paul Carr during the data analysis sessions. An overview of the Laue data analysis suite at Daresbury by John Campbell was followed by Marjorie Harding's presentation of small molecule applications of the Laue method. This included details of experimental conditions and data collection and processing strategies, examples of studies demonstrateing the validity of the method, and potential future applications. Trevor Greenhough described the Laue Crystallography of large unit cell problems, with particular reference to virus crystals and the deconvolution of spatial overlaps. The final lecture of the day was given by Keith Moffat of Cornell, detailing the principles and problems of time resolved crystallography. Particular interest was caused by the Cornell approach to wavelength normalisation. The day was rounded off by a somewhat boisterous workshop dinner at the Lymm Hotel.

Both Saturday and Sunday were taken up with 12-hour practical sessions, with the occasional break for meals and a session of research talks on Sunday afternoon. All participants were able to expose their own samples on 9.7 while data analysis, first of standard examples then of their own exposures, proceeded on both the Convex and DLVD. A total of around 50 samples were exposed over the weekend; such a diverse collection of crystals of biologically important macromolecules will probably never be assembled and destroyed in such a short space of time again. The majority of the samples brought by participants were unknown quantities as far as white-beam behaviour was

concerned. Nevertheless well over half gave good Laue diffraction patterns and several participants were able to make good progress on the analysis of their own exposures. It is a tribute to the Daresbury staff and the remarkably knowledgable and resilient instructors that all 22 "students" were able to collect and process data in the space of two 12-hour sessions. A highlight of the data analysis sessions was the realisation that Paul Carr had achieved a mastery of gnomonic projections sufficient to allow him to find crystal orientations in seemingly hopeless cases. He was persuaded by popular demand into giving a talk on the subject. Despite the length and intense nature of the practical sessions, the atmosphere in the Biology Support Lab seminar room, where all terminals, participants and instructors were housed, remained good-natured, humorous and extremely productive for all involved. I am sure the instructors also learned a great deal from the vast experience of the "students".

The research talks on Sunday afternoon began with a presentation by Fred Vellieux (Groningen) on his work with form I glycosomal GAPDH crystals usinf the Laue method. This structure has been solved from Laue data using molecular replacement method and the refined model of GAPDH from B. Stearothermophilus. The structure is now being refined and currently has an R-factor of 31.1% to 3.2 Åfollowing 250 steps of energy minimization with X-ray constraints. Emil Pai and Ilme Schlichting ((Heidelberg) then described their monochromatic and Laue diffraction studies of the nucleotide binding domain of the ras oncogene product p21. The protein has been crystallised complexed to a photolabile GTP-analogue which is stable against hydrolysis by the protein. Results of flash photolysis experiments at HASYLAB (X31) were presented. The structure determination of TEWL using Laue diffraction data was described by Lynne Howell (CEA, France). This first Laue data protein structure was solved by molecular replacement using the refined model of hen egg white lysozyme. Refinement using the simulated annealing technique has yielded a final R-factor of 19%. The final talk was given by Keith Moffat who described the Laue data analysis suite under development at Cornell.

The success of the workshop owes a great deal to those U.K. researchers and Daresbury staff who have established Daresbury as the leading centre for Laue Crystallog-

raphy. There is every indication that a great deal of interest and enthusiasm for Laue crystallography was generated by the meeting, and we can look forward to an increased flow of new ideas and results from the many groups who were represented. I could not end this report without naming those people without whose commitment and enthusiasm we would not have been able to proceed. The expert instuctors were: Inger Andersson (Uppsala), Sue Bailey (Keele), Paul Carr (Liverpool), Ian Clifton (Oxford), George Habash (Manchester), Steve Harrop (Manchester), Adrian Lapthorn (Liverpool Poly.), Steve Maginn (Liverpool) and Annette Shrive (Keele), with a little help from myself, John Campbell, Janos Hajdu and Marjorie Harding. The organisers wish to thank Shirley Lowndes, David Brown and Geoff Berry of Daresbury for the organisation of the programme and documents, and Kim Henrick and Ian Glover for preparation and support on DLVD/CONVEX and station 9.7 respectively. The ESF and the SERC are thanked for support.

Copies of the Workshop Programme, including abstracts of lectures and talks, are available on request.

Progress report on the structure determination of Deoxyuridine Triphosphate Nucleotidohydrolase (d-UTPase) from *E. coli.*

Gunilla Larsson, Per-Olof Nyman,	University of Lund
Eila Cedergren-Zeppezauer, Salam Al-Karadaghi,	University of Stockholm
Janos Hajdu	University of Oxford
Zbigniew Dauter, Keith S. Wilson,	EMBL Hamburg

The enzyme d-UTPase catalyses the hydrolysis of deoxyuridine-5'-triphosphate (d-UTP) to d-UMP and pyrophosphate. This reaction is of greatest importance to the cell. Decreased d-UTPase activity (measured on *E. coli* mutants having only a few percent of the activity of a wild type strain) causes severe damage, like elevated recombination frequency (hyper-rec), abnormal mutation rate and accumulation of high concentration of short DNA fragments as intermediates in the DNA replication. These phenomena all derive from the fact that uracil was incorporated in DNA as a result of too high concentration of d-UTP in the cell.

Very little biochemical data is available for this enzyme, but it was discovered early on that the substrate specificity for d-UTP is absolute. Gel-filtration experiments indicate that it is a tetrameric protein. Recent investigations show that it requires Mg^{2+} ion for activity. d-Utpase has a molecular weight of 14 kDa per monomer.

The structural gene for d-UTPase from *E. coli* (dut) has been isolated and the nucleotide sequence of the dut operon has been determined. Using an expression vector we have obtained an overproducing system that can provide about 600 times more d-UTPase than wild type bacterial strains.

Crystals of d-UTPase of various morphology and size grow very easily but only one form is suitable for high resolution X-ray analysis. These crystals belong to the space group R3, can be 1 mm large and diffract to about 1.6 Å resolution. Initially we could grow these large, well diffracting crystals without difficulty and started the data collection and search for heavy atom derivatives. The more we refined the preparation procedure for d-UTPase the poorer was the quality of the crystals and we noticed that we needed to store the enzyme for "maturation" for up to a year before nice large crystals appeared.

We have now solved the problem of the nonreproducible crystallisation of d-UTPase and we think it may be of interest for others. The "maturation" process which leads to the R3 form of crystals was found to be dependent on the buffer system, concentration of succinate ion, temperature and pH.

Electrophoretic and chromatographic analyses have shown that a homogenous solution of enzyme gradually develops into different populations of molecules. On an anion exchange column a freshly prepared enzyme solution shows one homogenous peak, whereas an aged preparation of enzyme gives rise to several peaks with longer retention times.

Likewise, we observe that on native gels and isoelectric focusing gels, in the presence of 8 M urea, an aged d-UTPase solution shows many bands in contrast to the freshly prepared sample. However, on a gel-filtration column and SDS-PAGE a single peak or band is present, regardless of the storage time.

Starting from a single population of molecules the enzyme in solution changes into more negatively charged forms of the protein. Dissolved R3 crystals have been shown to contain the most negatively charged components. We believe that the reason for this storage requirement is a deamidation reaction of asparagine and glutamine residues. In the deamidation reaction the side chain amide group is hydrolysed to form a carboxylic acid. This process is a commonly observed form of post-translational protein modification. The deamidation process can be accelerated by increasing the temperature and pH. By treating a freshly prepared d-UTPase sample for a few days at temperatures around 37 - 42° C in phosphate buffer at high pH and fine-tuning the succinate concentration, we can now control

the crystallisability of d-UTPase and reproducibly grow the desired crystal form. This treatment does not affect the enzymatic activity.

The crystals used for data collection were precipitated with PEG 8000 at pH 4 in the presence of 50mM of succinate, pyrophosphate and $MgCl_2$. A complete 1.9 Å data set was collected on X31 beamline at EMBL Hamburg, using the Image Plate scanner constructed in-house by J. Hendrix.

Two crystals were used, one to collect high resolution data set to 1.9 Å resolution, the second to cover the low resolution (2.6 Å) part which was to intense in the other set. For the first data set 51 images were collected with 2° oscillations and about 2 min exposure time per image.The wavelength used was 1.009 Å. For the low resolution data set 24 images were collected with 4° oscillations and about 8 times less exposure. The data were processed with the modified MOSCO package to produce 64120 intensities. A final R-merge of 6.9 % was obtained for all 13640 unique reflections (95 % complete). The space group was confirmed as R3 with cell dimensions a=b= 86.6 Å, c= 62.2 Å in hexagonal setting.

A single crystal soaked in ethyl mercury thiosalicylate (EMTS) was used to collect a complete 2.0 Å data set using the same conditions as for the native crystal, except that the wavelength was changed to 0.995 Å to maximise the anomalous diffraction effect of Hg. 42611 intensities were merged to 11730 unique reflections with R-merge of 5.9 %. A similar data set was also collected on a sample co-crystallised with Hg^{2+} ions.

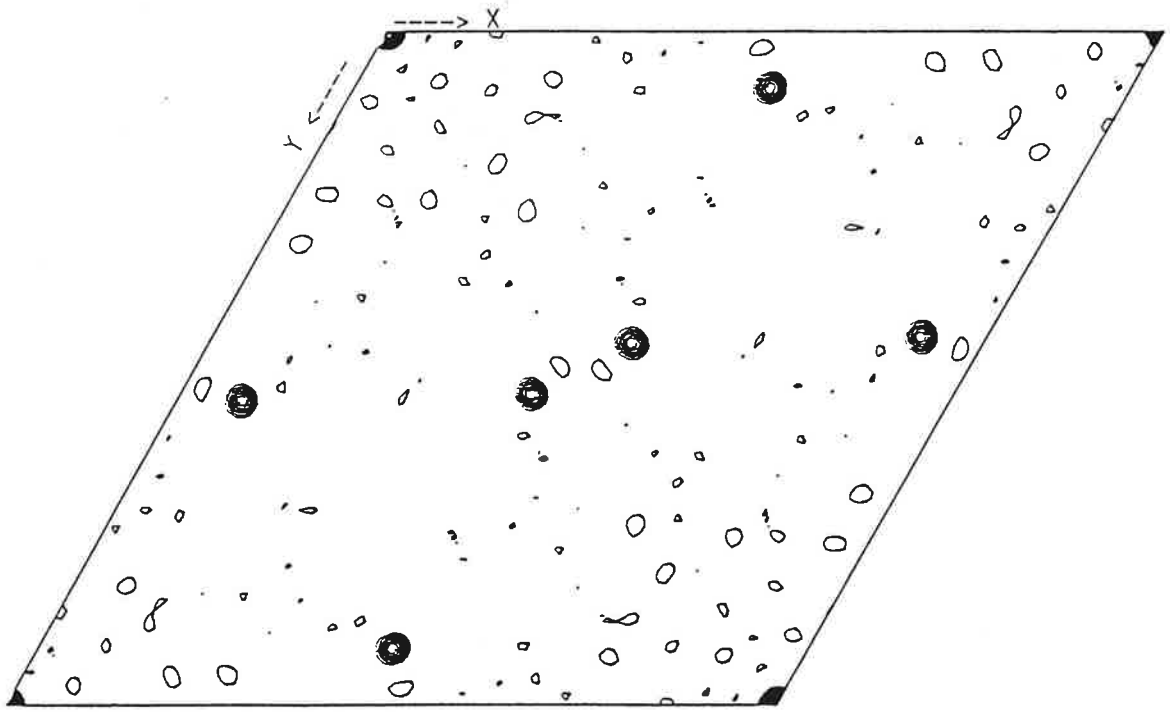
A second heavy atom derivative was prepared by soaking the crystals in 1 mM solution of K_2PtCl_4 for 7 h. The data to 2.1 Å resolution were collected using the same strategy as for the other sets with the wavelength of 1.05 Å. The R-merge value of 5.7 % was obtained by merging 33580 intensities into a set of 9910 unique reflections.

Both the Hg and Pt derivatives proved to be useful and produced very clear anomalous and difference Patterson maps with single sites of the heavy atoms. However, the data of the Hg-derivative obtained by co-crystallisation appeared to be non-isomorphous with the native, and produced very clear anomalous Patterson map but no peaks on the difference Patterson. The relative position of platinum with respect to the mercury atom was found on a difference map calculated on Pt and native amplitudes and phases based upon the Hg atom.

The phasing procedure based on the two derivatives resulted in the overall figure of merit 51 % at 2.1 Å and produced a map with clear continuity in the electron density and good contrast between protein and water regions. A map calculated for the opposite handedness was not interpretable.

At present we are building the protein chain into the map. It is clearl that the molecules are arranged in trimers around the three-fold axis.

The hydrolysis of d-UTP catalysed by d-UTPase might be a suitable reaction to follow by Laue diffraction methods. We have tested the R3 type crystals at the Daresbury synchrotron and were able to obtain high quality pictures on large crystals with an exposure time of 600 milliseconds. These data are being evaluated at present.



Z = 0 section of the anomalous Patterson map for Hg-derivative of d-UTPase ($\lambda = 0.995$ Å).

The high resolution structure of Xylose Isomerase from *Streptomyces albus*

Keith S. Wilson, Howard Terry, Zbigniew Dauter, EMBL Hamburg
Charles Collyer, Imperial College, London
Herbert Witzel University of Münster

The enzyme xylose isomerase from *S. albus* forms large (up to 3 mm) rhombo-dodecahedral crystals from a 100 mg/ml water solution in the presence of either 0.1 M $MgCl_2$ or other di-valent metal ions. The crystals are orthorhombic with space group I222 and cell dimensions **a**= 93.9 Å, **b**= 99.7 Å, **c**= 102.9 Å. The assymmetric unit contains one monomer of the protein of 40 kDa. Four monomers aggregate to form a tetramer located at the 222 symmetry site. The crystals diffract to about 1.6 Å resolution.

We have collected the complete data set on both the metal-free and Mg^{2+} crystals on the Hendrix/Lentfer image plate scanner on the X11 beamline at EMBL Hamburg. In both cases three data sets were collected. A 'high' resolution set (data from 1.65 Å for the metal-free crystals and data from 1.55 Å from Mg^{2+} containing crystals) was collected in ninety 1° oscillations.A 'medium' resolution data set (data from 2.0 Å) collected with 1.5° oscillations and a 'low' resolution set (data from 2.7 Å) collected with 3° oscillations. For the high resolution data the exposure time was between 1 and 2 minutes per image and the ratio of exposure time going from 'high' to 'medium' to 'low' was 40 : 6 : 1. The data were processed using the modified MOSCO package for image plates and merged with the CCP4 programs ROTAVATA /AGROVATA. The Rmerge for the metal free data was 5.8% for 54575 unique reflections and for the Mg^{2+} containing data 5.5 % for 72687 reflections.

As a starting model we have used the structure of xylose isomerase from *Arthrobacter* (Collyer, Henrick and Blow) after the appropriate rotation and translation. Refinement of the metal-free structure was performed using the Konnert-Hendrickson programs from the

CCP4 package. The R-factor dropped from an initial value of 55% to 34% after 10 cycles of refinement at 2.1Å resolution. At this stage a manual rebuild of regions showing well defined secondary structure and the insertion of some obvious sequence corrections was done. A further 20 cycles of refinement during which the resolution range was gradually expanded to 1.7Å resulted in the R-factor dropping to 28%. The difference Fourier map allowed us to insert about 100 water molecules and another 8 cycles of refinement lowered the R-factor to 25% at 1.65Å resolution. The model was subjected to a thorough rebuild of the whole chain and insertion of the correct amino acid sequence for the first 377 residues. 5 more cycles of refinement resulted in an R-factor of 22.5%. The Fourier map now showed a clear extension of the density at the C-terminus as expected from the sequence and 6 residues were built in to the model together with an additional 125 water molecules. The next 5 cycles of refinement decreased the R-factor to 18.7%. We are currently inserting the final 4 C-terminal residues, searching for more water molecules and correcting some minor errors in the flexible loop regions.

We intend to extend the resolution by using the Mg data. We have also collected and processed diffration data from crystals containing metal ions (Co^{2+}, Cd^{2+},VO^{2+}) and substrate/inhibitor complexes. Preliminary studies of the difference Fourier maps against the metal free data show clearly the metal binding sites.

X-RAY CRYSTALLOGRAPHIC STUDIES OF THE HIV-1 PROTEINASE

The Retroviral Proteinase Group
Department of Crystallography
Birkbeck College
Malet Street
LONDON, WC1E 7HX

Introduction

The HIV-1 retrovirus is one of a large family of enveloped viruses containing a single-stranded RNA genome which replicate via an obligatory DNA intermediate. Its notoriety stems from its role as the etiological agent of AIDS.

The retroviral genome consists of three major genetic elements that are arranged in the order 5'-gag-pol-env-3' (Krausslich, 1988). The gag gene codes for the structural proteins of the nucleocapsid and the env gene for the glyoproteins of the viral envelope. The pol gene codes for the enzymes involved in viral replication including a proteinase, a reverse transcriptase and an integrase. These proteins are expresses as part of large fusion polyproteins (gag, env and gag-pol, respectively) which, with the exception of env, are subject to processing by the virus-encoded proteinase. Thus, the HIV-1 proteinase is responsible for the cleavage of its own precursor, the gag-pol fusion polyprotein.

The HIV-1 proteinase in its mature form is a 99 amino acid subunit catalytically active as a dimer. The enzyme has been identified as an aspartic proteinase on the basis of several converging strands of information. Firstly, a conserved sequence Asp-Thr/Ser-Gly, conserved in all pepsin-like aspartic proteinases, is also found in the retroviral proteinases, (Toh, 1985). Secondly, weak competitive inhibition is afforded by the characteristic inhibitor pepstatin (Darke, 1989) and lastly, mutation of the essential aspartic acid to an asparagine leads to catalytically inactive protein (Kohl, 1988).

Attempts to develop antiviral drugs have so far centred on the viral reverse transcriptase but the proteinase may also prove to be a suitable target. This is so because the HIV-1 proteinase has been found to be essential for viral replication (Kohl, 1988). Thus, inhibitors of viral proteolytic processing which do not effect the metabolism of the host cell may prove to be of use as antivirals in the treatment of AIDS. In pursuit of this goal we have undertaken an X-ray crystallographic analysis of the native HIV-1 proteinase and initiated complementary studies of the enzyme complexed with high affinity inhibitors.

Methods

An plasmid containing a gene encoding a fusion polypeptide comprising an inhibitor methionine, a 7 residue junction sequence and a section of the pol gene product with a molecular weight of 115 kDaltons was expressed in *E. Coli*. This gene encompasses the proteinase, reverse transcriptase and integrase enzymes. The N-terminal sequence of the proteinase revealed the correct self-processing of the pol gene product. This recombinant material was provided by collaborators at Pfizer Central Research, UK and USA.

Tetragonal bipyramidal crystals of space group $P4_12_12$ were produced in 5 days using the hanging drop vapour diffusion method. Cell parameters

a=b=50.24Å, c=107.12Å, $\alpha=\beta=\gamma=90°$ were isomorphous with those produced elsewhere. (Navia, 1989) Diffraction data were collected from a single crystal over 2 days using a FAST area detector giving a total of 10,000 reflections. The merging R-value for this data was 0.11 giving 2370 independent reflections with $I>2\sigma(I)$ to a nominal maximum resolution of 2.7Å. Film data has since also been collected with the rotation method suing the synchrotron source at the Daresbury Laboratory. Comparative statistics for the two modes of collection are presented as Table 1. The crystals are characterized by their high mosaicity and short beam lifetime.

Solution of the Structure at 2.7Å Resolution

Using the program COMPOSER (Blundell, 1988) a model for molecular replacement was constructed from the known structures of pepsin-like enzymes and later from the homologous Rous sarcoma virus (RSV) proteinase. Several regions were specifically excluded from this model including the N- and C-terminal strands, the flap and the helix as these were in contention with the results of a previous analysis (Navia, 1989). Molecular replacement, performed with the CCP4 programs ALMN and TFSGEN, was reasonably straightforward given that only approximately two-thirds of the enzyme was present in the search model. The analysis provided the expected solution, namely a single protein subunit per asymmetric unit with the dimer generated by operation with the diagonal 2-fold of the cell.

A series of difference Fourier analyses, remodelling and refinement steps using the program RESTRAIN (Haneef, 1985) revealed those portions of the structure omitted from the initial phasing model. In each case the polypeptide could be followed in the difference electron density (calculated on the basis of a model from which they were excluded) without a break and with clear positions for the side-chains. In the later stages the molecular dynamics-based refinement program XPLOR was also used (Brünger, 1987). The final model includes individual atomic isotropic temperature factors and tight geometric restraints with 0.019Å rms deviation from ideal geometry. The final crystallographic agreement factor is 18.9% for a model containing protein atoms only against data collected to 2.7Å resolution.

Description of the Structure

The fold of the HIV-1 proteinase is basically very similar to that of a single domain of the pepsin-like aspartic proteinases, correcting an interpretation by another group based on 2.7-3Å resolution data which showed no helix and a rearranged basal β-sheet strand arrangement (see Figure 1) (Navia, 1989). In general the subunit consists of two motifs related by an approximate two-fold axis. Each motif is constructed from four anti-parallel β-strands (labelled a, b, c and d for the first motif and a', b', c' and d' for the second). These are organised into a distorted β-sheet with strands b, c and b', c' forming β-hairpins that are folded back over a basal sheet formed by the antiparallel arrangement of the two sets of N- and C- terminal strands. The major central sheet is formed by strands c, d', d and c'. This sheet forms the so-called Ψ-structures with the outermost pair of

strands on each side running parallel and the innermost pair running antiparallel.

The catalytic aspartic acid residue (Asp 25) is located at the end of the c strand preceding a loop whose local structure is stablised by a symmetrical and tightly hydrogen-bonded arrangement dubbed the 'fireman's grip', a feature absolutely conserved in all known aspartic proteinase X-ray structures (James, 1983; Pearl, 1984). Asp 29 plays an important role in hydrogen bonding interactions with peptidic ligands at the P3 positions (Schechter, 1967), replacing Thr 219 of the pepsin-like enzymes. This residue has another role in the retroviral proteinase structure in forming ion pairs with the almost invariant Arg 87 and Arg 8' (from the 2-fold symmetry-related subunit), this interactions forms part of the subunit interface. The conserved and buried Thr 31 interacts with Asn 88, a residue shown to be sensitive to mutation to amino acids of hydrophobic character.

In many respects the hydrogen bonding capacity of the enzyme in the active site is identical to that afforded by the pepsins as witnesses by the structures of many enzyme-inhibitor complexes solved in this laboratory (Blundell, 1987). An exception to this is in the residues of the β-hairpin flap formed by strands a' and b'. This is considerably shortened in the retroviral proteinase structure relative to the pepsin-like cases and, furthermore, there are, of course, two symmetry-related flaps extending over the retroviral proteinase active site rather than a single one in the pepsin-like enzymes.

However, the major difference between the two classes of proteinase arises in the symmetry of the retroviral enzyme. Whilst the crystal structure of the native enzyme is perfectly symmetric, binding of an asymmetric inhibitor will lead to a loss of symmetry. In this respect 2-fold symmetric inhibitors may afford an advantage over their non-symmetric counterparts in terms of gains in affinity.

Work in Progress

Refinement of the structure of the HIV-1 proteinase using synchrotron data of 2.2Å resolution is currently underway. We have obtained small crystals of an enzyme-inhibitor complex which may be suitable for data collection using synchrotron radiation. Several other co-crystallisation experiments are also in-progress. In addition, we have an active program for the in-house development of novel expression systems for the HIV-1 proteinase and its site-directed mutants for future structural investigation.

References

Blundell, T L et al (1988)Eur. J Biochem, 172, 513

Blundell, T L et al (1987) Biochemistry, 26, 5585

Brunger, A T et al (1987) Science, 235, 458

Darke, P L er al (1989) J Biol. Chem., 264, 2307

Haneef, I et al (1985) Acta Crystallog., 426

James, M N G and Sielecki, A R (1983) J Mol Biol, 16, 299

Kohl, N E et al (1988) Proc. Natl. Acad. Sci USA, 85, 4686

Krausslich, H G and Wimmer E (1988) Ann. Rev. Biochem 57, 701

Navia, M A et al (1989) Nature, 337, 615

Pearl, L H and Blundell, T L (1984) FEBS Lett, 174, 96

Schechter, I and Berger A (1967) Biochem. Biophys. Res. Commun., 27, 157

Toh H et al (1985) Nature, 315, 691

Figure Legend

1. Schematic diagram of the super-secondary structure of the HIV-1 proteinase. Well defined hydrogen bonds are shown by an arrow from the donor to the acceptor.

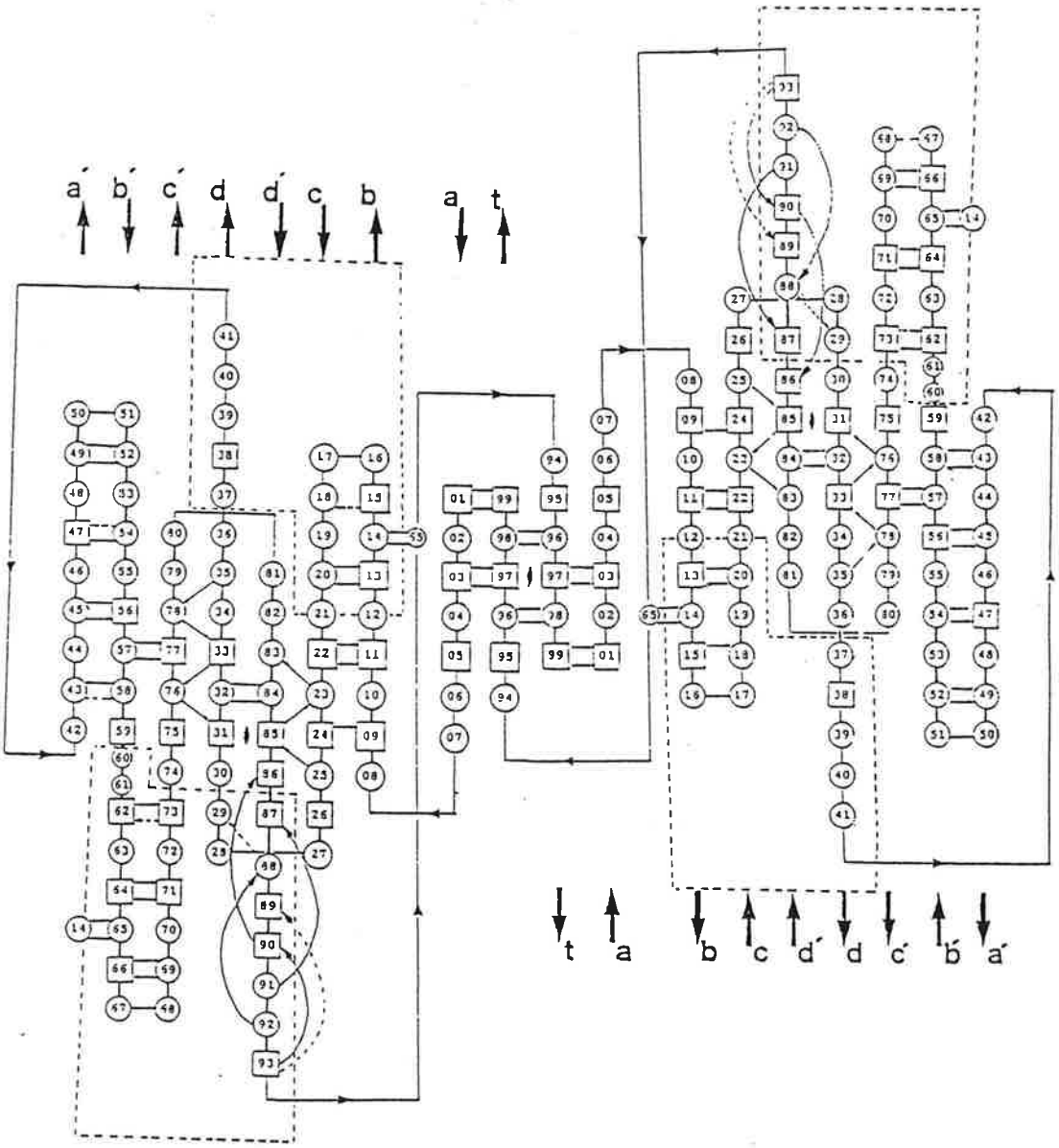


Figure 1

Table 1

Source	Resolution	HKL_{Total}	HKL_{unique}	$\% - complete$	R_{Merge}
FAST	2.7 Å	5918	2370	56.4	11.3
SRS PX7.2	2.2 Å	9759	3763	50.8	16.4

Native HIV-1 Proteinase X-ray data collection statistics

DETECTOR DIATRIBE

Paul Tucker
(European Molecular Biology Laboratory,
Meyerhofstrasse,1 D6900 Heidelberg)
EMail Tucker@DHDEMBL5

Preamble:

In the last months of 1987 both Nicolet (now Siemens) X100A (**X**) and Enraf-Nonius FAST (**F**) area detector systems were installed at EMBL in Heidelberg. In December 1988 we had installed a MAC Science DIP100 (**D**) image plate device. It may therefore be of some interest to the European crystallographic community to comment on our experiences with these three instruments.

For reference our configuration of the three systems is summarized in **Table 1.**

Table 1. Installation Summary

	F	**X**	**D**
Beam:			
Generator	GX21 (Elliot/Enraf)	GX18 (Elliot/Enraf)	MX18 (Siemens/MAC)
Focus	300µ	100µ	300µ
Normal Power	45kV 90mA	35kV 50mA	50kV 90mA
Monochromation	Graphite	Graphite Mirrors possible	Graphite
Collimation	Pinhole/selection	Pinhole	Pinhole/selection
Goniostat:			
Crystal	Kappa-Axis	Eulerian cradle with fixed chi.	Single axis
Detector	Swing Distance	Swing Manual distance	No swing Manual distance
Processors:			
Collection	FALCON	PCS(68010)	68000(VME crate)
Processing	µVAX II	µVAX II	NEC 9801
Disk Space	0.6GByte	0.6GByte	20MByte
Data Frame Size	0.27MByte	0.5MByte	5MByte

Fundamental Parameters:

A comparison of some fundamental characteristics of the (**X**) ,(**F**) and (**D**) systems is made in **Table 2.** On the assumption that we wish to achieve rapid, accurate data collection there are two parameters of prime importance. They are the number of resolution elements (which in turn is determined approximately by the point spread function (PSF) and the active area,**Table 2A**) and the efficiency of detection (**Table 2B**). For (**X**) and (**F**) the PSF was measured by illumination with CuKα radiation of a 25µ pinhole on the detector surface. For (**D**) a mask with 100µ holes 45mm from the image plate was used and the PSF is therefore an overestimate. Plots of the same source (collimated ^{55}Fe, calibrated with a scintillation counter and found to yield 4.2c/s) measured for the same time at the same distance from near the centre of each instrument are shown in **Figure 1.** The results of repeated mesurements are given in **Table 2B.** Overall (**D**) appears superior at present by virtue of the large active area. This is rather respectable for a device of about one half to one third the cost of the other

two systems. However the readout overhead (**Table 2C**) means it is impractical to collect small (0.1°) rotation frames and we might expect poorer signal to noise ratios in crystallographic data collections (see below). For our new (**F**) we expected a slightly better gain and a significantly lower noise than we have actually measured (but see the crystallographic results below). The global count rate limitation for (X) in **Table 2C** does not correspond to the manufacturers claim nor to what is assumed by the software. It is however rather unimportant if count rates are kept below 30KHz because the global count rate tends to be a smoothly varying function with rotation angle and its effect is taken out in the scaling of the data.

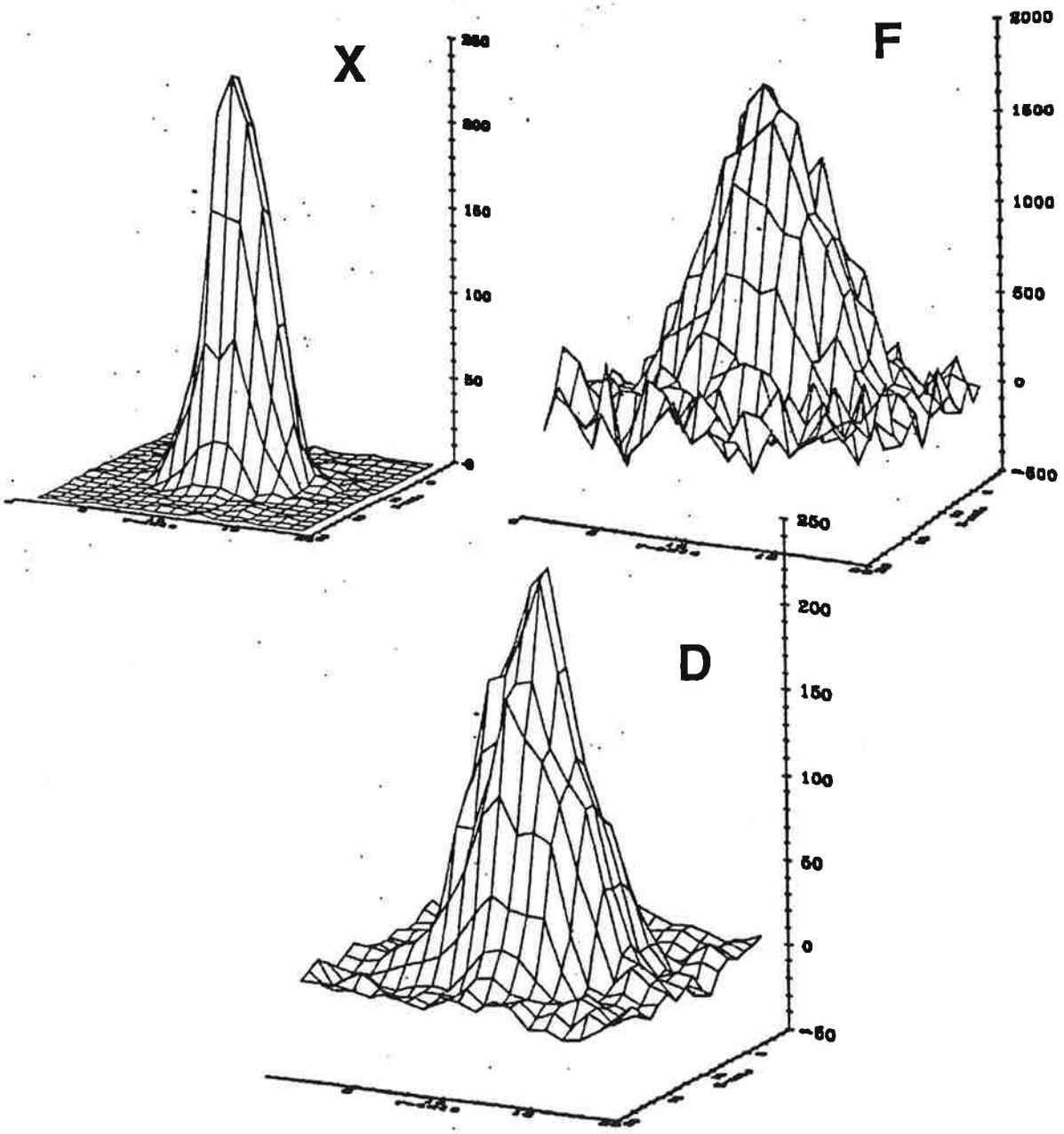


Figure 1. Measured profiles of an ^{55}Fe source for the same time and at the same distance on (X), (F) and (D). In the latter two cases a local background has been subtracted.

Table 2A.Point Spread Function and Active Area

	X	F	D
Binned:	512x512	512x512	1600x1600
Size:	100mm diam.	46.8x62.4mm	200mm diam.
Active Pixels:	236K	240K	2000K
FW1%M:	5x6 pixels[1]	6x7 pixels	8x8pixels(est)
Resoln.Elements:	8K	6K	31K
Variability:			
with position;	20% in detector X[2] Parallax:FW1%M doubles at angle of incidence of 15°	Broadening towards edge of active area	None found
with intensity;	None	Increased by 2 pixels at saturation	Unknown

[1] Approximately 4x5 pixels on first chamber. [2] This is on the original chamber.This had a large residual modulation in places with a peak to valley ratio approaching 1:4. The new chamber does not show this feature and the variation will therefore be less but has not yet been measured.

Table 2B.Detection Efficiency

	X	D	F
Incident photons:	10080	10080	10080
Transmission[4]	52%(Be window)	83%(paper)	83%(paper)
Expected photons:	5242	8366	8366
MS Units Observed	4803(88)	9592(484)[1]	97159(4076)
I/sigma(I)	54	20	24
Utility Factor [2]	432	610	140
Wavelength factor[3]		61(est)	48
Total	**432**	**670**	**188**

[1]The gain is misleading because the MS (mass store) units read out in cylidrical coordinates are distributed to an orthogonal grid.The result is that the 'true' number of MS units is about 1/4 of the number given and the photon equivalent is therefore about 4 [2]Defined as = (No. of Resolution Elements) x I/sigma(I) .[3]Factor to be added because measurment not made at wavelength used in crystallographic work .[4]Transmission calculated for ^{55}Fe source NOT CuKα

Table 2C.Count Rate and Dynamic Range Limitations

	X	F	D
Local Count Rate Limitations[1]:	225 c/p/s or 1.5KHz per resolution element	300 c/p/s gives ADC overflow at gain 20	Unknown
16bit overflow in:	290s	41s	Unknown
Global Count Rate Limitations:	1)13%deadtime at 30KHz 2)16%deadtime at 30KHz	None	None
Overheads:			
Erase			59s
Readout			178s
Frame Transfer	6s	2s	44s to DRAM
Display	5s	None	11s
Typical frame time	120s/0.2°	60s/0.1°	1800s/1.0°
Dynamic Range:			
Estimate	5×10^4	1×10^3 [2]	unknown
Observed F[2]	3×10^3	6×10^2	2×10^3

[1]When checked the response has been approximately linear up to this count rate,[2]Extendable by changing gain but at a cost in data collection speed.

Calibration:

The features of, and results from, the calibration of the instruments (**Table 3.**) deserve a few additional comments. The non-uniformity of response of (**F**) is large (apparently differing at the edges of the active area by some 30% of the value at the centre), but is calibrated ,taking into account the effects of spatial distortion. It is, at least crudely, verifiable and is corrected for by the processing software. One error is expected to be at edges of the active area were the PSF leads to an overestimation of the correction factor[1]. I would observe that the non-uniformity calibration is critical and is still being checked and improved. The (**X**) nonuniformity is small on our current chamber ,as indeed it should be , since no non-uniformity of response correction is measured by the calibration software and no correction for non-uniformity is made in the processing software. Some crystallographic checks of the response on (**X**) and (**F**) are shown in **Table 3C.** For (**F**) the discrepancies probably result from the PSF effect mentioned above, whilst for (**X**) they may reflect a less than satisfactory chamber in one smallish area.

Table 3A.Variations of Efficiency over detector surface/Non-Uniformity calibration.[1]

X Chamber 1)	Residual Modulation. Peak/Valley Ratio ca 1:4. 85% of maximum in specific regions.
2)	Residual Modulation negligable.Differences less than 2% except possibly in one smallish region.
Method	12bit ADC histogrammed in X and Y. Radial intensity falloff applied. Lookup table constructed such that X and Y dimensions of pixel adjusted to give smooth response over 8bit coordinate value. Therefore simply rebins events and no correction for non-uniformity made. Rebinning takes no account of spatial distortions and treats radial intensity falloff incorrectly.
Stability	Local variations with time (1st chamber)
F	60-70% of maximum towards edges of active area .Verified (except at detector edges where predicted overcorrection is observed).
Method	Spatial distortion taken into account. No account of differences in absorption (Ref. 1) due to different energies used for calibration and measurement.
Stability	Sensitive to temperature and magnetic field *at least* in the sense that the spatial distortion is sensitive to these factors.
D	30% variations at pixel level(fringing pattern due to sampling variations resulting from readout method). Lower than expected efficiency at both high and low radius. Crudely verified to within 5%.
Method	After applying sampling correction the radial dependence of efficiency is determined and smoothed. The correction is applied to the pixel by pixel sampling correction and the resultant file is applied to all measured frames. (Ref.8)
Stability	Unknown

[1](X) and (F) both use ^{55}Fe or fluorescent Fe source to give Flood Field . Verification of the calibration we have done crudely by scanning main beam across detector in one dimension and more thoroughly by repeated crystallographic measurements at different detector swing angles (see **Table 3C**) . For (D) a ^{55}Fe source is mounted on the read head and scanned across the plate .

For (X) and (F) the calibration of spatial distortion (**Table 3B**) seems adequate, for (D) it seems unnecessary. It needs to be done for each crystal-detector distance for (X) but seems rather crystal-detector distance independent for (F). In the past we have done our (F) calibrations every 2-3 months but now we do them at the crystal to detector distance and swing angle to be used in the experiment. The calibrations need take no more than 3hrs. We perform our (X) calibrations every week or when the crystal to detector distance is changed, the calibrations can however be performed rather more quickly than for (F).

Table 3B.Spatial Distortion Corrections

	F	X	D
Type of parameterisation:	Polynomial (order 7 usually)	Spline or Fourier Series	None(spatially linear)
RMS error to mask[2]: in X	0.1(mask	0.2[1]	0.9
in Y	limited)	0.2	0.6
RMS error to reflection coordinates[2]:	0.6	0.3[1]	1.5

Stability of calibrations:

F Sensitive to temperature,so time variations difficult to measure. Crystallographically valid several months. Sensitive to magnetic field.

X Sample-detector distance dependent. Crystallographically valid several weeks.

[1]Chamber 2. First Chamber always gave larger RMS error for X.[2] In pixels

Table 3C.Checking Detector Response

F (Lysozyme Crystal . 25 deg.of data)

Swing(°)	D(mm)	R(F^2)
25	75	5.3
15	75	4.8
5	75	4.7
-5	75	5.2
-15	75	4.9
-15	100	4.6
-15	60	5.1
	Overall	**5.8**

X (Lysozyme Crystal . 15 deg.of data)

Swing(°)	D(cm)	R(F^2)
-5	12.0	4.7
-15	12.0	5.3
-15	15.0	4.8
-15	9.0	6.2
	Overall	**5.6**

Z_{MS}	Z_{MS}.and Y_{MS}	<scale>[1]
1-32	>64 from edge	1.17
481-512	>64 from edge	1.22
33-64	>64 from edge	1.11
449-480	>64 from edge	1.08
128-384	128-384	1.02

X_{MS} or Y_{MS}	X_{MS} and Y_{MS}	<scale>[1]
X_{MS} 1-32	>64 from edge	1.01
Y_{MS} 1-32	>64 from edge	0.91
X_{MS} 481-512	>64 from edge	0.99
Y_{MS} 481-512	>64 from edge	1.00

[1]<scale> is the mean scale factor(over all reflections) between measurements of the same reflection recorded in different data sets at positions indicated in the preceding columns.

Time Variations:

One anticipated problem with (D) is decay of the latent image .We have looked for short time scale variations (to try to extrapolate to losses occuring during the readout period) but cannot find any greater than around 3%, although there is a clear longer time scale decay. Results of two experiments are shown in **Figure 2.** A problem on (F) has been the instability of the non-Xray background. This is shown in **Figure 3,** where the laboratory temperature is plotted as a function of time together with the non-Xray background measured 4 hours later. The correlation is extremely good and most importantly the difference in non-Xray background corresponds roughly to a

significant but weak reflection. The temperature stabilisation and the software checks on its constancy are, we think, the major reasons why our new FAST system gives so much better results than the initial one. For (X) we have noticed variations over a period of days of the residual modulation pattern but this is only likely to affect profile fitting methods.

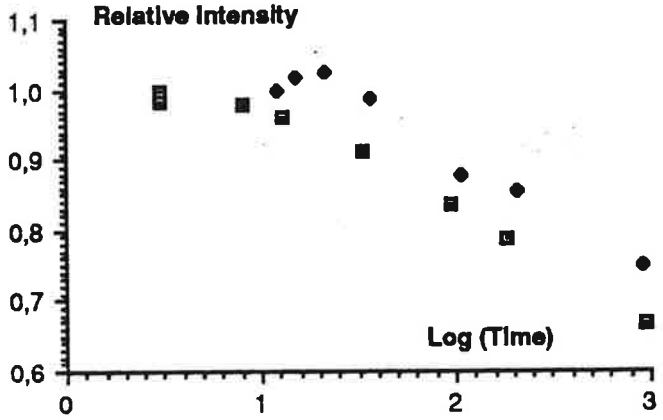


Figure 2. Results of two experiments recording the same oscillation range (lysozyme crystal) and leaving for a time period (in minutes) before readout, correction and integration. Intensities then scaled and the relative scale factors converted to fractional intensity loss plotted against the log of the time delay in minutes.

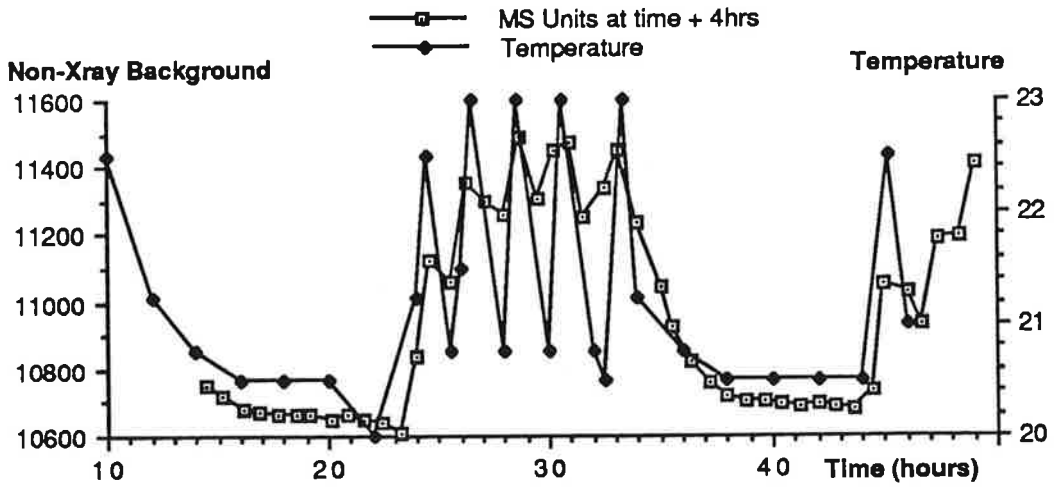


Figure 3. Plot of Non-Xray Background for the (F) system and laboratory temperature as a function of time, the effects of temperature are heavily damped and lag by about four hours.

Reliability and Usability:

Data collected in the first few months after the installation of (X) were unreliable (unless data frames had been archived and were reprocessed) resulting from a serious error in the XENGEN V1.1 software we were using at that time. (X) then performed reliably for 8 months before an I/O board and streamer tape drive on the PCS Computer needed replacing (down 2 time weeks). Some months later a fault began to appear (streaking of diffraction spots in the X direction) which became so bad that the instrument became unusable. A replacement chamber was required (downtime 6 weeks ,admittedly over the Christmas/New Year Period). More recently we had a wire to one of the preamplifiers come adrift . Hopefully the downtime awaiting replacement parts will be reduced when Siemens carries a full stock, including replacement chambers, in Europe.

In the initial months after installation (F) required personality module replacements and a replacement interface. We are on our third monitor. The down time is difficult to estimate because the instrument was not unusable just inconvenient and unreliable to use. Other minor faults have been corrected and upgrades made as part of the manufacturers field service. Eight months later these initial irritations faded to insignificance when, after concern on our part about data quality, the manufacturers found a generator installation error (corroded Ni filter left in the generator shutter housing) as well as a poor quality monochromator crystal. The consequent increase in incident beam intensity (by a factor of about 4.5) meant that most of our original comparative measurements were of dubious value. These measurements are, nevertheless, given below, together with measurements made after improvement of the main beam (**Table 6a**). Even so, concern about instability of the non-Xray background and the poorer quality data resulted in minimal use of the instrument during the first months of 1989. A new , thermally stabilised system (including short arm goniostat and new detector) was delivered in May this year. The new instrument (apart from some initial board faults) has been trouble free. The one problem we have had was primarily our fault in that a rather tight fit of the calibration grid resulted in my damaging the phosphor. Although the detector had to be returned to Delft, we were up and running again within three weeks.

With (D) there have also been problems which have so far been correctable by exchange of FAX messages with the company in Japan. In fairness these have mostly been connected with our desire to improve the efficiency of detection, which,, from time to time, has left us with a system whose response drifts with time. Other problems have been with the laser shutter and the rotation axis zero position. Our recent difficulty with an unreliable DECnet connection seems to have been solved. The company have just installed higher gain photomultiplier tubes so we expect a further improvement shortly. Although fairly stable now, the software has,over the year, been continually improved and this has made the system difficult to compare crystallographically with (X) and (F).

Software:

On (X) we have installed, for frame processing, the following software: XENGEN[2], BUDDHA[3] and XDS[4]. The (common) data collection software is described in Ref. 3. We have used primarily XENGEN because we obtained it first and most of our users are more familiar with it. Our experience with the XDS software software is still limited but in most cases it has appeared superior to the XENGEN software. We have, in one case, processed the same data frames with the three different software packages (90 degrees worth of Phospholypase C data collected for 120secs per 0.25 degree frame with a swing angle of 30.0 degrees). These results (**Table 4**) show a clear advantage for the XDS software. In previous comparisons of the

GENEX and BUDDHA software we have found very little to chose between these two packages. There a few comments of note here: XENGEN produces more reflections than BUDDHA or XDS, primarily because the active area is defined by the gamma background rather than an arbitrary radius and/or estimate of the beam stop shadow. The fixed 3-D integration box in XENGEN does however limit the density of spots on a frame and the range of allowed frame widths. I should point out that the crystal to detector distance was set to be optimal for the XENGEN package but could have been shorter for both BUDDHA and XDS. In consequence the obtainable overall resolution for the same sample is higher for the BUDDHA and XDS software. In principle the 2-D integration box used in the BUDDHA software might be expected to give a poorer signal to noise ratio but in practice we find little evidence of this. We have been able to process data with XDS that was unprocessable with XENGEN.

Table 4. Software on the X100-A system

Phospholypase C data collected at Heidelberg (Hough,Hansen)

	XDS	**BUDDHA**	**XENGEN**
Integration:			
Observations Processed	36505	34086	38607
Obs. not Integrated	597	1569	200
Obs.with bad profile	919	-	-
Obs.output	34989	32517	38407
Scaling:			
Obs.Rejected on Input	17	9	20
Obs. Read	34972	32508	38387
R(F^2)	**4.01**	**6.30**	**5.58**
$R_w(F^2)$	4.54	4.65	4.52
R(100Å-10Å)	2.4	1.4	1.4
R(2.24Å-2.18Å)	9.6	22.0	17.1
Obs. Used	32487	27952	36549
Rejected as outliers	1	74	271
Sigma Multiplier	0.7	1.0	1.1
Unique Reflections	14355	13116	15837
%completion to 2.1Å	77	70	83

In the XENGEN software there is the option of using intensities estimated either by a profile fitting algorithm or by straight background from peak subtraction. We have never found the profile fitting to make any significant improvement to **R** (although this *is* a poor test of improvement) nor have we expected it to do so given the large residual modulations over our first chamber. We shall look at this more carefully with data measured on the new (and virtually residual modulation free) chamber, but initial indications are that it does improve the internal consistency for weakly diffracting crystals.

For (**F**) only we have so far only installed MADNES[5]. However within this program there are different methods of integration available. It is, I think, generally accepted by users of this instrument that the ellipsoidal masking algorithm is preferable to the dynamic masking algorithm. The other choices relate to how the non X-Ray background and the non-uniformity of response are treated (In FASTology these are the NUFTYPs). On a data set where the non X-Ray background is very stable we have run the same data through with different NUFTYPs. The results (**Table 5**), in terms of **R** , show that there are probable differences and that simplest appears best. The MADNES software is continually changing and the result has been a number of newer treatments which are compared using another (weaker) data set in **Table 5**. With the stronger data set the Kabsch profile analysis yielded little effect, but with the weaker set

the effect is very obvious. On scaling together the data processed with different NUFTYPs we found 0 6 and 0 8 scaled extremely well together (except at very low resolution) but that the weaker reflections for 0 2 are overestimated relative to 0 6 and 0 8. There is also some evidence that 3 8 overestimates weak reflections relative to CAD4 data.

Table 5. FAST System (MADNES)

Lysozyme Data Collected at Groningen (Kalk,Stezowsky) and Heidelberg

NUFTYP	Max. Resoln.	Overall R(F^2)	R(1000Å-10Å)	R(2.50Å-2.43Å)
0 2(i)	2.2Å	4.42	2.24	7.04
0 6(ii)	2.2Å	5.13	2.33	8.07
0 8(iii)	2.2Å	4.74	1.74	7.93
0 2	2.0Å	4.65	2.05	6.81
0 8	2.0Å	5.09	1.77	7.93
0 8+PA(iv)	2.0Å	5.19	4.28	7.43
			R(10-7Å)	R(2.04-2.09Å)
3 8(vi)	2.0Å	9.06	2.36	46.14
3 8+PA	2.0Å	7.11	3.17	32.44
3 11(v)	2.0Å	9.45	2.24	75.54
3 11+PA	2.0Å	7.40	2.69	38.26
3 12(v)	2.0Å	9.15	2.31	81.07
3 12+PA	2.0Å	7.45	2.64	41.33

(i)Non-Xray image measured and subtracted pixel by pixel before applying non-uniformity.(ii)First and last image used to give a background plane which is subtracted pixel by pixel.(iii)Knowledge of the detector gain used to separate gamma from non-gamma background.Continually updated global background used.Overflows trapped.(iv)As for (iii) but applying the coordinate transformation described in Reference 4 and using part of that authors code to perform a profile analysis.(v)Application of a background plane fitting alogorithm.(vi)The PreNUFTYP 3 signifies a monitoring and correction for changes in DC level on an image by image basis.

Currently (**D**) runs the ELMS software[8] although we have run one batch of frames (transfered over DECnet to a μVAX) through XDS with similar results. Current software work indicates improvements of around 2% in **R** may be obtainable, but we have not yet checked for resulting systematic errors.

Cell Dimensions:

We tend not to use the instruments for determination of unit cell parameters but on (**X**) when running XENGEN or XDS we have often obtained inaccurate cell dimensions. I believe this is because these software systems assume the beam is normal to the calibration face plate. The (**F**) instrument usually gives cell dimensions as determined on a diffractometer or from precession photography.

Crystallographic Comparisons:

There are several means by which we can estimate and compare data quality from the instruments. These are

1) Comparison of Internal Consistency of a data set.

For this we use as a criteria

$$\mathbf{R}(F^2) = \sum | F^2 - \langle F^2 \rangle | / \sum \langle F^2 \rangle,$$

the summation being over symmetry equivalents after scaling. Where we have done this we have mostly used the same scaling software[6]. Comparisons are given in **Tables 6 and 7.** It must be stressed that the absolute **R** values are not very meaningful since they depend heavily on where one choses to set the resolution limit for a particular crystal. In **Table 7** crystals were usually from the same batch and of similar size. Also noted for obvious reasons are the mean data collection rate

(reflection/minute of exposure time) and the percentage of observations rejected in the calculation of **R** either for instrumental reasons (**F** software only) and/or as outliers from the distribution of observations.The high number of rejections in the (**F**) data (**Table 6A**) requires some explanation. Some 90-95% of these are weak reflections with error flags that indicate the reflection to be on the edge of the integration box. Since this is determined on the basis of a percentage of peak maximum the flags are statistically unreliable and there is probably no real cause to reject the reflections. Their inclusion, however, increases **R** by 1-2%. Comparing **Tables 6A and 6B** gives an indication of the enormous improvement evident for our new (**F**).

Table 6A.Data Collections on the same Crystal on X and F

Material	F/X	d_{min} Å	N_{meas} N_{Unique}	$R(F^2)$	R(F) vs Diffr.	Rate	% Rejects
Lysozyme 79x79x38							
Xtal 1	F	2.5	18808(3848)	9.1	7.4(4Å)	10	13
	X	2.5	13770(3546)	8.6	6.1(4Å)	10	5
Xtal 2*	F	2.0	15431(4439)	4.9		21	10
	X	2.0	37738(7767)	4.8		25	2
Rhodanese 156x49x42							
	F	2.0	8366(7047)	3.8	5.6	9	14
	X	1.9	17373(14281)	3.7	5.9	17	5
Phospholypase A2 47x47x103							
Xtal1	F	1.9	55321(8755)	9.2		12	30
	X	1.8	46366(10914)	4.4		15	6
	F	1.9	20555(6506)	6.3		12	25
Xtal 2	X	1.8	33899(8520)	3.8		11	6
	F	1.9	44165(7813)	5.7		11	27
	X	1.8	19294(6322)	4.0		14	8

* FAST main beam intensity improved

Table 6B.Data Collections on the same Crystal on X and F(New)

Material	F/X	d_{min} Å	N_{meas} N_{Unique}	$R(F^2)$	R(F) vs Diffr.	Rate	% Rejects
Lysozyme 79x79x38							
	F	2.0	5930(2414)	7.1		22	6
	X	2.2	7360(5154)	8.0		23	1
ROP 59x40x28							
	F	2.0	6883(3477)	4.2	7.7(4Å)	4	<1
	X	2.0	6895(4355)	6.8	11.2(4Å)*	4	6
CPA 52x60x47							
	F	2.0	34634(14295)	3.7		32	<1
	X	2.0	40056(16858)	5.9		26	3
Adenovirus DBP 79x76x67							
	X	3.0	10085(5968)	8.0		18	2
	F	2.6	12360(6821)	7.1		20	<1

* This is the only data set in Table 6 where crystal radiation damage may be limiting

2) Comparison with diffractometer data.
The assumption of course being that we expect, where diffractometer data is

measurable,to obtain better quality (absorption corrected) data therefore giving a reference against which to compare different area detector data sets. Where such comparisons have been made the data has been included in **Tables 6 and 7**, the **R** being on F (of the symmetry averaged data sets) rather than on F^2 It is reassuring to find that the trends do not conflict with the trends in internal consistency.

Table 7A.Data Collections on the same material on X and F

Material	F/X	d_{min} Å	N_{meas} N_{Unique}	$R(F^2)$	R(F) vs Diffr.	Rate	%Rejects
Adenovirus DBP * 79x76x67							
	F	2.8	20628(6030)	10.5		9	31
	X	2.7	47194(10121)	7.5		22	4
Collagenase* 112x112x166							
	F	2.3	51778(17792)	9.5	3.6(3Å)	20	21
	X	2.5	66164(18232)	9.4	3.9(3Å)	92	1
	X	2.1	85592(32147)	10.0		50	2
Seryl tRNA Synthetase 144x91x70							
	F	3.5	9860(25265)	8.9		9	12
	X	2.7	18724(30173)	6.3		17	<1
AAT 88x78x83							
	F	2.6	26520(11697)	10.5		32	15
	X	2.7	91055(26474)	6.2		19	14
DNase1/Actin * 133x56x110							
Pb Cocryst.	F	3.0	58642(13947)	8.4		24	32
	X	3.0	41563(15705)	6.6		29	3
Native	F	2.9	53687(14749)	8.3		22	35
	X	2.8	43281(16418)	10.5		30	5
Pb deriv.	F	3.0	51447(15158)	9.7		21	27
	X	3.5	28101(8853)	8.2		20	3
Phospholypase C * 90x90x74							
	F	2.1	57322(18304)	6.5		42	22
	X	2.2	91734(16193)	6.6		42	1

* FAST main beam intensity improved

Table 7.Data Collections on the same material on X and F(New)

Material	F/X	d_{min} Å	N_{meas} N_{Unique}	$R(F^2)$	R(F) vs Diffr.	Rate	%Rejects
Colicin 73x73x171							
	F	2.5	20175(7555)	8.3		15	<1
	X	2.6	20015(5343)	8.7		9	<1

We are still analysing newer data but, for the sake of completeness, some original lysozyme data will be considered further. We have two data sets measured on each instrument, the scaling and agreement of which are shown in **Figure 4** for (**X**) and **Figure 5** for (**F**). Scaling and agreement between instruments is shown in **Figure 6**. Finally scaling and agreement of the first crystal against CAD4 diffractometer data (to 4Å only) is shown in **Figure 7**. Scaling of data sets for (**X**) is fairly good and for (**F**) is reasonable except at higher resolution. Clearly the agreement between the two (**X**) data sets is rather better although this probably because, for the (**F**) data sets, the main beam intensities differed substantially. In scaling (**X**) *vs* (**F**) the trends are different for the two crystals and no conclusion can be drawn. Against limited resolution CAD4 data

(Xtal 1 only) both (X) and (F) show some underestimation of strong intensities. In terms of agreement (X) is clearly better than (F) at lower resolution. Note however that Hansen *et al* [7] have made a more thorough comparison on Phospholypase C data, which also suggests an overestimation of weak reflections on both instruments relative to diffractometer data.

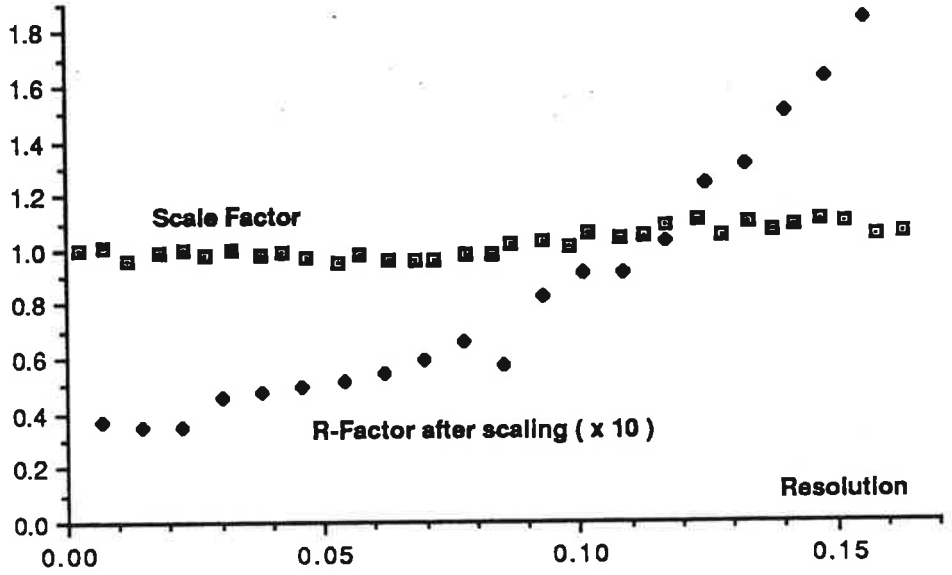


Figure 4. Scaling and agreement between two lysozyme data sets on the (X) instrument. Resolution here and elsewhere is $4\sin^2\theta/\lambda^2$

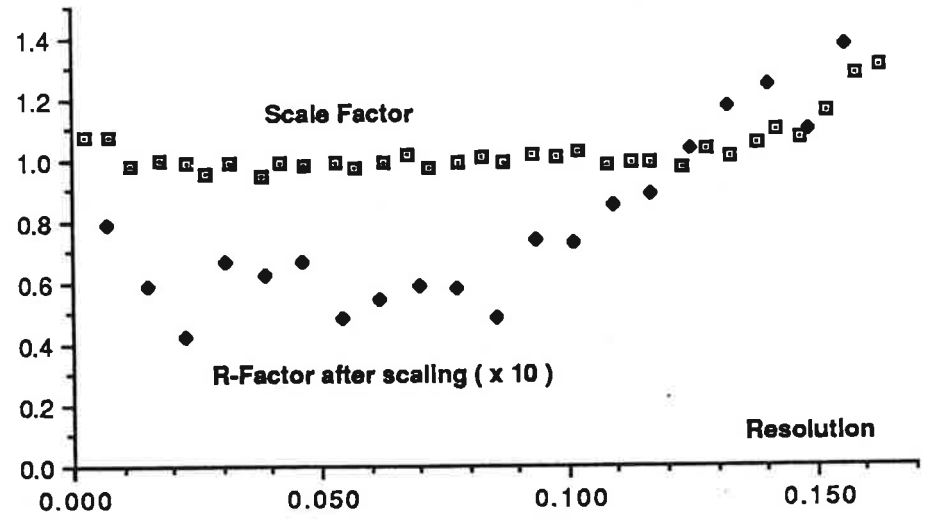


Figure 5. Scaling and agreement between two lysozyme data sets on the (F) instrument.

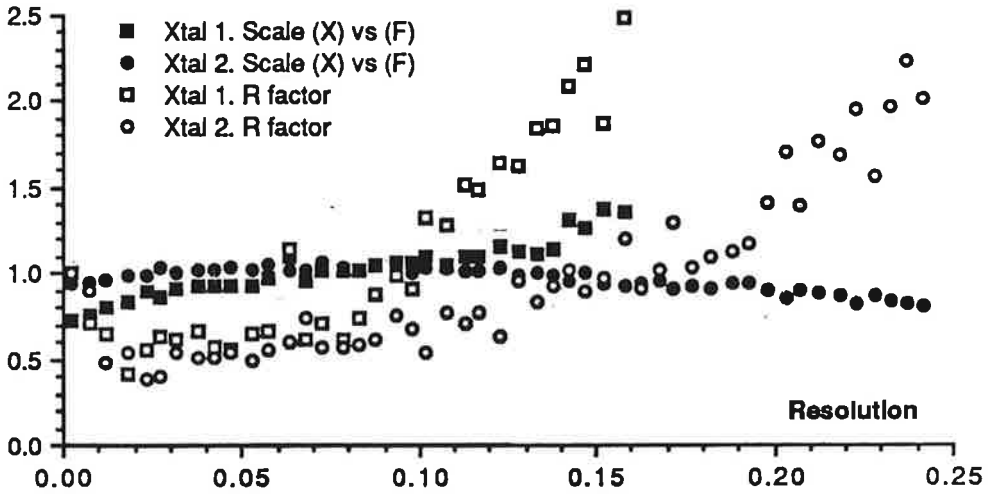


Figure 6 Scaling and agreement between data sets measured on two lysozyme crystals on each of (**X**) and (**F**).R factors are multiplied by 10.

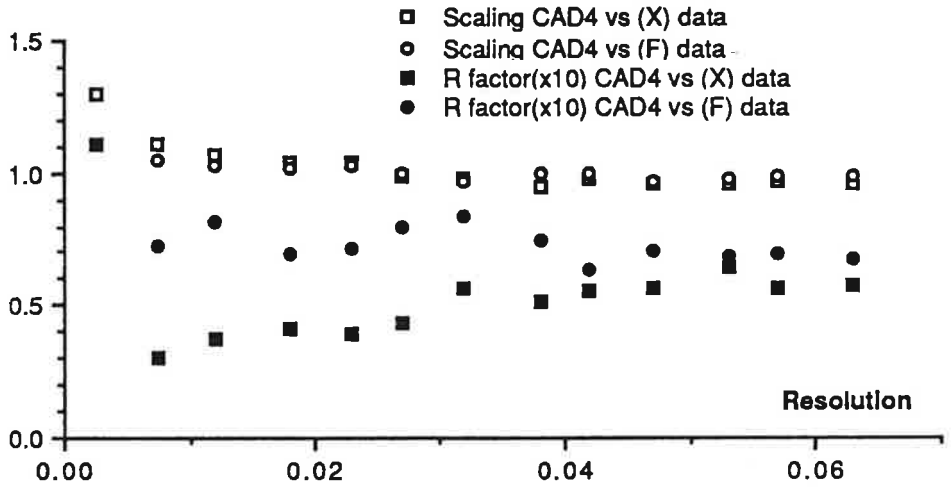


Figure 7. Scaling and agreement of CAD4 with (**X**) and (**F**) data sets on the same crystal.

3) Measurement of anomalous signal.

It seemed to us that a good test of data quality is to look to see what resolution the data contain an anomalous signal, for example by looking at anomalous difference Pattersons. Clearly from the point of view of a comparison a metal complex is preferable to a derivative soak. Therefore, although we do have some indicative data on this aspect , (**Table 8**) it is not in the form of a direct comparison. For the data in **Table 8** no effort was made to optimise conditions for accurate measurement of anaomalous signal. Further data under this head are being analysed. One should point out that despite our inability to detect anomalous signals on the (**F**) instrument others (notably those with some form of temperature stabilisation) have done so. We have still to try to repeat this sort of measurement on the new (**F**) instrument.

Table 8.Anomalous Signal

Instrument	Derivative	Sites	f''	Comments
Seryl-t-RNA-Synthetase (Cusack)				
F	Dy	1	8	No anomalous signal even at 6Å (isomorphous differences to 3.5Å)
X	U	8	16	Usable anomalous signal
X	Hg	2	9	No anomalous signal
Adenovirus DBP				
X	Dy	1	8	Anomalous signal to 5.5Å
X	Sm	1	13	Anomalous signal to 4Å
X	Au	1	8	No anomalous signal
DNAse1/Actin				
X,F	Pb,Hg			No useful anomalous data recorded
Colicin				
X	Hg			No useful anomalous data recorded

4) Comparison of Cullis R-Factors or Signal to Noise ratio in Isomorphous Difference Fourier Maps.

These are poor tests if done on different crystals because the occupancy in the derivative is difficult to control. We only have one comparative measurement at present, on the DNAse1/Actin complex and a Pb derivative and in this case we were almost certainly looking at a difference in derivatisation rather than an instrumental effect.

5) Refinement of data against coordinates of a known structure.

It is tempting to assume this is the ultimate test of the quality of a native data set. This is not actually the case unless comparing a very highly refined structure with data collected to lower resolution. Although we have no results to date we do have some tests underway.

Table 9.DIP100 Measurements

	Exposure Time(hrs)	Total Time	Resoln (Å)	N(obs) N(unique)	R(I)	Rejects %	Completeness %
Barnase (dGpC) 57x57x85	15	36	2.5	9262 (5397)	12.8	9	84[1]
StRS/BrATP 141x91x70	80	144	2.6	49325 (18962)	9.8	3	75[2]
Lysozyme 80x80x38	10	24	2.0	30493 (8159)	6.1	9	91[3]
.....							[4]
Elastase (complx) 52x58x75	45	60	2.1	47046 (13265)	7.1	<1	92
Aerolysin 104x104x222	20	30	4.0	10950 (6405)	14.2	<1	58
CPA 52x60x46	20	33	1.85	83279 (22022)	7.4	<1	91
CPA (-Zn)	20	33	1.85	83227 (22314)	11.8	<1	91

[1] *cf* 11.8% to 3.0Å on FAST. [2] 2 crystals. [3] Fullys only Partials only R(I) 9.3% 19509(7565) Rejects 14%. [4] The line represents a major software change.

Image Plate Device:

We are currently spending a little time checking the instrument calibration and setting up the software to make measurement routine. We have made some measurements, summarised in **Table 9**, which probably do not reflect the full capability of the instrument.
Although we anticipate poorer signal to noise ratios for larger oscillation frames, I am fairly certain that the data quality is software rather than hardware limited and this is being corrected. In particular scaling of the partial reflections against the fully recorded reflections has shown a systematic trend indicating that the beam description is too simple. The fully recorded reflections , however, scale rather well and agree rather well with data collected on (**X**) and our old (**F**). The results of such an experiment (again on lysozyme) are shown in **Figure 8.** The scaling and agreement with (**X**) are better, but we hope would be similar with data from the new (**F**).

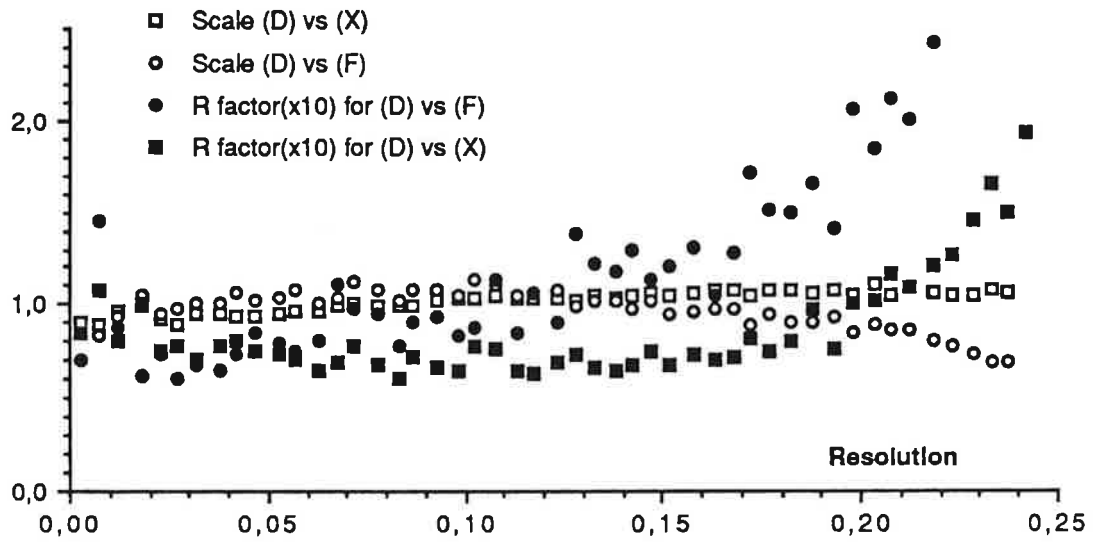


Figure 8. Scaling and agreement of (**D**) against (**F**) and (**X**) data sets.Same lysozyme crystal for (**F**) and (**X**) but different crystal for (**D**).

Postamble:

On the basis of the above data there is little to chose between (**X**) and (**F**) except to say that it has taken us much longer to get equivalent quality data from (**F**). Where we have worked with strongly diffracting crystals the results from (**X**) and (**F**) have always been comparable. We obtain improved results on (**F**) with a more intense beam which poses the question as to whether (**F**) would not be better run with a ß filter rather than a monochromator, however we anticipate enhanced radiation damage if we do this and will probably not try it until cryogenic work becomes more routine. We have a clear indication that temperature stabilisation of the (**F**) detector is essential for good quality data on weakly diffracting crystals.

The crystallographic comparisons are always unsatisfactory for protein crystals and I am still looking for the 'ideal calibration crystal'. What I would like is something with all axes in excess of 30Å, no radiation damage, high symmetry, hardness so that it may be ground to a sphere and a structure refinement to better than 1Å resolution. Any suggestions?

Finally I would point out that, where our local demands permit, we accept

visitors (obviously we cannot provide financing) wishing to make comparative measurements or simply wishing to become familiar with one of the instruments. There is likely to be some interest in the (**D**) system, which we have on loan for a limited period ,therefore contact me sooner rather than later.

Acknowledgements:

I thank my colleagues at EMBL as well as Bauke Dijkstra,Kor Kalk,Arnaud Ducruix,Ed Hough,Lars Hansen, Gil Shoham ,Rachel Nechushtai and John Stezowsky for providing some of the data used in this article.

References:

[1]Thomas,D. (1989),*Proc.R.Soc.Lond.* **A425** , 129 and in press.
[2]Howard,A.J. ,Gilligand,G.L. ,Finzel,B.C. ,Poulos,J.L. ,Ohlendorf,D.H. ,& Salemne,F.R. (1987) *J.Appl.Cryst.* ,**20** ,383.
[3]Blum,M. Metcalf,P. ,Harrison,S.C. & Wiley,D.C. (1987) *J.Appl.Cryst.* **20** 235.
[4]Kabsch,W. (1988) *J.Appl.Cryst.*,**21**, 916.
[5]Messerschmidt, A. & Pflugrath ,J.W. (1987) *J.Appl.Cryst.*, **20**, 306.
[6]Programs originating from the University of Groningen. Contact Bauke Dijkstra.
[7]Hansen,L.*et al*, to be published.
[8]Tanaka,I. *et al* ,to be published.

Misleading results of the self-rotation function arising from (systematically) incomplete data

Joerg Kallen and Richard Pauptit,
Biocenter, University of Basel,
CH-4056 Basel, Switzerland.

We have noticed that the self rotation function can give misleading results when data are (systematically) missing. We found that by supplementing the incomplete data set with data from an independent previous collection the error can be removed, and that the error can be reproduced when the data missing from the incomplete data set are excluded from a full data set. We would like to give two examples:

(1) Phosphoserine aminotransferase (PSAT) from E.coli:

PSAT crystallizes in $P2_12_12_1$ with one dimer per asymmetric unit. Diffractometer (CAD4) data revealed the orientation of the local dyad, which was subsequently confirmed by the heavy-atom positions. A FAST data set (unfortunately incomplete near k=0) at first gave the surprising self-rotation function results described below (Fig 1a,b). The $\kappa = 180°$ sections were produced by the POLARRFN program from the CCP4 program collection. F's in the range 8 Å to 4 Å and a sphere of radius 20 Å in Patterson space were used. ω is horizontal and ϕ vertical:

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v
0›	100	42	22	21	1	7	-	-	-	-	-	-	-	7	1	21	22	42	100
5›	100	42	22	21	2	6	.	-	-	-	-	-	-	2	4	21	21	29	46
10›	100	42	22	21	3	4	2	-	-	-	-	-	1	1	6	19	15	19	26
15›	100	42	22	20	4	1	3	-	-	-	-	-	1	3	4	8	6	13	22
20›	100	42	22	18	5	-	4	-	-	-	-	-	-	10	5	5	11	14	11
25›	100	42	21	16	7	-	5	1	.	-	-	-	.	6	1	28	12	17	19
30›	100	42	21	14	11	-	6	1	-	-	-	-	-	5	11	20	9	11	12
35›	100	42	21	12	14	4	5	3	-	-	-	-	-	8	8	11	5	9	7
40›	100	42	20	11	15	8	5	5	1	-	.	-	.	2	8	13	8	7	7
45›	100	42	20	9	14	6	4	3	2	-	-	-	-	-	3	8	10	4	-
50›	100	43	20	9	12	4	5	.	2	4	2	-	-	-	3	2	6	7	7
55›	100	43	20	9	11	5	9	6	2	6	4	-	.	6	4	10	14	13	7
60›	100	43	19	11	11	8	8	13	5	6	4	6	3	6	5	13	18	18	12
65›	100	44	19	13	13	9	3	9	5	7	6	11	4	3	7	9	11	13	19
70›	100	44	19	16	14	10	4	8	8	11	4	5	6	5	6	8	15	12	11
75›	100	44	19	19	15	11	8	13	10	14	11	7	6	4	7	11	17	22	22
80›	100	45	18	21	14	11	9	9	6	8	11	7	3	4	11	14	17	23	26
85›	100	45	18	22	11	11	13	5	6	9	4	4	11	7	5	20	26	27	46
90›	100	45	18	22	9	11	17	6	6	12	6	6	17	11	9	22	18	45	100

Fig 1a:
CAD4 data
(6627 reflections,
99% complete)

GOOD

peak at
$\omega = 75°$, $\phi = 25°$

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v
0›	100	67	43	33	22	17	10	5	.	.	.	5	10	17	22	33	43	67	100
5›	100	67	43	33	21	16	10	4	-	.	1	4	7	15	23	31	43	61	71
10›	100	67	44	33	19	13	9	1	-	1	2	5	6	14	21	30	42	54	59
15›	100	67	44	33	18	11	8	1	.	4	2	3	7	10	17	27	39	48	53
20›	100	67	44	34	18	10	6	1	4	4	2	2	5	14	19	27	39	45	47
25›	100	67	44	34	19	10	3	1	6	6	2	4	8	12	21	34	38	46	52
30›	100	67	43	34	20	9	.	2	5	6	3	1	8	9	17	30	39	47	52
35›	100	67	43	32	20	8	-	3	5	4	3	1	6	8	15	25	36	49	52
40›	100	67	42	30	17	5	-	2	4	5	6	1	3	4	14	23	34	49	52
45›	100	66	41	26	11	.	-	1	1	1	2	-	-	-	6	19	33	44	49
50›	100	66	40	23	6	-	-	-	-	-	.	-	-	-	2	17	29	42	52
55›	100	65	38	20	3	-	-	-	-	-	-	-	-	-	1	16	30	41	52
60›	100	65	36	18	1	-	-	-	-	-	-	-	-	-	-	14	33	43	52
65›	100	64	33	16	-	-	-	-	-	-	-	-	-	-	-	11	29	45	52
70›	100	63	31	14	-	-	-	-	-	-	-	-	-	-	-	9	29	46	47
75›	100	63	29	12	-	-	-	-	-	-	-	-	-	-	-	8	31	51	53
80›	100	62	27	11	-	-	-	-	-	-	-	-	-	-	-	7	29	51	59
85›	100	62	26	10	-	-	-	-	-	-	-	-	-	-	-	10	30	52	71
90›	100	62	25	10	-	-	-	-	-	-	-	-	-	-	-	10	25	62	100

Fig 1b:
FAST data
(5180 reflections,
78% complete)

BAD

The logical but wrong conclusion was that the FAST data were suspect, since they did not reproduce the correct solution for the orientation of the dimer. However, the results below (Fig 1c,d) suggest that it is the incompleteness of the data that gives an erroneous result for the self rotation function:

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v
0›	100	66	42	34	27	25	18	15	11	12	11	15	18	25	27	34	42	66	100
5›	100	66	42	33	26	23	18	13	9	9	10	14	15	21	26	32	41	59	72
10›	100	66	42	33	23	18	16	8	5	7	8	10	12	17	22	30	40	53	58
15›	100	66	42	33	20	14	12	4	2	5	2	3	8	11	16	26	36	45	50
20›	100	66	43	33	18	10	5	.	1	-	-	-	1	11	16	25	37	42	43
25›	100	66	43	33	17	7	-	-	-	-	-	-	1	6	19	33	36	45	49
30›	100	66	43	32	17	4	-	-	-	-	-	-	.	4	16	30	38	46	51
35›	100	65	42	30	16	2	-	-	-	-	-	-	-	3	14	25	38	49	54
40›	100	65	41	27	13	-	-	-	-	-	-	-	-	-	13	24	36	51	55
45›	100	65	40	24	8	-	-	-	-	-	-	-	-	-	6	21	35	47	52
50›	100	64	38	20	3	-	-	-	-	-	-	-	-	-	3	19	33	44	55
55›	100	63	36	18	1	-	-	-	-	-	-	-	-	-	4	20	35	44	54
60›	100	63	34	16	-	-	-	-	-	-	-	-	-	-	4	20	37	45	51
65›	100	62	31	14	-	-	-	-	-	-	-	-	-	-	1	17	32	43	49
70›	100	61	29	13	-	-	-	-	-	-	-	-	-	-	-	15	31	44	43
75›	100	60	27	12	-	-	-	-	-	-	-	-	-	-	-	12	31	48	50
80›	100	60	25	11	-	-	-	-	-	-	-	-	-	-	-	9	27	47	58
85›	100	59	24	11	-	-	-	-	-	-	-	-	-	-	-	10	28	50	72
90›	100	59	23	11	-	-	-	-	-	-	-	-	-	-	-	11	23	59	100

Fig 1c:
CAD4 data, excluding reflections not measured in FAST data set (5134 reflections, 77% complete)

BAD

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v	v
0›	100	43	24	22	2	8	.	-	-	-	-	-	.	8	2	22	24	43	100
5›	100	43	24	22	2	7	1	-	-	-	-	-	.	4	6	21	23	32	46
10›	100	43	24	22	4	5	3	-	-	-	-	-	1	3	10	22	18	22	28
15›	100	43	24	22	5	3	5	-	-	-	-	-	1	4	9	13	13	19	28
20›	100	43	24	22	8	1	5	-	-	-	-	-	-	12	8	13	19	21	21
25›	100	43	23	21	10	.	5	-	-	-	-	-	1	5	15	31	19	24	29
30›	100	43	23	19	13	1	2	-	-	-	-	-	-	4	10	21	15	18	21
35›	100	44	22	17	15	4	-	-	-	-	-	-	-	7	9	12	8	14	14
40›	100	44	22	14	15	6	-	-	-	-	-	-	2	5	11	17	12	12	11
45›	100	45	22	13	13	5	.	-	-	-	-	-	-	1	5	13	15	12	3
50›	100	45	22	12	11	3	2	-	-	-	-	-	-	3	6	9	12	15	11
55›	100	46	23	13	10	5	6	3	-	-	-	-	2	11	9	16	20	19	14
60›	100	47	23	14	12	9	6	10	-	-	-	2	4	9	10	18	23	25	21
65›	100	47	22	16	14	11	2	4	-	-	-	5	2	5	13	14	16	22	29
70›	100	48	22	18	16	11	3	2	1	2	-	-	2	5	12	14	20	20	21
75›	100	49	22	20	17	11	6	6	4	5	4	.	2	3	12	16	23	28	28
80›	100	49	21	21	16	12	5	2	-	-	2	2	.	3	14	16	19	26	28
85›	100	49	21	22	14	13	9	.	-	.	-	-	7	7	9	22	27	30	46
90›	100	49	21	22	13	14	14	3	-	3	-	3	14	14	13	22	21	49	100

Fig 1d:
FAST data, supplemented by CAD4 data (6673 reflections, 100% complete)

GOOD

peak at
$\omega = 75°$, $\phi = 25°$

(2) Porin from E.coli:

This membrane protein crystallizes as sets of trimers exhibiting non-crystallographic three-fold symmetry. Data collection from a putative derivative was not completed for technical reasons, nonetheless the data were examined for utility. A strong peak ($\omega = 57°$; $\phi = 30°$) corresponding to a local three-fold could be observed in the self-rotation function, (Fig 2a), some 20° away from the values obtained for the native protein ($\omega = 54°$, $\phi = 51°$; Fig 2b). This result appeared so convincing that a number of hypotheses were formulated that could account for the tilt of the trimer upon derivatization without changing the unit cell parameters, and experiments were designed to obtain a "tilted" native data set. In light of the PSAT results described above it appeared that the shift might be an artifact of incomplete data. Indeed, when the missing

Rotation angle kappa =120
polar coordinates of rotation Axes are phi (vertical) and omega (horizontal)

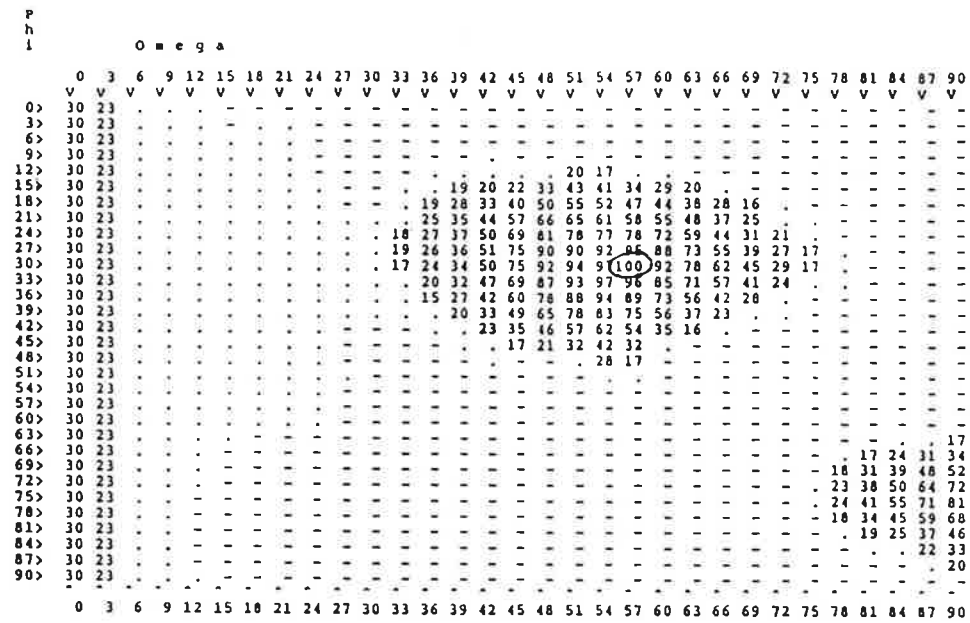


Fig 2a

Rotation angle kappa =120
polar coordinates of rotation Axes are phi (vertical) and omega (horizontal)

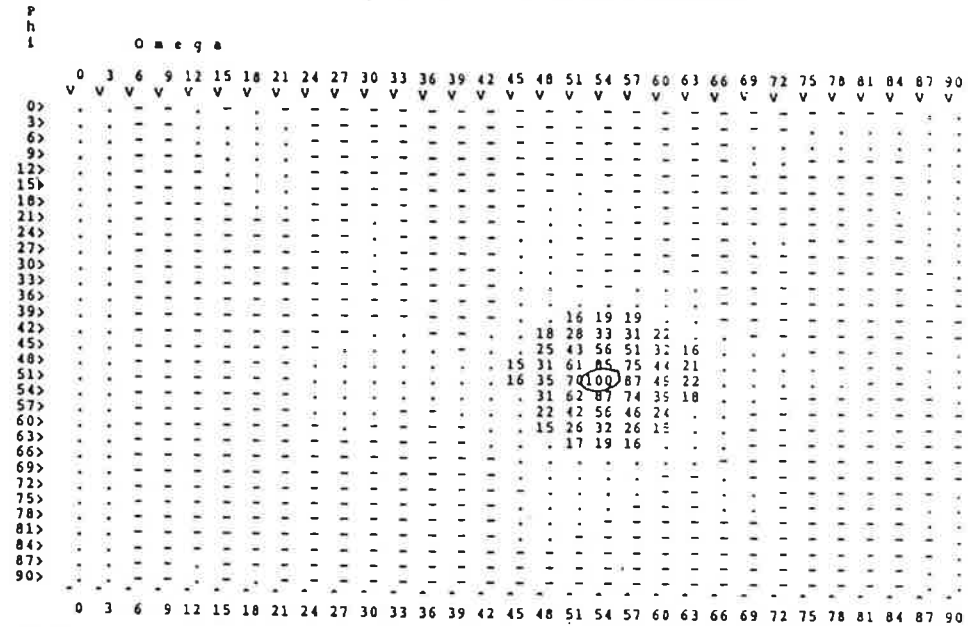


Fig 2b

Fig 2c

Rotation angle kappa =120
polar coordinates of rotation Axes are phi (vertical) and omega (horizontal)

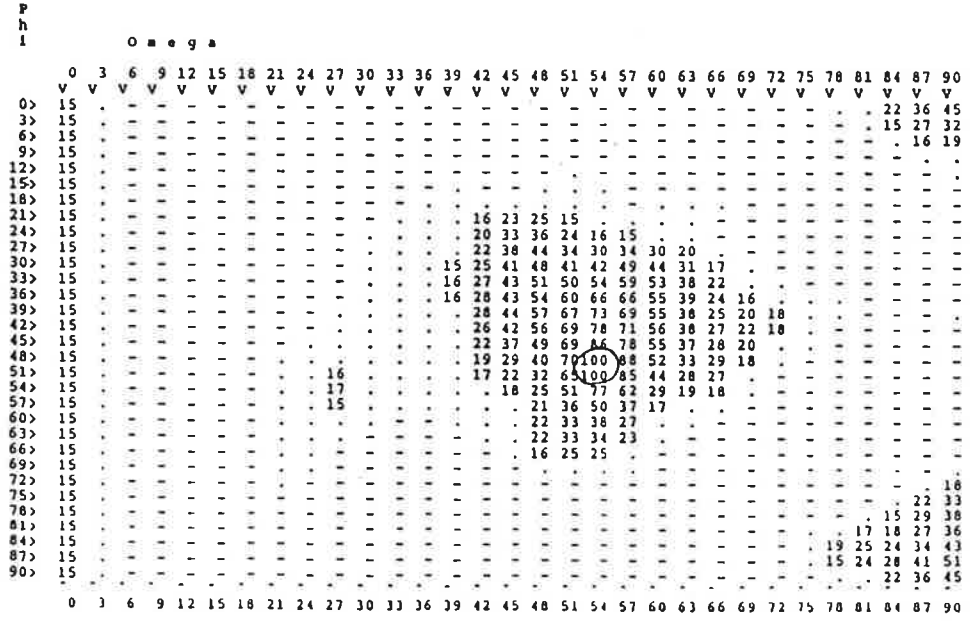
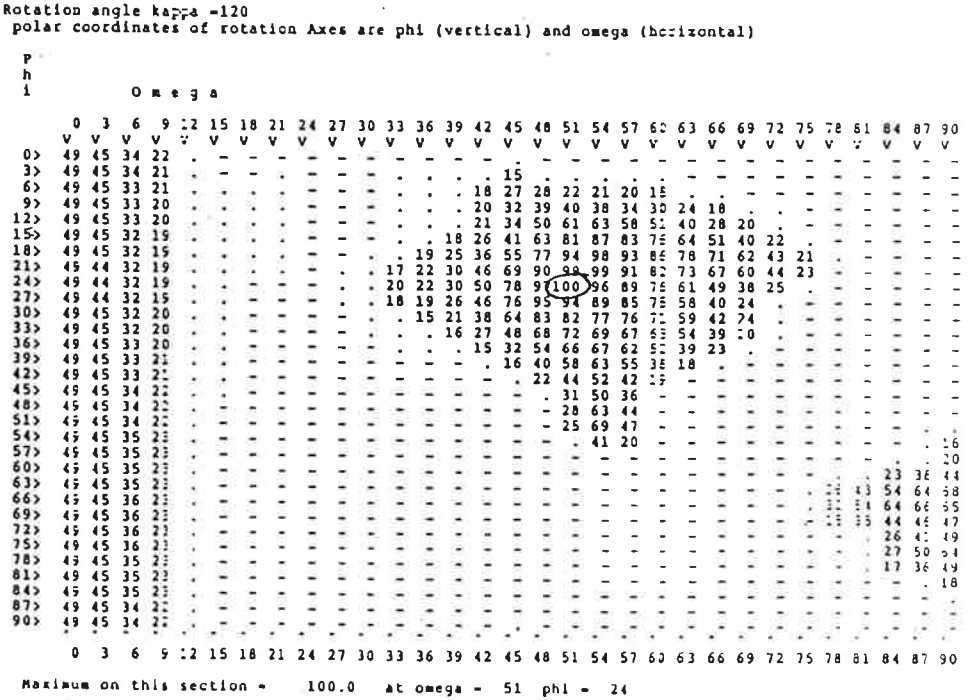


Fig 2d

Rotation angle kappa =120
polar coordinates of rotation Axes are phi (vertical) and omega (horizontal)



data were included using data from a native data collection, the peak appeared at the long established triad position (ω = 54°, ϕ = 51°, Fig 2c). Furthermore, when the missing data were removed from the complete native data set, a shift of the peak to a similar erroneous position (ω = 51°, ϕ = 24°, Fig 2d) was observed.

The κ = 120° sections in Figs 2a-d were produced by program POLARRFN using data between 15 Å and 8.5 Å resolution, and an integration radius of 50 Å of which the inner 20 Å were smoothed.

Fig. 2a putative derivative data, 61% complete
Fig. 2b native data, 99% complete
Fig. 2c derivative data supplemented by native data
Fig. 2d native data excluding data missing in Fig 2a

The Determination of d_{min}, $\lambda_{min.}$ and $\lambda_{max.}$ from Laue Photographs Using Measurements of the Clear Gaps Surrounding Zone Lines in Their Gnomonic Projections

Cruickshank[1] D.W.J., Carr[2] P.D. and Harding[2] M.M.

1. Chemistry Department, UMIST, Manchester M60 1QD
2. Chemistry Department, University of Liverpool, P.O. Box 147, Liverpool, L69 3BX

ABSTRACT

The width of the clear envelopes surrounding principal zone lines on the gnomonic projections of Laue photographs can be used to determine the 'soft limits' $d_{min.}$ and $\lambda_{min.}$ for a crystal of known unit cell. If the detector is capable of measuring reflections out to the maximum Bragg angle, the value of $\lambda_{max.}$ may also be determined.

The determination of 'soft limits'

In a recent review of the recording and analysis of synchrotron x-radiation Laue diffraction photographs, Helliwell et al. (1989) show that the choice of values of the so called 'soft limits' of $\lambda_{min.}$, $\lambda_{max.}$ and $d_{min.}$ is important, and that in particular the resolution limit $d_{min.}$ is "especially critical". They also described the method by which these values are usually obtained, i.e. trial-and-error matching of the Laue photograph (or an enlarged print of it) with patterns predicted using the LGEN routine. In practice this technique is rather time consuming and somewhat subjective. The method described in this paper requires a simple measurement to be made from the gnomonic projection of the pattern followed by a straightforward calculation.

Figure 1 shows the gnomonic projection in relation to the Laue geometry and reciprocal lattice. The spots on the Laue photograph (or other detector image) have angular coordinates $(2\Theta,\beta)$ where 2Θ is the Bragg angle and β is the azimuthal angle. The gnomonic projection transforms the angular coordinates to $(\pi/_2 - \Theta, \pi+\beta)$ and represents them on a plane normal to the incident beam. This may be regarded, for a given crystal orientation, as a transformation from beam directions (a beam is a diffracted x-ray beam) to ray directions (a ray is a central line in the reciprocal lattice passing through the origin (0,0,0) and reciprocal lattice points (h,k,l), (2h, 2k, 2l) ..... (nh, nk, nl)). Thus the gnomonic projection may be regarded as an inverted fish-eye view of the reciprocal lattice as seen from the origin.

We can assume that the limit of the clear envelopes bordering a zone line in the gnomonic projection corresponds to the intersection of two 'delimiting planes' with the external surface of the accessible region of reciprocal space. These delimiting planes of reciprocal lattice points (Bravais) lie parallel directly "above" and "below" a given zone. The surface of accessible reciprocal space corresponds to the $^1/\lambda_{min.}$ sphere when of $\Theta < \Theta c$ (where $\Theta c = \sin^{-1}(\lambda_{min.} D^*/2)$) or to the $D^*=1/d_{min.}$ sphere for values of $\Theta > \Theta c$. We can derive expressions (Cruickshank, Carr and Harding 1990)) that relate the minimum width of the clear gaps on the gnomonic projection to either $^1/\lambda_{min.}$ or D^* depending on the Bragg angles of the diffracted spots.

If $\Theta < \Theta c$ the nearside (high Θ side) gap can be used to determine $\lambda_{min.}$ using the following expression:

$$\lambda_{min.} = \frac{2\,Y_o\,|P|\sin\psi}{(Y\delta - 2Y_o\cot\psi + \mathrm{cosec}^2\psi)} \qquad (1)$$

Where Y_o = minimum width of gap envelope on gnomonic projection

$1/P$ = interplanar distance for this zone ($P[uvw]$ is distance from the origin to the point uvw in the direct lattice)

ψ = inclination of zone axis to the x-ray beam. For unit projection distance, the perpendicular distance from the projection centre to the zone line is cot ψ.

If $\Theta > \Theta c$ the gap can be used to determine D* using the expression:

$$D^* = \frac{2}{P(Y_{o+} - Y_{o-})\sin^2\psi} \qquad (2)$$

where $Y_{o+} - Y_{o-}$ = the total width of the gap from both sides of the zone line.

Equation (2) is an approximation valid when $P^2\,D^{*2}_{max}\sin^2\psi >> 1$. If the detector is capable of measuring spots out to the maximum Bragg angle Θ_{max}, we can use Bragg's Law to determine λ_{max}:

$$\lambda_{max} = 2\sin\Theta_{max}/D^*_{max} \qquad (3)$$

Figure 2 represents a Laue photograph collected from a crystal of phospholipase c on station 9.5 of the SRS along with the corresponding gnomonic projection. The constructions show the minimum width and cotψ measurements for the $[1\bar{2}1]$ and $[4\bar{5}3]$ zone lines. Substituting these values in (1) and (2) gives:

$$\lambda min = \frac{2(1.55)(618)\sin(2.56°)}{(1.55)^2-3.1(22.4) + \text{cosec}^2(2.56°)} = 0.2 \text{ Å}$$

$$D^* = \frac{2}{(214)(0.88)\sin^2(8.92°)} = 0.44 \text{ Å}^{-1} \quad (dmin = 2.2 \text{ Å})$$

Because the diffracted spots from the crystal extended to beyond the edge of the film, it was not possible to use equation (3) to determine $\lambda_{max.}$ in this case.

The values calculated above were used to predict the diffraction pattern using the LGEN routine. This gave a good match on the first attempt, and led to satisfactory processing of the films through the rest of the program suite. Thus we find this method a quick and objective way of determining the values of $d_{min.}$ and $\lambda_{min.}$

The gnomonic projection may also be used to refine the direct beam position (the zone lines are bent and tilted if the direct beam position is in error) and is exceedingly useful when trying to index Laue patterns both from crystals of known and unknown unit cells. (See below) Programs for producing and displaying gnomonic projections from optical density scanned images have been written and are available from PDC.

Indexing the Laue pattern from a crystal of unknown cell in a general orientation

For crystals in a general orientation the gnomonic projection is a series of straight lines (zones) which often intersect to give several conspicuous triangles. The intersection of these zones occur at the

positions of prominent nodels (See figure 2). If one picks three of these prominent nodels which form the apices of such a triangle (ABC on figure 2) and arbitrarily assigns them indices of (100), (010) and (001), it is then possible to identify four other prominent spots which will have indices of (011), (101), (110) and (111). These are marked as points D,E,F and G on figure 2.

One can use the directions of A , B and C from the origin of reciprocal space to determine the inter axial cell angles $\alpha^*,\beta^*,\gamma^*$ corresponding to the index assignment that has been arbitarily made. Similarly the ratio of the cell lengths a^*/b^* can be determined from the ratio of the sines of angles between rays (010)B, (110)F and (100)A. The ratios for b^*/c^* and c^*/a^* can also be calculated.

Hence it is possible to derive a (non-conventional) unit cell from a triangle on the gnomonic projection. A cell reduction procedure allows this to be converted into its conventional cell. The GENLAUE program can then be used to refine this cell and it's orientation. This unit cell will contain a scale factor relating the absolute cell edges to those determined from the cell reduction procedure. This scale factor can be determined if $\lambda_{min.}$ is known.

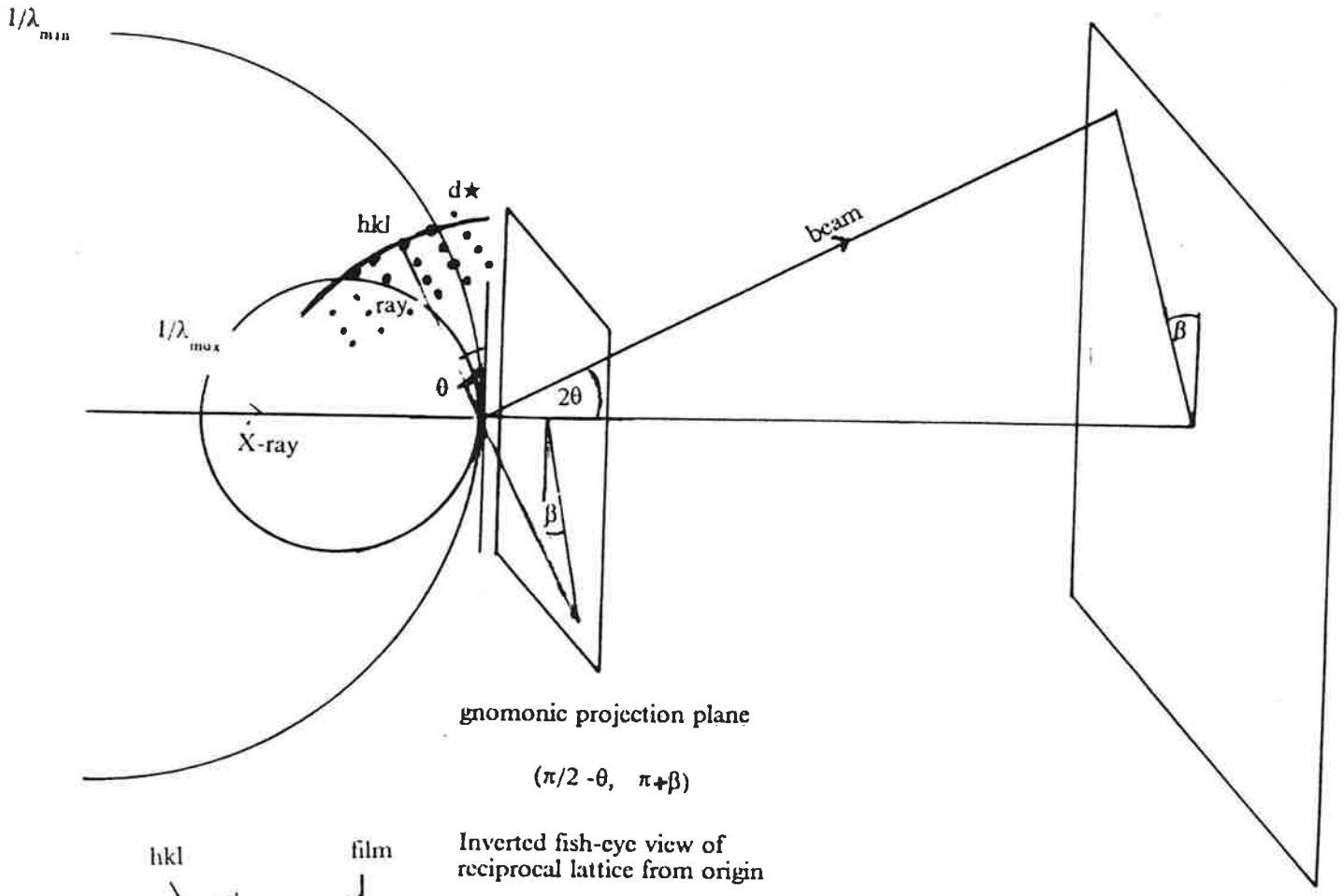
References

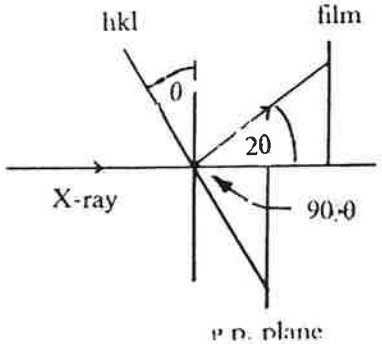
Helliwell JR, Habash J, Cruickshank DWJ, Harding MM, Greenhough TJ, Campbell JW, Clifton IJ, Elder M, Machin PA, Papiz MZ and Zurek S (1989) J Appl Cryst 22, 483 - 497.

Cruickshank, D.W.J, Carr, P.D. and Harding, M.M. (1990) In prep.

Figure 1

Laue Geometry showing Gnomonic Projection

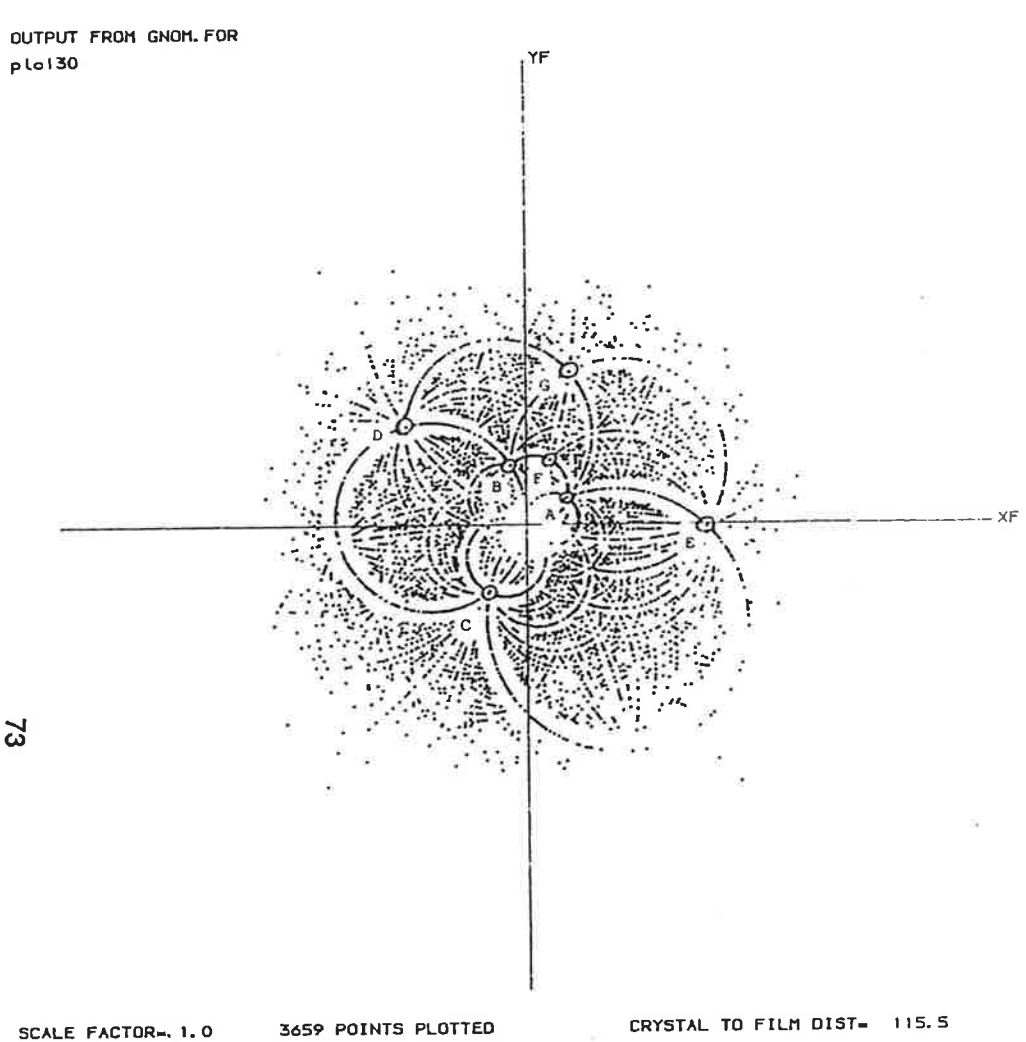


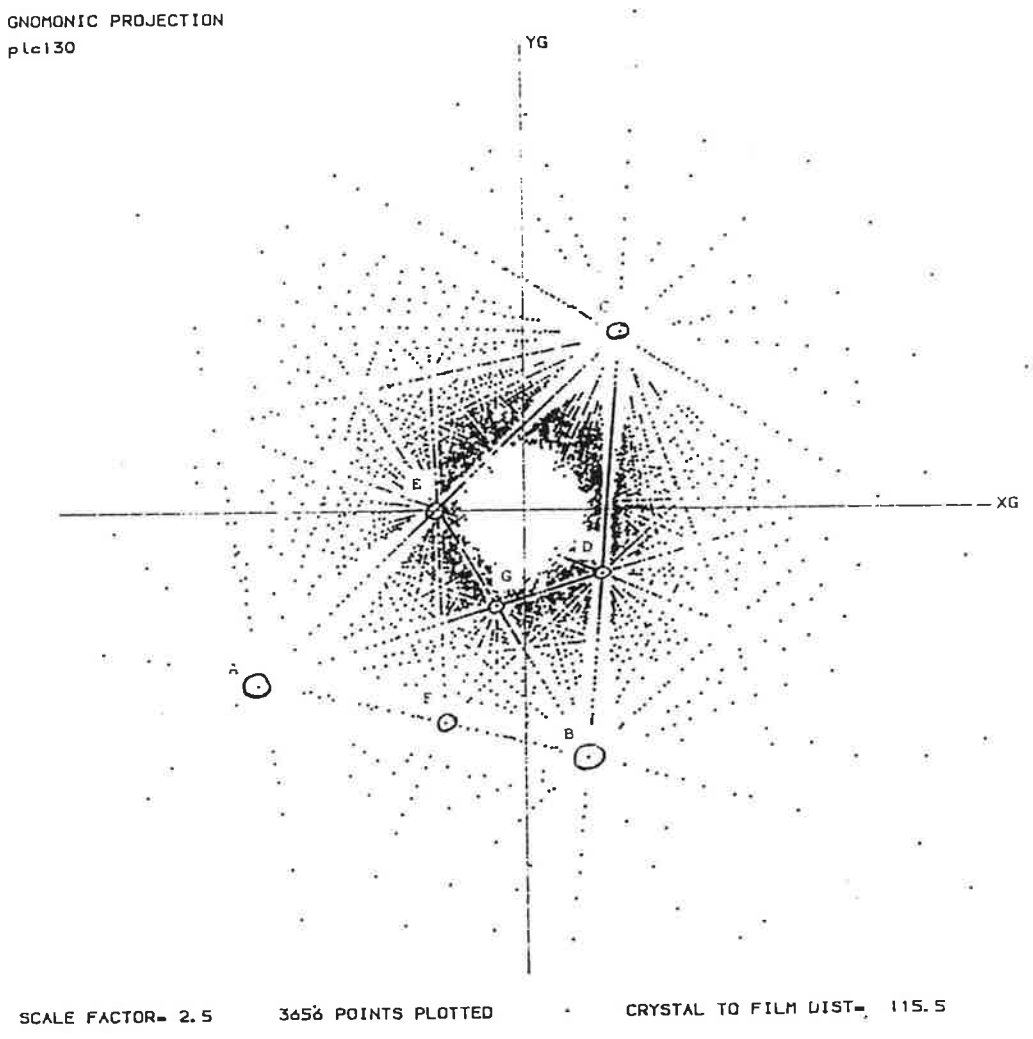


E.D. plane

Figure 2

Film Image





Experiences in introducing the Ardent Titan supermini in a crystallographic/macromolecular modelling site.

E.E. Eliopoulos
Astbury Dept. of Biophysics, Leeds University,
Leeds LS2 9JT, England.

1. Introduction.

The TITAN is a parallel/vector processor supermini computer integrated with high performance graphics. Although faster (P3) processors have recently become available we are using the original P1 processors. Each P1 processor combines an integer RISC processor running at 16MHz with a peak performance of 16 MIPS and a 16 MFlop 64bit vector unit with 32 Kb cache. It can incorporate up to 4 Memory Array Boards (8-32Mb each) with a bus bandwidth of 256Mb/s. The enhanced version of the graphics subsystem includes two boards (standard and expansion) offering 1280x1024x24 Image memory, 16 bit Z-buffering, up to 52 colour planes (24 double buffered + 4 overlay planes with the expansion board) and dynamic stereo option. It offers a hardware draw rate of 600k 3D vectors/s and an animation rate of 10k Gouraud-shaded triangles per frame at 15 frames per second. The operating system is Unix, compatible with V.3 AT&T UNIX and Berkeley 4.3 UNIX. X-windows is supported and provides the basis for other graphics software like Dore. System V file system is supported together with Ardent's fast filing system (AFFS) offering 1Mb/sec peak performance with disk striping (reading from two disks simultaneously). Both SCSI (as standard) and VME (with extra VME interface) devices are supported. Support for Local Area Network includes yellow cable and RG-58 cable Ethernet with a Network Filing system (NFS) and TCP/IP running simultaneously. Our configuration is 2xP1 processors, 32Mbytes of Memory, 2 380Mbyte SCSI disks, graphics (standard + expansion) boards, 120Hz (1280x1024) Sony colour monitor with dynamic stereo option (by Tektronix).

2. Performance Tests.

As usual there is a difference between the quoted peak performances (by the manufacturer) and how the machine performs in its working environment. *Table 1* below indicates performances on standard benchmarks in comparison with some other popular computers.

Table 1

All benchmarks have been performed during working hours in our laboratory with machines under a normal load.

Benchtest	VAX 750	µVAX 3600	TITAN
Dhrystones/sec	1674	8010	24437 (1xP1)
Whetstones (MIPS)	0.52	1.93	7.05 (1xP1)
Linpack (MFlops)	--	--	6.068
Disk Access Kbytes/s	--	--	841
FTP rate Kbytes/s	--	--	190

For further detailed benchmarks see reference (1).

3. Applications.

The major points of interest in introducing a new computer into an established scientific environment are how easy is to port applications from other computers (e.g. VAX/VMS), whether applications run faster and turn round is shorter, whether novel applications can be implemented with ease to make use of the extra power and the availability of established applications for this computer.

3.1 Portability.

As far as Unix non-graphics applications are concerned, apart from ome system specific routines (eg. timing, dynamic memory allocation), portability is straightforward for both high (Fortran and C) and low level languages. In addition both the FORTRAN and the C compilers cope very well with VAX/VMS FORTRAN and C implementations making the task of transferring applications easy. The problems usually start with parallelisation (less with vectorisation) of ported serial code. In some cases (e.g. subroutines with non-standard transfer of argument types) automatic vectorisation (which works well with real arrays) may fail, with no indications of the problem. Parallelisation of serial ported code (i.e code with many subroutine or function calls or frequent i/o) might seem to go through the compiler; but not only is there no gain in performance, there might actually be a loss since multiple processors are tied up by the same application which actually runs serially. Our general conclusions in porting non graphics code to the TITAN are that there is a 3x gain over the VAX 3600 in CPU time in most applications without vectorisation and parallelisation. For calculation intensive applications with matrix and repetitive operations there is a considerable gain using automatic vectorisation which can be increased by adjusting some of the code manually. Parallelisation can be very useful for custom made applications or when firmware is used (e.g. provided libraries, Dore graphics) but the compilers will not perform the *miracle* of changing a serial to a parallel application. Documentation/examples on parallelisation are still under development and debugging tools are no better than those offered by any Unix system. Non-graphics applications will run from remote nodes, under batch or from X-windows.
Graphics firmware on the Titan consists of an X (or PEX) base from which Dore (Ardent's own graphics firmware) or PHIGS (+) run. Since both are very different from the previous static/plotter oriented line graphics, portability is non existent. Those familiar with X and C will have no problem writing the basic functions (manipulating windows, checking interrupts, or creating menus) but there is still no facility that will do this automatically. For those with no such knowledge, they have to rely on Ardent's provided examples for various applications, the demo programs and the DUI (Dore User Interface) code. In fact the most time-consuming part of writing a program that will display a macromolecule in Richardson style is actually **opening the window** and **connecting the dials**. Dore offers a dynamic rendering environment which is scene/object oriented. Objects are constructed from geometric primitives and added to the scene using appearance and geometric attributes. Various studio objects (cameras, lights) can be readily included in the scene. Once

defined, the common Dore database is used for different styles of rendering (e.g. wireframe, spacefill, shadows, reflections, ray tracing) which can be done interactively and can be dynamic (real time) or not (ray tracing in batch). So the effort of programming lies more in decribing the object rather than how to represent it. The price to pay for this convenience is memory and processor usage. Since Dore is optimised to use the vector and parallel options of the compilers, processor usage is not a major problem; but if only limited memory is available then the relatively slow swapping between memory and disk can be a severe limitation. *Figure 1* illustrates the use of the commercial macromodelling software BIOGRAF (V2.6) to calculate a wire model and then a space filling model for 1400 atoms with 32 Mbytes of memory available. Clearly the application would benefit from additional memory. A Fortran interface is provided for Dore but since X handling is done in C , and nearly all examples are written in C, programming in FORTRAN requires considerable dedication and effort. An example of the power of Dore is given by the demo software and the more specialised **Chemviewer**, a collection of applications from different groups concerned with molecular calculations and graphics applications. Some debugging tools are available for Dore (if you can find them) but unfortunately there is no documentation for them. The best source of information is the courses that Ardent are running on Dore and programming optimisation.

3.2. Performance.

In Table 2 comparisons are made between the Titan and the computers on which the applications were originally run. In most cases manual intervention was minimal for vectorisation. The STELLAR vector/parallel version of XPLOR was implemented with minor modifications and further manual optimisation is underway. Implementations of the molecular mechanics, dynamics and crystallographic refinement programs EMPMDS, EMPTOR, EMPXTL, EMPNMA and SIMMIN (I. Haneef) are described in references (5,6). Times for the serial highly optimised SHELX86 program and the serial character manipulating program CLUSTAL show that there is not always a gain in performance in automatically vectorising/parallelising the programs. In the case of the molecular mechanics package ENERGY where i/o statements are used within the time-consuming energy function calculating loops, automatic parallelisation offers marginal gains.

3.3 Availability of software.

Computational chemistry software listed in a recent Ardent directory includes AMBER, AMPAC, BIOGRAF and POLYGRAF, DSPACE, FTNMR and NMR1 and 2. Crystallographic modelling and ab initio calculation packages are expected shortly and no doubt the recent merger of Ardent and Stellar will enhance the availability of software for both graphics and non graphics applications. We have mainly been using BIOGRAF (8), a molecular simulation program that integrates interactive (stereo) graphical display and analysis with molecular mechanics and dynamics. Although memory intensive, it offers an easy to use environment for model building, structure comparison and presentation. So far it has no options for map fitting and direct linking of user applications is cumbersome. Development of graphics software in Leeds is so

far directed to molecular representation using the Dore and DUI firmware. Other molecular modelling sites (eg Scripps, UCSF) have developed very elegant graphics applications using Dore.
Mathematical Libraries available in vector/parallel form include NAG, MATLAB (bundled including Dore representation), FFT routines, GLIM, Mathematica and Math Advantage. CAD software include ABAQUS, ANSYS, DYNA3D, MARC, MOVIE.BYU etc.

4. Projects underway.

The major use of the machine is for interactive macromolecular modelling with examples from small enzyme inhibitors to large protein-protein and protein-nucleic acid complexes. The non-graphics CPU intensive applications include sequence database searches and alignments, secondary structure predictions, molecular dynamics simulations and crystallographic refinements and more recently, tumour growth simulations.

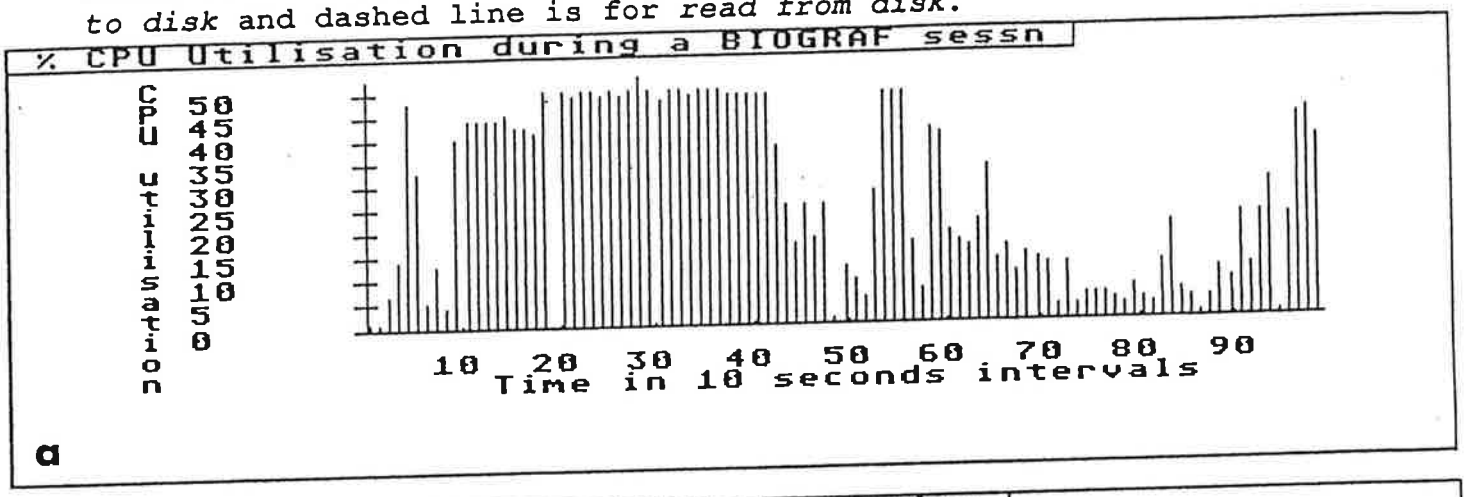
5. A personal view.

From the hardware point of view the Titan is an extremely reliable machine. So far in the first sine delivery we have only had one minimal hardware failure which was rectified the same day. Support has been very good from Ardent U.K. and I hope it will continue to be so, now that support is based in Holland. As far as firmware is concerned, the UNIX implementation is satisfactory but help facilities, while adequate are not very good.
X+ implementation in version 2.2 is still basic and it lacks the useful development tools that are becoming popular in UNIX-based graphics workstations. Both C and Fortran compilers are fast, with tools to help in manual vectorisation and parallelisation but only the standard debugging facilities. The power of the machine is evident in both graphics and non-graphics applications and as long as memory requirements are not limiting, several users can work happily through terminal ports, ethernet and the graphics console without the load affecting performance too much. Communications in the local area network have been very reliable easing the problem of limited available disk storage. However it is the outstanding graphics performance that probably make the Titan stand out. The dynamic stereo implementation through X+ is flicker free and although while in stereo mode the resolution drops by half (512x1280) it is still very useful especially for complicated structure manipulations. Once the first barriers with X+ applications are over, Dore is not difficult to program and indeed very powerful.

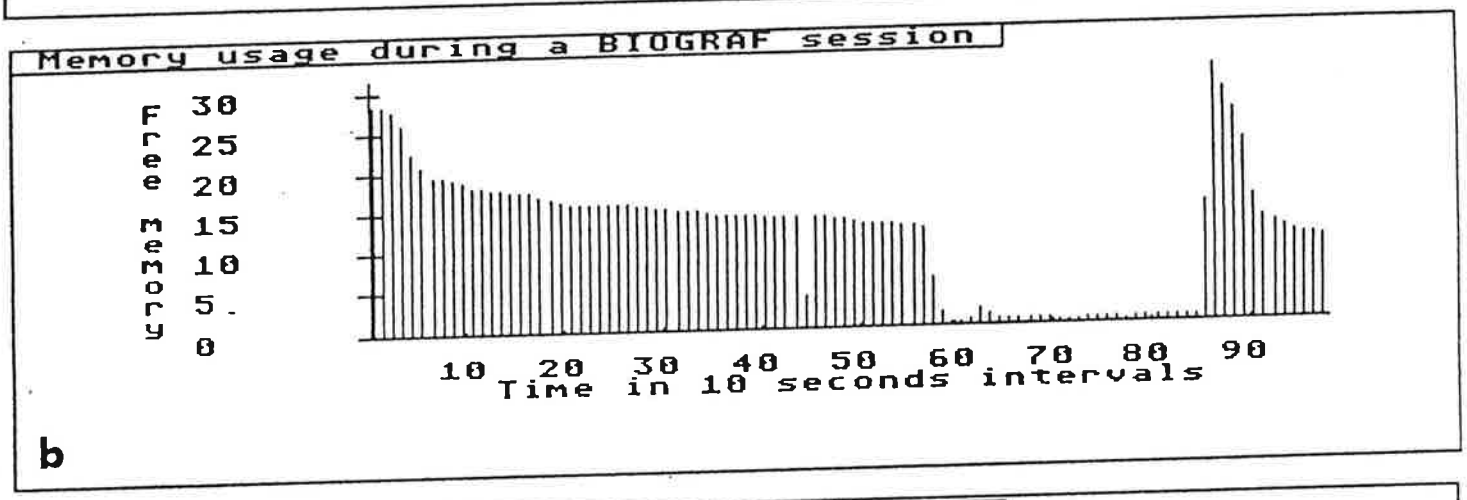
Acknowledgements. I would like to thank Dr. I. Haneef and Mr. D. Parry-Smith for their benchmarks and applications.

Figure 1

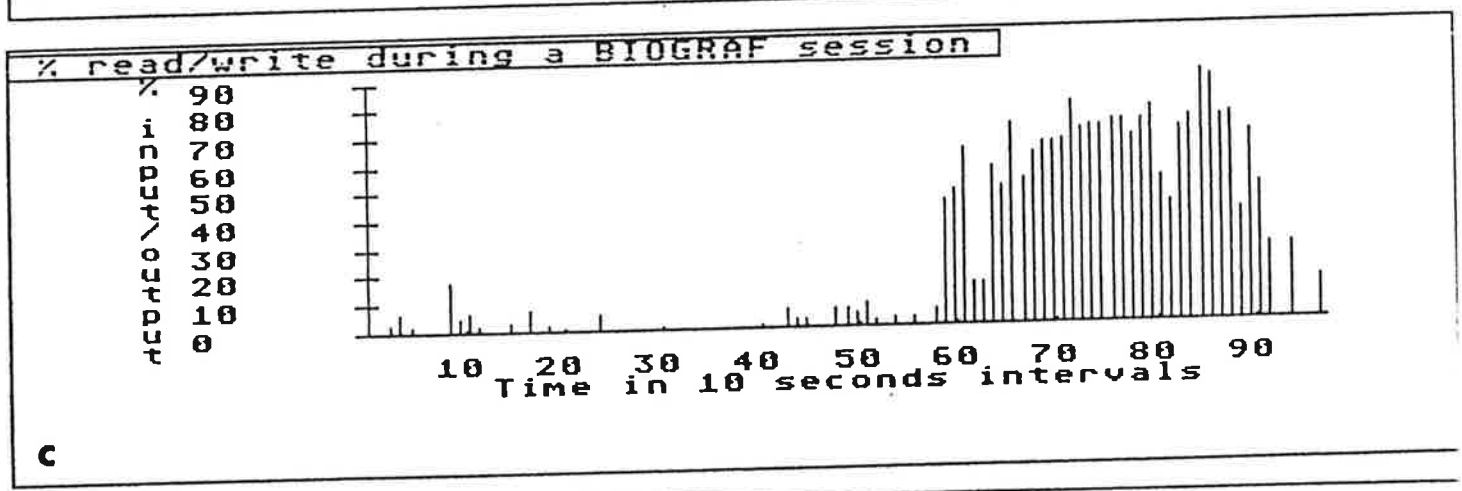
CPU, Memory and i/o use is shown during a 20 minute BIOGRAF session (in 10 second steps) . From login in (t=0), X-windows (t=5), start BIOGRAF (t=10), read, calculate and display and manipulate 6000 atom wire models (t=20), calculate 1400 atom space filling model (t=53), display space filling model (t=58), reset to wire models (t=88). In figure 1d solid line is for *write to disk* and dashed line is for *read from disk*.



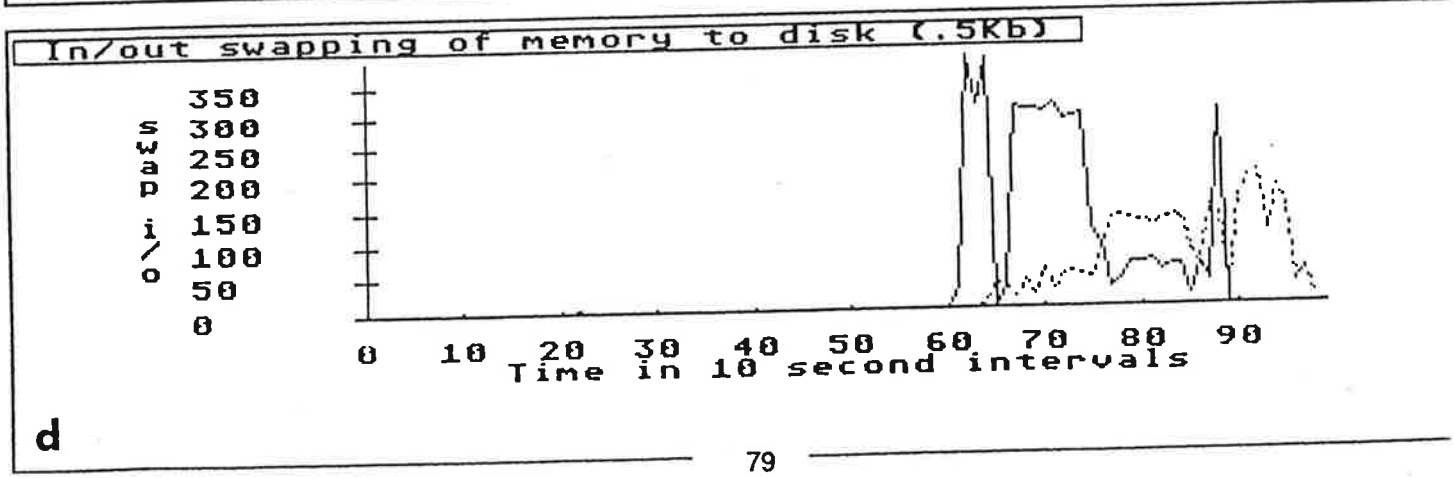
a



b



c



d

Table 2

Time comparisons for ported applications between the Titan and other computers. All CPU timings in seconds unless otherwise indicated. O1 is serial optimisation for the Titan compilers, O2 is O1 + vectorisation, and O3 is O2 + parallelisation.

Application	µVAX 3600	TITAN	Amdahl V8	CRAY-XMP
CLUSTAL (2) mult. alnm. of 21 seqs (170aa)	97	35 (O1) 86 (O3)	--	--
SWIMM (3) 69 seqs dbase scan with 9aa search motif	73	71 (O1) 47 (O2) 50 (O3)	--	--
ENERGY 2x20 atom targeting, 5 s.p. energy minimis.	53	25 (O1) 22 (O3)	7	--
EMPMDS (5) energy minim of APP 368 atoms	5842	1295 (O2)	--	59
Long MD simulation 459 atoms	--	108000 t.a. 1 day		7500 t.a. 1week
SHELX86 (4) 214 atom 4 cycl anisotropic ref	--	836 (O1) 715 (O3)	127	--
XPLOR (7) 2482 atoms 4ps 4000steps	--	216350 t.a. 2.5 d		14240 t.a. 6.2 d

t.a. indicates turn-around time.

References.

1. How the Transputer Stacks up to Other Processors. A Comparison of Performance on Several Application Programs. Stiles G.S. SERC/DTI Transputer Initiative Nov.1989.

2. CLUSTAL Higgins D.G., Sharp P.M. Gene, **73**, 1988, 237.

3. A Protein Sequence/Structure Database. Akrigg D. et al., Nature, **335**, 745.

4. SHELX86 Sheldrick, G. M. 1976.

5. Proceedings of the Workshop on Parallel Computing, Daresbury Lab.,Nov 88.

6. SIMMIN, by I. Haneef,J. Compt. Chem. 1990.

7. XPLOR Brunger, A.T. J.Mol. Biol.,**203**, 1988, 803

8. BIOGRAF BioDesign Inc. (1987)

Report of the Working Group on

MACROMOLECULAR CRYSTALLOGRAPHY AT THE ESRF

ESRF Users Meeting
held 20-22 March 1989
Grenoble, FRANCE

Convenor: Professor J.R. Helliwell

---

1. Introduction

The Working Group discussed the new science that could be done at the ESRF, its implementation and the general requirements to ensure a successful research programme.

A number of key areas of innovative scientific work were identified. The technology for these topics is already proven, but the exploitation either depends absolutely on the ESRF, or in some of the cases would be only marginally possible without the ESRF. The topics are:

- time-resolved studies of enzyme reaction and viral states in the crystal;
- structure determination of extremely large molecular assemblies such as viruses or other multimacromolecular complexes;
- smaller crystals would be brought in range for data collection;
- the solution of the phase problem by multiple wavelength anomalous dispersion techniques (so-called MAD method).

Technically both the Laue and monochromatic geometries would be exploited for quantitative structure analysis.

Policy issues that were considered relevant were:

- itemising new science
- extrapolations from existing science
- likelihood of success
- identifying dependent technologies (e.g. optics, detectors, biotechnology)
- the amount of usage of a given instrument
- the continued need for the national facilities.

Existing discussion papers, e.g. the relevant section of the Red Book, had been circulated to participants within the Working Group about two

months before the meeting. The newsletter of the European Association of the Crystallography of Molecular Biology (EACMB), February 1989 issue, carried a copy of the Red Book chapter on macromolecular crystallography and a cover note updating various topics, results and achievements. This newsletter is distributed to essentially all macromolecular crystallography laboratories in Europe.

2. Beam Line Requirements

The following beamline recommendations were made by the Working Group:

1. Multipole wiggler (for Laue and monochromatic work)
   100% usage expected.
2. Undulator (for monochromatic and quasi-monochromatic work)
   50% usage expected.
3. Bending magnet (for rapidly tunable monochromatic)
   50% usage expected.

The order given is the order of priority. However, the undulator should be seen to be of nearly equal priority to the multipole wiggler. In the Red Book the undulator was the highest priority. The change in priorities reflects the rapid and dramatic development of the Laue method for quantitative structure analysis in the last 2-3 years.

2.1 Multipole wiggler beam line

This is to provide the broadest possible wavelength band pass (i.e. $\simeq 0.2 < \lambda < 2$ Å) for Laue diffraction; a large region of reciprocal space will be stimulated, small changes in cell parameters in a time sequence can be tolerated and the sample is held stationary. It has been shown that the multiplicity distribution is not adversely affected by using a broad band pass of wavelength. The spectrum should be smooth to allow accurate wavelength normalization.

The impact of the ESRF will only be realised if advantage is taken of the low emittance. Point focussing of as broad a wavelength bandpass as possible will be essential (although $0.6 < \lambda < 2$ Å is a compromise

wavelength range to be considered). The diameter of the focus should be ≃ 0.1 to 0.2 mm. Subsidiary sources will have to be collimated out in the focal plane. Time resolutions will as a reuslt be improved into the micro to millisecond regime for an enzyme crystal for example, considerably surpassing what is possible on the national facilities of millisecond to seconds exposure times.

If focussing is not realized then exposure times at the ESRF will not be significantly better than on the national machines. Indeed the latter may well be better because of the closer distance of approach to the tangent point at the shield walls. The unfocussed white multipole wiggler beam also emanates from two separate sets of sources which would be a problem.

The possibility of a single bunch of electrons delivering enough photons for a Laue exposure in ~ 100 picoseconds may well completely avoid radiation damage to the specimen. An idea of the flux required at the sample for this experiment can be arrived at by multiplying the following factors:- (sample scattering efficiency)$^{-1}$ x (number of reflections) x (counts for required precision)/(area of sample). For the protein crystal case, typically 10^4 x 10^5 x 10^4/(0.3 mm)^2, i.e. 10^{13} photons/bunch into an area of 0.1 mm^2. Recent experiments at CHESS had yielded a diffraction pattern corresponding to a limited data set from lysozyme using an undulator on such a time scale (i.e. with such a single bunch). At the ESRF perhaps a tapered undulator should be looked at to try and broaden the band pass compared to the CHESS experiment. The multipole wiggler discussed in this section probably would not deliver enough flux onto the specimen. In the Working Group some arguments were advanced on each side about the possible reduction of radiation damage by single bunch operation. Most members believed that no effects of radiation damage could affect X-ray scattering within a nanosecond.

Rapid Laue photographs could be used for kinetic crystallographic studies, which are also discussed in the next section. The general remarks made there are also applicable to kinetic studies using the wiggler.

Use of a (graphite or multilayer) monochromator arrangement may be

particularly suited for (scanning) Laue applications to structures with broadened crystal mosaicity. The addition of a monochromator to a point focussing mirror system on the multipole wiggler would allow a variety of other experiments to use the instrument. The monochromatic beam would be fully tunable at a high intensity for the study of large unit cells and/or small crystals. Two crucial areas were identified for research and development:

- stable, precise point focussing optics for Laue and monochromatic modes under very high (kWatt/mrad) heat and specific heat loadings;
- area detector development. Charge coupled devices (CCD's) with 2000 x 2000 pixels in an area of 100 x 100 mm^2 would be reasonable for Laue or monochromatic data acquisition. Since the large overall size is unlikely to be readily available, smaller CCD's will be used in conjunction with photographic film and the image plate.

The European Molecular Biology Laboratory has expressed an interest in developing this multipole wiggler beam line.

2.2 Undulator Beam line

This beam line will be optimal for very large unit cell crystallography (e.g. cell edges up to 2000 x 2000 x 2000 $Å^3$). The flux at the specimen per 0.1 mrad will be much higher than on the multipole wiggler beam line discussed above. Currently studied virus crystals (cell edge $\leq$ 400 Å - 1000 Å) have a narrow mosaic spread (e.g. $\leq 0.02°$) so that the very small beam divergence of the undulator would be very beneficial. Focussing (1 : 1) of the beam to $\approx$ 0.1 mm diameter spot at the sample would be a typical need.

Radiation damage to virus crystals is reduced considerably at shorter X-ray wavelengths. For example on the SRS wiggler protein crystallography workstation use of a 0.9 Å wavelength beam leads to 3-4 times more data per sample than use of 1.5 Å beams elsewhere. The wavelength of 0.9 Å is selected to enhance the photographic film sensitivity using the bromine K absorption edge of the silver bromide grains. By the time the ESRF comes on line the image plate is likely to have replaced film as detector in these studies. Hence, a

wavelength for the undulator fundamental of 1.0 Å would be much easier to realize than 0.9 Å and yet minimise radiation damage almost as effectively.

Small protein crystals (10 μm or less) could be studied extremely well on this instrument. Focal spots in this smaller size range would be beneficial. Freezing of the sample to prolong its lifetime in the beam will be obligatory. Diffuse scattering from macromolecular crystals would use this station. Use of a tiny pinhole (10 μm ?) before the sample would essentially prevent the Bragg spots from obscuring the diffuse pattern. If not adjustable, a range of pinhole sizes will be required. Accurate preset alignment will be crucial, together with a means for accurately positioning the sample to better than 5 microns.

Kinetic crystallography studies using a beam of wide spectral width ($\delta\lambda/\lambda = 0.1$) would create a facility to study order-disorder states in a rapid sequence on the millisecond time-scale. Such states cause considerable streaking in broad band pass Laue patterns. All crystal kinetic experiments require special techniques to initiate the process without delays due to diffusion within the crystal. "Caged" chemical reagents whose reactivity is initiated by a u.v. or X-ray flash are available, but will require more development.

The overall usage of the instrument would depend critically on the number of very large unit cell (virus) projects coming forward. It is very difficult to predict how many new virus crystallography groups may be established in Europe in the next 6-7 years. However, significant usage for preliminary studies of tiny crystals can be foreseen. The usage for kinetic studies is hard to predict initially, but if successful would grow rapidly.

We anticipate that the initial demand for this instrument might be about 50% of the available time. As a result this beam line is considered to be lower priority than the multipole wiggler instrument.

Research and development areas are:
- stable, precise, point focussing optics for monochromatic and quasi-monochromatic (narrow band pass Laue) modes;

- charge coupled device area detector development (2000 x 2000 pixels in an area of 100 x 100 mm^2). Image plates will still also be needed.

2.3 Bending Magnet

An ESRF bending magnet surpasses any of the national facilities bending or wiggler magnet light sources. It enjoys a low emittance and will be relatively simple in its exploitation for rapidly tunable/multi-wavelength anomalous dispersion (MAD) work (compared with a tunable undulator or a monochromatised multipole wiggler). The establishment of a diffractometer for MAD work on an ESRF bending magnet will act as a European focus for this important technique for solving those structures that are not tractable by multiple isomorphous or molecular replacement methods. The technical needs for the instrument are well known (1 eV bandpass, 0.9 Å $< \lambda <$ 1.8 Å tunable monochromator range, rapid tuning facility, etc.).

We anticipate 50% usage of this beam line.

3. General Points

In order to extract the best science from the ESRF in our scientific area a centre/laboratory at the ESRF site should be established for synchrotron radiation related biological research. This would allow

- purification of molecules, genetic engineering, laboratory for handling pathogens (of live virus) etc., and
- the development of techniques for stimulating biological reactions in crystals and populating intermediate states for time-resolved studies.

Software development should be encouraged and stimulated whenever possible. The following were identified for attention:

- rapid analysis of results on the beam line involving all stages from data processing (especially Laue) through to calculation of structure. This would guide the sequence of time resolved work

actually undertaken in a sequence of particular experiments.

- software to enhance multi-λ phasing techniques.

4. Instrument Design and Development

It is proposed that a European project team or working group is set up for these instruments with urgency. The purpose of this group is to coordinate expertise from the participating countries in the area of macromolecular crystallography. The project team would work under the auspices of ESRF and perhaps the European Association for the Crystallography of Molecular Biology. The working group would detail the design and identify groups which could develop or supervise the construction of the instrument. This would include the insertion devices, optics, sample environment, goniometry, detectors, computers and software.

5. Detector Development

Resources for detector development need to be considerable if the investment in the machine itself is to be properly realised. For protein crystallography on the multipole wiggler and undulator instruments, it has been identified that charge coupled devices (CCD's) and image plates are likely to be of most use. On the bending magnet station where the most accurate measurements are needed, photon counting will be required.

For kinetic crystallography there exists probably the most demanding situation. Although in the initial stages of ESRF a few time slices in such studies would be very exciting, there will quickly be a need for many more time frames. Hence, one can imagine, for example, that a CCD with 2000 x 2000 pixels (in area) and 12-16 bits per pixel dynamic range would need to deliver 100 or even 1000 time frames each of 1 msec width every 1 msec. This is preumably an impossible data transfer rate specification. Clearly however, performance approaching this will be of interest to the user.

On the undulator, a CCD device should be harnessed for the virus data collection requirement. The huge number of simultaneously diffracting

reflections places stringent demands on the detector aperture and number of pixels. Although 2000 x 2000 pixels over an area of 100 x 100mm^2 or 50 x 50 mm^2 would initially be useful a need for 10000 x 10000 pixels over continuous areas five times as large would be a requirement.

The limitations of current image plate and associated scanner technology fall into two categories. Firstly, spatial resolution could be enhanced, at the expense of some sensitivity, by making a thinner active layer; however, these plates are available only with a preset specification from the manufacturer. Secondly, the largest sized plate available is 40 x 20 cm^2; with an off-line scanner a larger detection area can be realised by tiling several plates together but this scheme is not possible with the on-line scanner arrangement; hence a more modest development project is the utilisation of several plates and an off-line scanner with software tailored to deal with the tiling arrangement. The large detector area so realised would be invaluable for the virus data collection, probably as a more viable alternative to the 10000 x 10000 pixel CCD detector mentioned above.

oOo

SINGLE BUNCH EXPOSURES IN LAUE MODE

J. Hajdu

Need:
- 10^{13} photons at the sample
- 0.2 – 2.0 Å λ-range (0.4)
- 0.2 mm beam diameter (collimation)

Get: no radiation damage, ultimate time resolution.

Details:

1. An average macromolecular Laue pattern consists of ~ 10^5 reflections. Need about 10^4 photons/reflection for good statistics.
   Every 10^4th photon is diffracted.
   This means 10^{13} Photons/Laue Image.

2. An average protein crystal has ~ 10^{16} unit cells. A small proportion of all unit cells will see the beam (probably true).

Consequences:

1. If the exposure time (bunch length) is ~ 100 ps, then no (or practically no) radiation damage will be observed in the image since there will be no time for the crystal to collapse for the chemical and transport processes to take place.

2. "Ultimate" time resolution.

EUROPEAN SYNCHROTRON RADIATION FACILITY
B.P. 220, 38043 GRENOBLE CEDEX, FRANCE

E S R F U S E R S' M E E T I N G 20–22 March 1989

LIST OF PARTICIPANTS BY GROUP

Group			SAC Responsible	ESRF Responsible
1. Diffraction				
Group 1	(Macromolecules)			
J. Helliwell	Convenor	(UK)	A. Liljas	A. Miller
[redacted]		[redacted]		
H. Bartunik		(D)		
D. Blow		(UK)		
[redacted]		[redacted]		
R. Fourme		(F)		
J. Hajdu		(UK)		
A. Hordvik		(Nor)		
[redacted]		[redacted]		
H. Iwasaki		(Japan)		
J. Jansonius		(CH)		
A. Kvick		(USA)		
D. Stuart		(UK)		
J.-C. Thierry		(F)		
K. Wilson		(EMBL)		
Group 2	(Materials Science Crystallography)			
M. Marezio	Convenor	(F)	G. Heger	A. Miller
Ch. Bärlocher		(CH)		
P. Becker		(F)		
P.J. Brown		(ILL)		
A. Catlow		(UK)		
A. Cheetham		(UK)		
A.N. Christensen		(Den)		
P. Coppens		(USA)		
R. Fita Rodriguez		(E)		
A.N. Fitch		(UK)		
H. Fjellväg		(Nor)		
H. Fuess		(D)		
M. Lehmann		(ILL)		
F. Mo		(Nor)		
I. Olovsson		(Swe)		
J.-P. Weber		(CH)		

Advisory group on synchrotron radiation

1. At the EACBM's Como workshop, on May 18 1989 an advisory group on synchrotron radiation was formed, consisting of

D. Blow (London, presently in Grenoble)
R. Fourme (Orsay)
J. Helliwell (Manchester)
D. Stuart (Oxford)
J-C. Thierry (Strasbourg)
K. Wilson (Hamburg).

For the time being, D. Blow was appointed its spokesman.

2. It is intended to add a member from Italy as soon as possible.

3. The group is informing the ESRF that it is available to consult with them on relevant issues. As a first step they wish to consider with the ESRF plans for detectors on the biology beamlines.

4. The group is informing EACBM participants through the newsletter of its existence, and of its willingness to receive information and opinions relevant to the exploitation of synchrotron radiation.

5. It is hoped the group will meet again at the Laue workshop and the MAD workshop.

IUCr COMMISSION ON BIOLOGICAL MACROMOLECULES

Policy on Publication and the Deposition of Data from Crystallographic Studies of Biological Macromolecules

I. Preamble

1. Crystallographic analyses of protein, nucleic-acid and virus structures produce an extraordinary amount of information, and these results are widely recognised as having unique scientific value. Available information transcends that which can be recorded in usual scientific publications, and the Protein Data Bank is often used as a supplementary repository for such results. As in all science, it is imperative that sufficient information be made available so that the structural results can be reproduced and verified.

2. The importance of preserving the fundamental data and results from diffraction studies is recognised alike by producers and users of this information. There are, however, concerns that results from the early stages of analysis will be inaccurate in detail and that investigators should have the opportunity to complete the analysis and interpretation of their data. On the other hand, an open-ended protection of authors' interests conflicts with the general scientific good and it creates the risk that valuable data will be lost forever. Accordingly, the deposition policy promulgated below stipulates immediate deposition of atomic coordinates and diffraction data supporting publications on structure, but it provides for the possibility of a specified delay in the release of this information for public use.

II. Policy

1. The Commission on Biological Macromolecules of the International Union of Crystallography endorses a deposition policy for crystallographic studies to permit independent verification of the results and to preserve the primary data for future use. Scientific publications reporting results from crystallographic determinations of macromolecular structure should be accompanied by a

deposition of atomic coordinates and structure factor information at a level appropriate to the description given in the paper.

Specific provisions of the policy are elaborated below:

a. Provisions for Atomic Coordinates. Two different levels of description arise with respect to the coordinates of macromolecular structures. In the case of chain-tracing descriptions, the alpha-carbon coordinates for proteins or phosphorous positions for nucleic acids are appropriate for deposition. If the interpretation presented depends on atomic details as shown in figures of side chains or numbers derived from atomic coordinates, then the full coordinate list should be deposited. Atomic displacement parameters (B-values) and occupancy factors that are part of a model should also be deposited. Investigators might choose to flag regions of a structure that are judged to be particularly unreliable or subject to revision.

b. Provisions for Diffraction Data. Native structure factor magnitudes should be deposited to the limit of Bragg spacings stated in the paper. The deposition of additional data used in phase determination (heavy-atom isomorph data, Bijvoet mates, multiple wavelength measurements, etc) is also encouraged. In the case of structure reports that do not involve atomic models (eg low resolution studies) both structure amplitudes and phases used in Fourier syntheses that are reported should be deposited.

c. Provision for Publications in Methodology. The policy applies to reports on structural results. Those papers that describe purely advances in methodology are exempt from this policy even if diffraction data or structural results were required for their development.

d. Provision for Manner of Deposition. The Protein Data Bank at Brookhaven National Laboratory is recognised by the Commission on Biological Macromolecules of the International Union of Crystallography as the appropriate repository for results from macromolecular crystallography. Accordingly, data should be deposited in machine form as instructed by the Protein Data Bank.

e. Provision for Delayed Release. It is the intention of this policy that the deposition of data associated with a scientific publication should occur concurrently with publication of the article. Nevertheless, provision is allowed for the authors to request a delay in the release of the deposited data. For deposited coordinates this delay is not to exceed one year from the date of publication. For deposited structure factors, the requested delay can be up to four years from the date of publication.

f. Provision for Enforcement. The provisions of this policy require inclusion in the publication of a statement to the effect that "the atomic coordinates and structure factor data described here have been deposited in the Protein Data Bank at Brookhaven".

III. Journals

The Commission recommends that this policy be communicated to all the relevant Scientific Journals and that they be urged to adopt its provisions.

IV. USER OBLIGATIONS

1. The Commission hopes that the practice of depositing coordinates used in structural description will be extended to publications based spectroscopic data (eg nuclear magnetic resonance, EXAFS) and from theoretical and modelling studies.

2. Users of deposited data should cite the primary references, as well as the Protein Data Bank, when making use of the data.

17/2/1989

International Union of Crystallography (IUCr)

Commission on Synchrotron Radiation

The Executive Committee of the IUCr has decided, after detailed consultation with the crystallographic community involved in this field, to set up a Commission on Synchrotron Radiation on an ad interim basis. A General Assembly of the International Congress of Crystallography will be invited formally to establish the Commission, approve its terms of reference and elect its members. I have been asked to form the ad interim Commission, and I now invite comments and further suggestions on its draft terms of reference which are set out below.

The Commission on Synchrotron Radiation shall be concerned with specific applications of synchrotron radiation to various areas of crystallography and diffraction including:

(i) Crystallography of small molecules and large molecules in the areas of very high resolution, large unit cells, microcrystals, reduced radiation damage, kinetic crystallography and multi-wavelength anomalous dispersion phasing.

(ii) Fibre diffraction and small angle scattering including time resolved studies.

(iii) X-ray topography.

(iv) EXAFS (extended x-ray absorption fine structure).

(v) Powder diffraction.

(vi) Diffuse scattering.

(vii) X-ray optics and detectors of particular relevance to the utilisation of the unique properties of synchrotron radiation.

(viii) Magnetic scattering.

(ix) High pressure diffraction.

The Commission will serve several functions as follows:

(i) One of the main functions of the Commission would be to organise sessions and invite speakers at IUCr Congresses. It will initiate and assist in the organisation of other meetings that will catalyse developments and innovations in the subject.

(ii) Cataloguing of information on the available synchrotron radiation sources and the instrumentation relevant to the above topics. This information will be made freely available and so improve the future planning, use and effectiveness of the global resources available for the community of crystallographers and diffractionists represented by the IUCr.

(iii) It will provide a forum for comparing and contrasting the policies of the various centralised synchrotron facilities.

(iv) It will facilitate scientific and technical studies aimed at improving standards of sources, equipment and procedures.

(v) It will try to provide a mechanism whereby potential users of synchrotron radiation from anywhere in the world can be directed to the appropriate facility and helped with gaining access there.

The Commission will be distinct from and complementary to the other Commissions of the IUCr. There will be parallel interests with the Commission on Neutron Diffraction, which deals with centralised neutron facilities, and in instrumentation with the Commission on Crystallographic Apparatus. There will also be

cooperation with the Commission on Crystallographic Teaching in the preparation of a booklet on synchrotron radiation.

Professor J.R. Helliwell,
Department of Chemistry,
University of Manchester,
Manchester M13 9PL,
England.

Tel: 061-275-4686
Fax: 061-275-4598

November 1989

LCCMB Laboratoire de Cristallographie et
URA 232 - CNRS Cristallisation des Macromolécules Biologiques

POSDOCTORAL POSITION IN PROTEIN CRYSTALLOGRAPHY

Applications are invited for a postdoctoral position at CNRS Marseille, France, to undertake Protein Crystallography in the field of Enzyme/ Inhibitor structural studies.

A crystallographer, with a recent PhD and a particular interest in enzyme mechanisms, in computing and molecular graphics techniques would be especially welcome.

The current projects in the Lab., headed by J. Fontecilla-Camps, and covering several aspects of macromolecular structure, include:

** bacterial redox proteins, scorpion neurotoxins, pancreatic and malt alpha -amylases, lipases...

Laboratory facilities include an area detector (Xentronics), two graphics systems (Iris Silicon graphics) and adequate computing facilities (Vaxes).

The post is funded for one or two years and salary will follow the CNRS rules (if PhD, around 14000,00 F / month).

The Marseille area, the entrance of Côte d'Azur, and with the proximity of Aix-en-Provence, offers a wide range of cultural and recreational activities (especially all kinds of sea activities, but also climbing...) and pleasant weather in all seasons.

Interested applicants should send their curriculum vitae and two letters of recommendation to:

HASER Richard, Directeur de recherches au CNRS,

LCCMB-CNRS
Faculté de Médecine Nord,
Bvd. Pierre Dramard,

13326 MARSEILLE Cedex 15
FRANCE

Phone: Lab. 91 65 79 47 FAX 91 65 75 95

Further information can be requested at the above adress or by phone outside lab. hours on 91 75 22 17.

LCCMB-CNRS, Faculté de Médecine - Secteur Nord -, Bvd Pierre Dramard
13326 MARSEILLE CEDEX 15 FRANCE TEL: 91.65.79.47 FAX: 91.65.75.95

Post-doctoral position in protein crystallography at Lund University, Sweden

A postdoctoral position (forskarassistent) for a maximum of four years is available at the department of Molecular Biophysics, Lund University, Sweden. The department which is very new (current staff 8 persons) has been established at the Chemical Centre of the university. Good grants have been obtained and the speciallized equipment will by beginning of next year include a Rigaku rotating anod, an area detector, an oscillation camera, VAX computers and work stations and an Evans & Sutherland PS390. In addition the department is well equipped for bacterial growth as well as biochemical work.

The current projects include proteins in the protein biosynthesis machinery (ribosomal proteins and factor proteins). New isoenzymes and site directed mutants of carbonic anhydrases are analyzed in order to explore the enzyme mechanism as well as folding and stability properties of the enzyme. Several bacterial peptidases are analyzed in order to explore the possibility that yet other structures and active site constructions can be utilized for peptide bond breakage. Several exciting new projects have also started to emerge in collaboration with local biochemists. The postdoctoral coworker is expected to participate in one or several of these projects.

The applicant should have well documented experience in several aspects of protein crystallography. The applicant should also have received a Ph D degree (or corresponding) no more than five years ago. The salary level starts at approximately 13,000 Sw. Cr. per month. Applicants should get in contact with Anders Liljas for instructions how to make the formal application to the University.

Prof. Anders Liljas
Department of Molecular Biophysics
Chemical Center
Lund University
Box 124
S-221 00 Lund
Sweden

Telephone: 46-46-104681 FAX: 46-46-104543 Telex: 33533
Bitnet: MBFYS SELDC52

UNIVERSITY OF HOUSTON

DEPARTMENT OF BIOCHEMISTRY AND BIOPHYSICAL SCIENCES

HOUSTON, TEXAS 77204-5500

Kurt L. Krause
Assistant Professor of Biochemistry

Phone: 713.749.7590
FAX: 713.749.3239

November 14, 1989

Dr. Kim Henrick
Daresbury Laboratory
Daresbury
Warrington
WA4 4AD
England

Dear Dr. Henrick:

The University of Houston is continuing to expand its program in structural biochemistry. This fall I joined the Department of Biochemistry and I am in the process of setting up a crystallography laboratory which will include an area detector/rotating anode combination along with a high speed diffractometer. The University is interested in recruiting another faculty member to share this equipment and I would appreciate you bringing this position to the attention of any suitable candidates. The position is being advertised as follows:

> X-ray Crystallographer - Tenure-track assistant professor to begin Fall 1990. Candidate should have a strong record of research in the determination and functional interpretation of biomolecular structure. Position is associated with the Institute of Molecular Design, which offers access to modern area detector and graphics supercomputing facilities as well as extensive opportunities for collaboration. Send CV, research plans, and three letters of recommendation to Professor J. Andrew McCammon, Department of Chemistry, University of Houston, Houston, Texas 77204-5641.

Thanks and best regards.

Sincerely yours,

Kurt L. Krause

POSTDOCTORAL RESEARCH FELLOWSHIP

IMPERIAL CANCER RESEARCH FUND Protein Strucrure Laboratory LONDON U.K.

Applications are invited for a postdoctoral research fellowship to work with Dr Paul Freemont in the newly formed Protein Structure Laboratory, in the ICRF central London laboratories, on the determination of protein structure by X-ray crystallography.

The project will involve studying two associated cytosolic calcium binding proteins (CaBP), p8 and p14 which are expressed by circulating monocytes and neutrophils and are members of a larger family of CaBP's named the S100 family, that may be involved in intracellular signal transduction. In collaboration with Dr. Nancy Hogg's laboratory at the ICRF we have recently developed a large scale purifications procedure for the proteins which has allowed us to characterise the association as being a heterodimeric complex of 1:1 stoichiometry. We have also shown that p14 is phosphorylated in a calcium dependent manner and have isolated the phosphorylation site to the C-terminal region of the molecule. Initial crystallisation trials have yielded small crystals which are not yet suitable for high resoltuion X-ray analysis. The main aims of the proposed project are to obtain larger crystals of the complex and to initiate X-ray crystallographic studies on crystals of the complex.

The Imperial Cancer Research Fund is one of the largest independent cancer research institutes in Europe employing approximately 400 scientists and clinicians.The Protein Structure Laboratory is a newly formed laboratory at the ICRF and has a number of collaborations with other ICRF scientists. We have extensive biochemistry facilities for protein purification and crystallisation and are in the process of purchasing a Rigaku RU200 rotating anode generator. Future developments will include the purchase of an area detector system. For data collection we currently use the facilities at Birkbeck college by agreement with Prof. T.L. Blundell which includes a FAST area detector system. The ICRF has extensive computational facilities including a VAX 8700 cluster, a DAP array processor and an IRIS 4D turbo molecular graphics system for model building and electron density map interpretation.

Suitable candidates should have a good honours degree and are expected to have completed their PhD in some aspect of (protein) crystallography before taking up the post, which is for 3 years starting in February 1990.

Applicants should send a curriculum vitae, and the names and addresses of two referees to Dr P.S. Freemont from whom further details on the position can be obtained, at the Protein Structure Laboratory, Imperial Cancer Research Fund, 44 Lincolns Inn Fields, London WC2A 3PX U.K.

Telephone: 01-269-3291; JANET/BITNET: p_freemont@icrf.ac.uk

POSTDOCTORAL RESEARCH FELLOWSHIP

A senior postdoctoral position is available, in virus crystallography, for work on SV40, polyomavirus, and rotavirus structures. Experience with synchrotron data collection, non-crystallographic-symmetry phase refinement, and other methods of very-large-unit-cell crystallography highly desirable. Send CV with names of two or three referees to

Prof. Stephen C. Harrison
Howard Hughes Medical Institute
Harvard University
7 Divinity Ave.
Cambridge MA 02138

FACULTY POSITION IN X-RAY CRYSTALLOGRAPHY

MRC GROUP IN PROTEIN STRUCTURE AND FUNCTION

University of Alberta, Edmonton

The Medical Research Council of Canada Group in Protein Structure and Function at the University of Alberta invites applications for a Faculty position immediately available in the Department of Biochemistry. Applicants should have a strong background in protein/macromolecular crystallography and will be expected to develop a research project that will complement and contribute to the overall research efforts in the group. Appointments can be made at any level from Assistant to Full Professor depending upon the qualifications.

The MRC Group in Protein Structure and Function was established in 1974 and has undergone two successful renewals. In addition to X-ray crystallography, the major techniques of physical biochemistry are included in the armament used by the group members for the study of protein systems: hydrodynamic and spectroscopic methods, 2D nuclear magnetic resonance spectroscopy, protein micro-sequencing techniques, solid phase synthesis of peptides, recombinant DNA and protein engineering. The X-ray crystallographic laboratories are supported by excellent facilities, two single crystal diffractometers, two rotating anode generators, an SDMD system area detector with two detectors, two molecular graphics systems. The computing facilities at The University of Alberta are excellent. The University mainframe is an Amdahl 5870 operating under MTS. Access to supercomputing facilities is also possible at the Universities of Toronto (Cray) and Calgary (Cyber 205).

In accordance with Canada Immigration regulations, priority will be given to Canadian citizens and permanent residents. The deadline for the positi is November 30, 1989. Applicants should forward curriculum vitae, list of publications and names of three referees as soon as possible to either:

Professor C.M. Kay

Co-Director, MRC Group

in Protein Structure and Function

Department of Biochemistry
University of Alberta,
Edmonton, Alberta, Canada T6G 2H7
or
Professor M.N.G. James,
Department of Biochemistry.
University of Alberta
Edmonton, Alberta
Canada T6G 2H7

FACULTY POSITION in MACROMOLECULAR CRYSTALLOGRAPHY

Nature of Position:

A position is available in macromolecular crystallography at assistant to full professor level; entry level applicants are encouraged. Appointment is through the Institute of Molecular Biology in the Departments of Chemistry and Biology. The Multiwire Area X-ray Diffractometer Facility is available for use.

Contact:

Please send resume with research proposal and have three letters of recommendation sent to

R. H. Kretsinger, Chair Search Committee,

Department of Biology,

University of Virginia,

Charlottesville, VA, 22901.

MEETING

THE APPLICATION OF CHARGE-DENSITY RESEARCH TO CHEMISTRY AND DRUG DESIGN

A Nato Advanced Study Institute

San Feliu de Guixols, Costa Brava, Spain

17-27 April 1990

Objective:

The aim is to instruct in the methods of charge-density analyses by X-ray and neutron diffraction and explore the use of this information in Chemistry and Drug Design.

Topics:

The basic theory relating crystal diffraction data to the electron distributions and the electrostatic potential of molecules will be given. This will be followed by a discription of experimental methods using X-ray, neutron and synchrotron X-ray radiation. Threoretical calculations of electrostatic properties will be discussed. Applications of these data and concepts to the understanding of hydrogen bonding, chemical reactivity, and drug design will be described.

Costs:

There is no registration fee. The cost of accomodations, double occupancy, and half-pension meals for the period of the Institute is 55,000 pts (%450 US) per person. A limited number of single rooms are available with an additional charge of 20,000 pts (%160 US).

Grants are available to assist with living expenses for predoctoral and junior postdoctoral participants from NATO countries.

Applications:

Application forms for the Institute are available from:

Professor G.A. Jeffrey

Department of Crystallography

University of Pittsburgh

Pittsburgh, PA 15260 USA

FAX: 412-624-1882

TELEX: 199126

PHONE: 412-624-9300

Professor J.F. Piniella

Departament de Cristallografia

Universitat Autonoma de Barcelona

08193 Bellaterra, Barcelona

Spain

FAX: 34-3-581-2003

PHONE: 34-3-581-1609

Applicants are advised to apply as soon as possible and certainly before December 15, 1989.

GENES PROTEINS & COMPUTERS

An International Conference on

COMPUTING IN MOLECULAR BIOLOGY

Chester, UK 18-20 April 1990

Scientific Programme:

There will be six plenary sessions, a poster session and an industrial exhibition. Topics covered include the Human Genome Project, sequence analysis, structure analysis, applications of computer architectures and novel software technologies/ algorithms.

Contributions are invited for poster presentations on these topics.

Speakers who have so far accepted are:

P. Pearson	S. Wodak	T. Blundell	D. Moss	R. Fuchs
N. Paton	S. Gardner	A. Bleasby	J. Collins.	

Details of the conference fee, accomodation charges etc, are available on request.

For further information contact:

G.P.C. Secretariat,
SERC Daresbury
Laboratory,
Warrington WA4 4AD, UK.
Tel. 0925-603235 or 0925-603305
Fax: 0925-603174

REGISTRATION FORM

CRYSTALLOGRAPHIC COMPUTING SCHOOL

SATELLITE OF

XV IUCR INTERNATIONAL CONGRESS

INTRODUCTION TO THE SCHOOL:

The school will cover topics on the crystallography of both small and large molecules, like data collection, program packages, direct methods, maximum entropy, multiple isomorphous replacement, phasing by anomalous scattering, molecular replacement, density modification, refinement (including molecular dynamics), modelisation, graphics and data bases.

The following speakers have already agreed to participate:

G. Bricogne	A. Brunger	D. Collins	P. Evans	C. Giacovazzo
C. Gilmore	J.Hajdu	H. Hauptman	J. Helliwell	A. Jones
W. Kabsch	J. Karle	A. Leslie	P.Main	A. Olson
A. Podjarny	R. Read	D. Sayre	D. Schwarzenbach	G. Sheldrick
W. Steigemann	J. Sussman	K. Watenpaugh	D. Watkin	K. Wilson

It will take place in Bischenberg, in the Vosges mountains near Strasbourg, from July 29 to August 5 1990. Strasbourg is easily accesible from Paris by train, private car or plane, and transportation will be provided from Strasbourg to the school site.

The registration fee, including inscription, full meals and boarding, is FF 4200 (around SF 1000 or US$ 700 at current exchange rates). Financial aid is available for a very limited number of students.

The total number of available places is 125 (students and lecturers).

IMPORTANT DATES:

Deadline for sending this form (postmark date): 1 feb. 1990.

Confirmation of place in the school will be sent by 15 february 1990.

Final registration and payment should be sent by 15 march 1990.

FILL IN THE FOLLOWING (type or clear print) :

Name:

Adress:

Position (faculty, postdoc, student):

Telephone:

BITNET adress:

Financial aid necessary (yes or no):

Note: If you cannot attend the school unless you have financial support, please answer yes . Remember that the number of places and the amount of financial aid is limited.

INCLUDE WITH THIS FORM:

1) A short CV, marking the three most important publications.

2) A statement of current research activities, marking in particular the way they will benefit from the school.

3) For students, a letter of thesis advisor.

If possible, send the filled form by bitnet to SCHOOL@FRIBCP51.

Mail your complete answer to :

J.C. Thierry.
Crystallographic Computing School
IBMC.
15, Rue Descartes.
67084 Strasbourg CEDEX FRANCE

If you have any particular enquiries, the telephone number is: (33)88-41-70-25 and the Fax number is(33)88-61-06-80.

New translation and packing functions

VAGIN A.A.

Institute of Crystallography, Academy of Science of the USSR, Leninsky pr.59, 117333, Moscow, USSR.

The purpose: From the known structure of some protein to obtain the structure of a homological protein or of the some protein packing in a different crystal form, in general with a different space group.

Suppose we know the orientation of the molecule in the unknown structure (for example from rotation function). To determine the position of the molecule in crystal we need to use a translation function. Usually the Crowther and Blow function is applied / 1 /. In 1981 in the paper of Y.Harada et al./ 2 / a new translation function was suggested. A similar function was worked out in 1982 in Institute of crystallography in Moscow / 3 /. This algorithm is more easy and efficient than that of Crowther and Blow. It is widely used in our laboratory.

Let Rm(x) be the electron density of the known structure in the same orientation as the unkown structure.

$$Rm(x) = \sum_{\bar{h}} Fm(\bar{h})*\exp[-2pi*\bar{x}*\bar{h}]$$

$Fm(\bar{h})$ is structure factor the known structure placed in the unit cell of unknown structure in some position.

Let the space group for unknown structure is defined by n symmetry operators.

$$\bar{x}j = [Aj]\bar{x}0 + \bar{d}j \;,\; j = 0,1,\ldots,n-1$$

[Aj] is the rotation matrix. $\bar{d}j$ is the vector of translation.
$R0(\bar{x})$ is electron density of the model in a point $\bar{s}$.
$R0(\bar{x}) = Rm(\bar{x}-\bar{s})$
$Rj(\bar{x})$ is the some electron density derived by j-th symmetry operator.
$Rj(\bar{x}) = R0([Aj](\bar{x}-\bar{d}j)) = Rm([Aj](\bar{x}-\bar{d}j)-s)$
$R(\bar{x})$ is entire electron density in the unit cell.

$$R(\bar{x}) = \sum_j Rj(\bar{x}) = \sum_{t\ j} Fm(\bar{t})*\exp[-2pi*\bar{t}*[A]^{-1}*\bar{x}]*\exp[2pi*\bar{t}*[A]^{-1}*\bar{dj}]*\exp[2pi*\bar{t}*\bar{s}]$$

Structure factors for this crystal form are

$$F(\bar{h}) = \int R(\bar{x})*\exp[2pi*\bar{h}*\bar{x}]*d\bar{x}$$
$$= \sum_t \sum_j Fm(\bar{t})*\exp[2pi*\bar{t}*[A]^{-1}*\bar{dj}]*\exp[2pi*\bar{t}*\bar{s}]*\int \exp[2pi*(\bar{h}-\bar{t}*[A]^{-1})*\bar{x})]*d\bar{x}$$

The Integral is not equal to zero only when

$$\bar{h} = \bar{t}*[Aj] \text{ or } \bar{t} = \bar{h}*[Aj]$$

Let us denote

$$Gj(\bar{h}) = Fm(\bar{h}*[Aj])*\exp[2pi*\bar{h}*\bar{dj}]$$

Then we may write the following equation

$$F(\bar{h}) = \sum_j Fm(\bar{h}*[Aj])*\exp[2pi*\bar{h}*\bar{dj}]*\exp[2pi*\bar{h}*[Aj]*\bar{s}]$$
$$= \sum_j Gj(\bar{h})*\exp[2pi*\bar{h}*[Aj]*\bar{s}]$$

Let us define the translation function as

$$T(\bar{s}) = \int [Pobs(\bar{u})-P(\bar{u},\bar{s})]\, d\bar{u}$$
$$= \int P^2obs(\bar{u})\, d\bar{u} + \int P^2(\bar{u},\bar{s})\, d\bar{u} - 2\int Pobs(\bar{u})*P(\bar{u},\bar{s})*d\bar{u}$$

or

$$T(\bar{s}) = \int Pobs(\bar{u})*P(\bar{u},\bar{s})*d\bar{u}$$
$$= \int \left(\sum_{\bar{h}} F^2obs(\bar{h})*\exp[-2pi*\bar{h}*\bar{u}]\right) * \left(\sum_t F(\bar{t})*F^*(\bar{t})*\exp[-2pi*\bar{t}*\bar{u}]\right) d\bar{u}$$
$$= \sum_{\bar{h}} \sum_j \sum_k F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h})*\exp[-2pi*\bar{h}*([Ak]-[Aj])*\bar{s}]$$

Where $Pobs(\bar{u})$ and $P(\bar{u})$ are Patterson funtions for unknown structure and for the model.

This translation function reaches the maximum when the vector $\bar{S}$ corresponds to the position of the model in unknown structure.

Let us expand translation function in the Fourier series.

$$T(\bar{s}) = \sum_{\bar{H}} A\bar{H}*\exp[-2pi*\bar{H}*\bar{s}]$$

$$A\bar{H} = \sum_h \sum_j \sum_k F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h}) \text{ , for only such } \bar{h}: \bar{H}=(\bar{h}*[Ak]-[Aj])$$

Coefficients AH may be calculated by following procedure:

1. The list of values of $F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h})$ and $\bar{H}$ for each $\bar{h}$,J,k is constructed.
2. The list is sorted out by Index $\bar{H}$.
3. Values of $F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h})$ with the some $\bar{H}$ are summed up to obtain $A\bar{H}$. They are collected together in list after sorting.
4. The translation function is calculated by FFT program.

Let us consider the expression for $T(\bar{s})$. If k equals zero and j is fixed we shall obtain the usual Crowther and Blow function.

$$T1(\bar{s}) = \sum_{\bar{h}} F^2_{obs}(\bar{h}) * Gj(\bar{h}) * G0^*(\bar{h}) * \exp[-2pi * \bar{h} * (\bar{s} - [A]] * \bar{s})]$$

Here we can see an advantage of the new function:
1. All symmetry operators are used simultaneously increasing a signal to noise ratio.
2. The new function gives the position of the model explicitly in the unit cell where as that of Crowther and Blow indicates the position relatively to an symmetry operator.

Let the electron density equal 1 inside the envelope of the model and zero in other points. We may consider it as the electron density map of low resolution or that calculated with high temperature factors. Let us suppose that Rk, which is the electron density of the model corresponding to k-th symmetry operator, is overlaped with Rj.

The overlapping is determined by

$$\int Rj(\bar{x}) * Rk(\bar{x}) * d\bar{x}$$

Let us define the packing function as

$$P(\bar{s}) = -\int \sum_{j} \sum_{\substack{k \\ j \neq k}} Rj(\bar{x}) * Rk(\bar{x}) * d\bar{x}$$

Minus is introduced to make it maximal for the least overlapping. By transformations similar to that done above we may obtain the expression for the packing function.

$$P(\bar{s}) = -\sum_{\bar{h}} \sum_{j} \sum_{\substack{k \\ j \neq k}} Gj(\bar{h}) * Gk^*(\bar{h}) * \exp[-2pi * \bar{h} * ([Ak] - [Aj]) * \bar{s}$$

It may be used autonomously for the search of molecular positions in init cell or for analysis of translation function peaks, for example by normalization of translation function.

$$T'(\bar{s}) = T(\bar{s}) * P(\bar{s})$$

Let us return to translation function. If we take as a model the sphere of small size (with diameter about 3 A) and as observed structure factors the differences between structure factors for the native protein and for heavy-atom derivative then the function would be similar to that used for heavy atom search in vector space. Indeed the translation function in this case is the sum of values of the difference Patterson function for all interatomic vectors generated by a heavy atom in the position s. Advantages of this function over function in direct space are calculation speed and the absence of approximation errors for the Patterson function as a grid of points.

If in calculation of translation function a model is a sphere with diameter roughly equal to that of molecule, and Fobs are experimental structure factors for native protein then this function may be used to search a rough position of the molecule in unit cell, assuming that at given resolution the molecule is not too much distinguished from a sphere. Suppose that the model consists of two parts each being placed in init cell undependently.

$$Rm(\bar{x}) = Rm1(\bar{x}) + Rm2(\bar{x})$$

Then the translation function may be consider as a function of two shift vectors that is as a function of six variables. For its calculation a six-dimansional FFT should be required.

$$T(\bar{s}1,\bar{s}2) = \sum_{H1}\sum_{H2} A\bar{H}1\bar{H}2 * \exp[-2pi*\bar{H}1*\bar{s}1]*\exp[-2pi*\bar{H}2*\bar{s}2]$$

$$AH1H2 = \sum_{h}\sum_{j}\sum_{k}$$

$$= F^2obs(\bar{h})*G1j(\bar{h})*G2k^*(\bar{h}) + \quad \text{for } \bar{h} : \bar{H}1=\bar{h}*([Ak]-[Aj]),\ \bar{H}2=0$$

$$+ F^2obs(\bar{h})*G1j(\bar{h})*G2k^*(\bar{h}) + \quad \text{for } \bar{h} : \bar{H}1=0,\ \bar{H}2=\bar{h}*([Ak]-[Aj])$$

$$+ 2F^2obs(\bar{h})*G1j(\bar{h})*G2k^*(\bar{h}) \quad \text{for } \bar{h} : \bar{H}1=-\bar{h}*[Aj],\ \bar{H}2=\bar{h}*[Ak]$$

If the position of one part of the model is known we would have a usual translation function for the other part.

By similar way one may derive a packing function for two models. Using small spheres one may search in heavy-atom structures positions of two heavy atoms simultaneously.

However, if one wants to search positions of two subunits of protein molecule using spherical approximation then regions where the spheres overlap would have electron density of double value. For this reason the electron density should be modified:

$$Rm(\bar{x}) = Rm1(\bar{x}) + Rm2(\bar{x}) - Rm1(\bar{x})*Rm2(\bar{x})$$

This expression would give electron density equally filling the space inside the combined envelope. Translation function in this case may by written as:

$$T(\bar{s}1,\bar{s}2) = \sum_{\bar{H}1}\sum_{\bar{H}2} A\bar{H}1\bar{H}2 * \exp[-2\pi i*\bar{H}1*\bar{s}1] * \exp[-2\pi i*\bar{H}2*\bar{s}2]$$

$$A\bar{H}1\bar{H}2 = \sum_{\bar{h}}\sum_{j}\sum_{k}$$

$= F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h}) +$ for $\bar{h}$: $\bar{H}1=-\bar{h}*([Ak]-[Aj])$, $\bar{H}2=0$

$+ F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h}) +$ for $\bar{h}$: $\bar{H}1=0$, $\bar{H}2=\bar{h}*([Ak]-[Aj])$

$+ F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h}) +$ for $\bar{h}$: $\bar{H}1=-\bar{h}*[Aj]$, $\bar{H}2=\bar{h}*[Ak]$

$+ F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{h}) -$ for $\bar{h}$: $\bar{H}1=\bar{h}*[Ak]$, $\bar{H}2=-\bar{h}*[Aj]$

$-\sum_{\bar{m}} F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{m})*Gk^*(\bar{h}-\bar{m})$ for : $\bar{H}1=-\bar{h}*[Aj]+\bar{m}*[Ak]$, $\bar{H}2=(\bar{h}-\bar{m})*[Ak]$

$-\sum_{\bar{m}} F^2obs(\bar{h})*Gj(\bar{h})*Gk^*(\bar{m})*Gk^*(\bar{h}-\bar{m})$ for : $\bar{H}1=\bar{m}*[Ak]$, $\bar{H}2=-\bar{h}*[Aj]+(\bar{h}-\bar{m})*[Ak]$

$-\sum_{\bar{t}} F^2obs(h)*Gj^*(\bar{m})*Gj(\bar{h}-\bar{m})*Gk(\bar{h}) -$ for : $\bar{H}1=\bar{h}*[Ak]-\bar{m}*[Aj]$, $\bar{H}2=-(\bar{h}-\bar{m})*[Aj]$

$-\sum_{\bar{t}} F^2obs(h)*Gj^*(\bar{m})*Gj(\bar{h}-\bar{m})*Gk(\bar{h}) +$ for : $\bar{H}1=-\bar{m}*[Aj]$, $\bar{H}2=\bar{h}*[Ak]-(\bar{h}-\bar{m})*[Aj]$

$+\sum_{\bar{n}}\sum_{\bar{t}} F^2obs(\bar{h})*Gj(\bar{l})*Gj(\bar{h}-\bar{l})*Gk^*(\bar{m})*Gk^*(\bar{h}-\bar{m})$

for : $\bar{H}1=\bar{m}*[Ak]-\bar{l}*[Aj]$, $\bar{H}2=(\bar{h}-\bar{m})*[Ak]-(\bar{h}-\bar{l})*[Aj]$

where $\bar{t} = \bar{l}*[Aj]$, $\bar{n} = \bar{m}*[Ak]$

At last, let the model to consist of n part. Translation and packing functions may be defined for this case as above. To calculate such functions 3n-dimensional Fourier transformation would be needed.

References

1. R.A.Crowter & D.M.Blow, Acta Cryst. 23, (1967) 544-548.
2. Y.Harada, A.Lifchitz & J.Berthou, Acta Cryst. A37, (1981) 398-406.
3. A.A.Vagin, Ph.D.Thesis (1982) /in russian/.

A.A. Vagin is a Royal Society Guest Fellow at the Krebs Institute, University of Sheffield.

New crystallizations of phosphorylase *b*

N.G. Oikonomakos, A.C. Papageorgiou, D.D. Leonidas,

D. Barford* and L.N. Johnson*

*National Hellenic Research Foundation, 48 Vas. Constantinou, Athens 11635, Greece and *Laboratory of Molecular Biophysics, South Parks Road, Oxford OX1 3QU, UK.*

Glycogen phosphorylase is an allosteric enzyme that exists in two interconvertible forms, *b* and *a*. Phosphorylase *b* is inactive but can be allosterically activated by AMP (or IMP) and by phosphorylation of one specific serine residue [1]. The AMP activation or the phosphorylation of the enzyme can be understood as a conversion from a dimeric T-state (low affinity) to a dimeric R-state (high affinity) [2]. High concentration of substrate anions and several anions high on the Hofmeister series such as sulphate are also able to activate phosphorylase *b* to a considerable extent [3,4] and this activation can be further stimulated by AMP [5]. *In vitro* activation of the enzyme is accompanied by a dimer to tetramer conversion.

The growth of a monoclinic crystal (I) of phosphorylase *b* (space group $P2_1$ with unit cell dimensions a =119 Å, b =190 Å, c =88.2 Å, β =109.35° and one tetramer of M_r of 390,000 per asymmetric unit) in the presence of 1.0-1.2 M of ammonium sulphate was reported some time ago [6-8] and its three-dimensional structure has been recently described [9]. The X-ray analysis at a resolution of 2.9 Å revealed that sulphate mimics the substrate phosphate by binding to the serine phosphate site, resulting in

localized changes in tertiary structure. These changes are coupled to large changes in quaternary structure which directly affect the AMP and the Ser-14 phosphate site and indirectly the catalytic site [9]. It appears therefore that ammonium sulphate works as an activator in place of phosphate at the phosphorylation site at Ser-14.

Phosphorylase *b* from rabbit skeletal muscle has been crystallized in a new monoclinic form **(II)** in the presence of the activator AMP (1-2 mM) from 1.2-1.4 M-ammonium sulphate by the microdialysis method. Under the conditions of crystallization phosphorylase *b* exists as a tetramer and is practically saturated by the nucleotide *(K_d=3 μM)* (D.D. Leonidas *et al.*, unpublished results). The crystals belong to a monoclinic space group $P2_1$ with unit cell dimensions a =119 Å, b =190 Å, c =176 Å, β =110°. In the new crystal form the c -dimension is twice the c -dimension of the previous crystal form **(I)** and this results in a doubling of the cell volume. The increased cell has two tetrameric molecules per asymmetric unit instead of one tetramer as in the first cell. The crystals diffract X-rays to a resolution of at least 2.8 Å and are suitable for X-ray structure analysis. Since the new crystal is the first crystal of phosphorylase co-crystallized with the physiological activator AMP, it will be of interest to compare with the high affinity binding mode observed when AMP is diffused into preformed R-state crystals. This structure has been refined (D. Barford, unpublished results). A similar increase in the c -axis is observed in the R-state phosphorylase-glucose-1-P complex and analysis of this crystal form is now at the refinement stage (D. Barford, S.-H. Hu and L.N. Johnson).

Crystals of phosphorylase *b* have been obtained by using 1.0 M tartrate instead of ammonium sulphate. These crystals may be used as an alternative in studying catalysis, since tartrate is less likely to bind at the catalytic site but it is not yet known if they correspond to type I or type II monoclinic crystal forms.

Crystals of phosphorylase *b* reconstituted with pyridoxal, 5'-deoxy-pyridoxal or pyridoxal-5'-diphosphate in place of the natural cofactor pyridoxal-5'-phosphate were produced with perfect monoclinic morphology following the usual procedure. We have been able to cocrystallize these derivative proteins in the presence of AMP as well. These crystal forms are interesting because they may shed light on interactions made by modified cofactors.

In an attempt to search for other favourable crystal forms of R-state phosphorylase we have been able to crystallize pig muscle phosphorylase *b* in the presence of AMP. Pig muscle phosphorylase *b* does not show any tendency to tetramerize under conditions where rabbit muscle phosphorylase *b* does [10]. These crystals have not yet characterized.

Acknowledgements. The crystallographic work on the preliminary characterization of the new crystal forms was supported by a grant (N.G.O.) from the Royal Society.

References

[1] Graves, D.J. and Wang, J.H. (1972) in: *The Enzymes,* 3rd edn, Vol 7 (Boyer, P., ed.), pp. 435-482, Academic Press, New York.

[2] Johnson, L.N., Hajdu, J., Acharya, K.R., Stuart, D.I., McLaughlin, P.J., Oikonomakos, N.G., and Barford, D. (1989) in: *Allosteric Enzymes,* (Herve, G. ed.), pp. 81-127, CRC Press, Boca Raton, Florida

[3] Buc, H (1967) *Biochem. Biophys. Res. Commun.* **28**, 59-64.

[4] Engers, H.D. and Madsen, N.B. (1968) *Biochem. Biophys. Res. Commun.* **33**, 49-54.

[5] Sotiroudis, T.G., Oikonomakos, N.G. and Evangelopoulos, A.E. (1979) *Biochem. Biophys. Res. Commun.* **90**, 234-239.

[6] Mathews, F.S. (1967) *Federation Proc.* **26**, 831.

[7] Madsen, N.B., Honikel, K.O. and James, M.N.G. (1972) in: *Metabolic Interconversion of Enzymes*, (Wieland, O., Helmreich, E. and Holzer, H., eds.), pp. 55-71, Springer, Berlin.

[8] Fasold, H, Ortanderl, F., Huber, R., Bartels, K. and Scwager, P. (1972) *FEBS Lett.* **21**, 229-232.

[9] Barford, D. and Johnson, L.N. (1989) *Nature*, **340**, 609-616.

[10] Oikonomakos, N.G., Melpidou, A.E. and Johnson, L.N. (1985) *Biochim. Biophys. Acta*, **832**, 248-256.

CRYSTAL PACKING OF PROTEIN MOLECULES

A.Teplyakov, B.Vainstein

Institute of Crystallography, Moscow

One of the main principles of the molecular crystal formation is the principle of maximum filling of the space or what is the same the principle of close packing of the molecules. For the organic crystals close packing ("bulge" to "concave") can be reached by planes and axes of symmetry with translational component as the most suitable paking elements of symmetry. It is known that the space groups $P2_1/c$, $P2_12_12_1$, $P2_1$ are the most frequent among the organic crystals [1].

Only 65 space groups (without center and planes of symmetry) out of 230 are possible for the protein crystals because of the chirality of the amino acids. So the screw axes should be the main packing elements of symmetry as well as the simple translation. The rotational axes of symmetry usually correspond to the symmetry of the molecules.

To check these considerations we performed the statistical analysis of the distribution of 436 crystal forms of globular proteins over the space groups (table 1). The analysis is based on the X-ray structure studies published during last 10 years.

About one half of all crystal forms belong to the space groups $P2_12_12_1$, $P2_1$ and C2. This fact suggests that the main principles of the crystal formation are the same for organic molecules and proteins.

For the hexagonal crystals the tendency to close packing

results in the predominance (18:4) of the crystal forms with the axes 6_1 (6_5) compared to those with the axes 6_2 (6_4). It can be explained by the absence of the twofold screw axes in the space groups $P6_2$ ($P6_4$) and $P6_222$ ($P6_422$). There are such axes in the space groups $P6_3$ and $P6_322$, but there are no screw threefold axes - only rotational, and this fact excludes the possibility of close packing.

The similar conclusions one can make for the tetragonal crystals. The number of the crystal forms with the axes 4_1 (4_3) is 4 times greater than that with the 4_2 axes.

The possibility of crystallization in the space groups P2, P222, P3, P4, P6 which do not contain screw axes is very small. Among all 436 crystal forms under inspection only 3 such cases were found: hemerythrin [2] in the space droup P4, superoxide dismutase [3] and ornithine decarboxilase [4] in the space group P6. It is necessary to note that in the cases of hemerythrin and ornithine decarboxilase symmetry of the crystals coincides with that of the molecules.

Many protein molecules consist of two or more identical subunits, i.e. exist as dimers, trimers, etc. These subunits are related by the axes of symmetry which can coincide with the crystallographic axes in crystals. An oligomeric protein can be crystallized in several crystal forms so that the molecules can occupy various symmetry positions. For example, tetrameric molecule of catalase with the own symmetry 222 can occupy the asymmetric position in the space groups $P3_121$ [5] and $P2_12_12_1$ [6], the position on the crystallographic axis 2 in the space group $P3_221$ [7] and the position with symmetry 222, corresponding to the molecular symmetry, in the space group $P4_22_12$ [8].

Statistical analysis of the crystal forms of oligomeric proteins shows that the higher the symmetry of the molecule, the more probable its position on crystallographic axis (table 2). The asymmetric position is occupied by more than 60% of dimers, about half of tetramers and by only two hexamers out of 16. Molecules containing 8, 12 and 24 subunits with the corresponding symmetry 422, 622 and 432 lie on the crystallographic axes of symmetry in 12 cases out of 14.

As well as monomers oligomeric molecules are crystallized mainly in the space groups $P2_12_12_1$, $P2_1$ and C2 (table 3). It is interesting that the molecules crystallized in the space group C2 almost always lie on the crystallographic twofold axis.

Tetramers with the molecular symmetry 222 can occupy the positions with symmetry 2 and 222, although the former - more frequently. For hexamers the positions with symmetry 2, 3 and 32 are equally probable.

The comparison of tables 1 and 3 shows that almost all cubic and rhombohedric crystal forms contain protein molecules in the positions with symmetry corresponding to that of the molecules.

The analysis of distribution of the protein crystal forms confirms the validity of the principle of close packing for globular proteins. Symmetry of the molecules results in the appearance of the crystal forms with the rotational axes of symmetry.

Table 1. Distribution of 436 crystal forms of proteins over the space groups.

1.	$P1$	15	
2.	$P2$	0	
3.	$P2_1$	55	
4.	$C2$	46	
5.	$P222$	0	
6.	$P222_1$	1	
7.	$P2_12_12$	19	
8.	$P2_12_12_1$	105	
9.	$C222_1$	14	
10.	$C222$	5	
11.	$F222$	2	
12-13.	$I222$, $I2_12_12_1$	9	
14.	$P4$	1	
15-16.	$P4_1$, $P4_3$	8	
17.	$P4_2$	2	
18.	$I4$	3	
19.	$I4_1$	1	
20.	$P422$	2	
21.	$P42_12$	2	
22-23.	$P4_122$, $P4_322$	6	
24-25.	$P4_12_12$, $P4_32_12$	23	
26.	$P4_222$	3	
27.	$P4_22_12$	5	
28.	$I422$	5	
29.	$I4_122$	2	
30.	$P3$	0	
31-32.	$P3_1$, $P3_2$	3	
33-34.	$P312$, $P321$	3	
35-38.	$P3_112$, $P3_121$, $P3_212$, $P3_221$	34	
39.	$R3$	7	
40.	$R32$	7	
41.	$P6$	2	
42-43.	$P6_1$, $P6_5$	7	
44-45.	$P6_2$, $P6_4$	1	
46.	$P6_3$	4	
47.	$P622$	2	
48-49.	$P6_122$, $P6_522$	11	
50-51.	$P6_222$, $P6_422$	3	
52.	$P6_322$	6	
53.	$P23$	0	
54.	$F23$	0	
55.	$I23$	1	
56.	$P2_13$	4	
57.	$I2_13$	0	
58.	$P432$	0	
59.	$P4_232$	1	
60.	$F432$	2	
61.	$F4_132$	1	
62.	$I432$	1	
63-64.	$P4_132$, $P4_332$	2	
65.	$I4_132$	0	

Table 2. Distribution of crystal forms of proteins over the number of subunits in the asymmetric part of the unit cell.

Molecules	Number of subunits in the asym. part				
	1	2	3	4	more
Monomers	175	37	5	7	0
Dimers	30	38	-	7	2
Trimers	2	-	0	-	0
Tetramers	6	16	-	19	2
Hexamers	3	5	6	-	2
Octamers	1	3	-	3	1

Table 3. Distribution of crystal forms of oligomeric proteins over the space groups. Own symmetry of the molecules is in the brackets.

Dimers (2)		Tetramers (222)		Hexamers (32)		Trimers (3)	
in the pos. with symm. 1		in the pos. with symm. 1		in the pos. with symm. 1		in the pos. with symm. 1	
$P2_12_12_1$	16	$P2_12_12_1$	9	$P2_12_12_1$	1	R32	1
$P2_1$	8	$P2_1$	3	$P2_1$	1	$F4_132$	1
$P6_1$	5	P1	3				
$P4_12_12$	3	$P2_12_12$	2	in the pos.			
P1	2	F222	2	with symm. 2		Octamers (422)	
$P2_12_12$	2	$P4_12_12$	1				
I222	2	$P3_121$	1	C2	2	in the pos.	
$P4_122$	2			$P4_222$	1	with symm. 1	
$P3_121$	2	in the pos.		$I4_122$	1	P1	1
C2	1	with symm. 2		$P3_121$	1		
$C222_1$	1			$P6_122$	1	in the pos.	
R3	1	$P2_12_12$	5			with symm. 2	
P6	1	$P3_121$	4	in the pos.			
$P6_322$	1	C2	2	with symm. 3		C222	1
		$P222_1$	1			$C222_1$	1
in the pos.		$P4_22_12$	1	P321	2	$P4_12_12$	1
with symm. 2		$P4_12_12$	1	R3	1		
		$P6_122$	1	R32	1	in the pos.	
C2	10	$P6_222$	1	$P2_13$	1	with symm. 4	
$P3_121$	5						
$C222_1$	4	in the pos.		in the pos.		P4	1
$P4_12_12$	4	with symm.222		with symm. 32		I4	1
$P2_12_12$	2	C222	2			P422	1
I222	1	I222	2	R32	1		
$P4_22_12$	1	$P4_22_12$	2	$P6_322$	1	in the pos.	
$P4_222$	1	$P42_12$	1	$P4_132$	1	with symm.422	
$P2_13$	1	I422	1				
$P4_232$	1	I23	1			I422	1

References

1. Belsky V.K., Zorky P.M. Kristallografia (Rus.), 1970, 15, 704-709.
2. Stenkamp R.E., Sieker L.C., Jensen L.H., Loehr J.S. J. Mol. Biol., 1976, 100, 23-34.
3. Parge H.E., Getzoff E.D., Scandella C.S., Hallewell R.A., Tainer J.A. J. Biol. Chem., 1986, 261, 16215-16218.
4. Momany C., Hackert M.L. J. Biol. Chem., 1989, 264, 4722-4724.
5. Vainstein B.K., Barynin V.V., Gurskaya G.V., Nikitin V.A. Kristallografia (Rus.), 1967, 12, 860.
6. Gurskaya G.V., Karpuhina S.A., Lobanova G.M. Biofizika (Rus.), 1971, 16, 553.
7. Eventoff W., Tanaka N., Rossmann M.G. J. Mol. Biol., 1976, 103, 799-800.
8. Marie A.L., Parak F., Hoppe W. J. Mol. Biol., 1979, 129, 675-676.

Structure Determination of Haloalkane Dehalogenase

Sybille M. Franken, Henriëtte J. Rozeboom, Kor H. Kalk and Bauke W. Dijkstra

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

The enzyme haloalkane dehalogenase has been isolated from the nitrogen-fixing bacterium Xanthobacter autotrophicus GJ10 which is able to grow on 1,2-dichloroethane as a sole carbon and energy source [1]. The enzyme catalyzes the cleavage of halogenated alkanes releasing a halide ion and the corresponding alcohol. Neither oxygen nor cofactors are needed, suggesting a nucleophilic substitution with water as reaction mechanism. The enzyme was cloned, expressed and purified to homogeneity [2]. It consists of a single polypeptide chain with a molecular weight of 35,000. The determination of the nucleotide sequence [3] revealed the existence of 4 cysteine residues. As the enzyme is inhibited by thiol reagents such as $HgCl_2$, iodoacetamide and p-chloromercuribenzoate, it has been suggested that a cysteine residue is involved in the catalysis.

Substrates for haloalkane dehalogenase are many alkanes with a chain length of one to four carbon atoms containing one or two chlorine or bromine atoms. Among the compounds a number of major pollutants can be found. Dehalogenase offers a simple possibility for detoxification. However, drawbacks are its relatively low activity (K_m is 1.1 m Mol, V_{max} is 10 µmol/min per mg of protein) and its limited substrate specificity. Therefore a long term goal of our collaborative protein-engineering project is to make the enzyme suitable for application in the degradation of halogenated pollutants. To obtain more insight into the reaction mechanism and to learn which factors determine the substrate specificity we started a crystallographic investigation of the three-dimensional structure of haloalkane dehalogenase [4].

Pure enzyme material was generously made available by Dr. D. Janssen and his coworkers. The protein could be crystallized at room temperature in hanging drops containing bis-tris buffer (pH 5.6-6.8) and ammonium sulfate as precipitant (60-65% saturated). Because of pH-dependent variations in cell dimensions all crystals were soaked in a standard mother liquor (pH 6.2, 64% ammonium sulfate) before being mounted for X-ray experiments. The space

group of the crystals is $P2_12_12$ with unit cell dimensions of a = 94.8 Å, b = 72.8 Å and c = 41.4 Å. With one molecule per asymmetric unit the solvent content is 39%. The crystals diffract to 2.1 Å resolution.

Heavy atom derivatives were prepared by soaking the native crystals for one day in the corresponding solutions. Diffraction data for native and derivative crystals were collected at room temperature on our FAST area detector. For each data set one single crystal was used (see table 1).

Table 1 Data collection

	maximum resolution	total/unique observations		R_{merge}*	$R_{to\ native}$#	completeness to 2.7 Å
native I	2.3	23332	8684	3.79 (F)		75%
native II	2.4	29258	10866	7.1 (I)		94%
native I + II	2.3		11506			94%
$K_2[PtCl_4]$ (1m Mol)	2.44	21645	7251	5.69 (F)	17.1% (F)	76%
$UO_2(NO_3)_2$ (3m Mol)	2.45	26231	7986	4.83 (F)	14.0% (F)	82%
$Na[Au(CN)_2]$ I (3m Mol)	2.44	19092	7435	4.99 (I)	17.2% (F)	78%
II		34819	9467	5.57 (I)	17.0% (F)	90%

$$*R_{merge} = \frac{\sum_{hkl} \sum_{N} |F_i^{hkl} - \langle F^{hkl} \rangle|}{\sum_{hkl} \langle F^{hkl} \rangle} \qquad \#R_{to\ native} = \frac{\sum |F_{PH} - F_P|}{\sum F_P}$$

The gold and platinum sites could easily be localized in isomorphous and anomalous difference Patterson maps, whereas the uranyl site was determined from a cross difference Fourier map. The problem of the correct hand could be solved by using difference Fourier techniques as well. The heavy atom parameters were refined and "best phases" were calculated (program PHARE). The results of the refinement are given in table 2.

Table 2 Results of the heavy atom refinement

derivative	site	x	y	z	real/anom occupancy		B(Å²)	resolution range (Å)	overall phasing power*
$K_2[PtCl_4]$	1	.42	.30	.14	.75	.65	24.1		
								3.3 - 20	1.37
	2	.06	.24	.47	.90	.85	78.8		
$UO_2(NO_3)_2$	1	.16	.12	.55	.43	.35	7.0 (fixed)	3.1 - 8	1.11
$NA[Au(CN)_2]$	1	.31	.44	.73	.94	1.29	7.8	2.7 - 20	2.64

overall figure of merit for 7703 reflection (2.7 - 20 Å) .69

* phasing power = F_H/E

A native Fourier map was calculated and the molecular boundaries were clearly visible. To enhance the features of the electron density we applied a solvent flattening [5]. During the whole procedure one molecular envelope was used which was calculated using A. Leslie's algorithm [6] with an integration radius of 7 Å and an assuming solvent content of 32%.

In this map the complete polypeptide chain could be traced and the sequence was fitted into the density. There was no density for the first two and the last three residues. The model is currently refined at a resolution of 2.5 Å. We hope to be able to locate the active site in the refined model.

Haloalkane dehalogenase shows the characteristic features of an α/β protein [7]. In the center of the molecule an eight-stranded β-sheet is located. This sheet is surrounded by α-helices. The dimensions of the globular molecule are about 45 x 45 x 40 Å³. A stereopicture of the C_α-tracing of the current model is shown in figure 1.

References

[1] Janssen, D.B., Scheper, A., Dijkhuizen, L. & Witholt, B. (1985) Appl. Environm. Microbiol. 49, 673-677.

[2] Keuning, S., Janssen, D.B. & Witholt, B. (1985) J. Bacteriol. 163, 635-639.

[3] Janssen, D.B., Pries, F., van der Ploeg, J., Kazemier, B., Terpstra, P.

& Witholt, B. (1989)
[4] Rozeboom, H.J., Kingma, J., Janssen, D.B. & Dijkstra, B.W. (1988) J. Mol. Biol. 200, 611-612.
[5] Wang, B.C. (1982) Methods Enzymol. 115, 90-112.
[6] Leslie, A.G.W. (1987) Acta Cryst. A43, 134-136.
[7] Richardson, J.S. (1981) Adv. Prot. Chem. 34, 167-339.