

DARESBURY LABORATORY
INFORMATION QUARTERLY
for
PROTEIN CRYSTALLOGRAPHY

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

Number 20

June 1987

Contents

Editorial	1
Fast detector performance (Uli Arndt, MRC Cambridge)	3
Analysis of Reflection data collected using the Phase III Software for the FAST area detector (Anne Bloomer et al., MRC Cambridge)	17
Progress report on the FAST diffractometer and a new method of response calibration of an area detector using single-crystal diffraction intensities (Glaucius Oliva & Ian Tickle, Birkbeck)	21
A note on the performance of the Xenonics/Nicolet area detector installed at York University (Zygmunt Derewenda, York)	29
The area detector software discussion group (John Campbell, SERC Daresbury)	33
Progress report on the X-ray structural analysis of diferric rabbit serum transferrin (Peter Lindley et al., Birkbeck)	37
Program to contour protein channels and cavities (David Barford & Louise Johnson, Oxford)	41
MOPAC for beginners (Steven Glover, Leeds)	45
A new dialysis crystallisation technique (David Thomas, Sheffield)	49
An improved ribbon algorithm for proteins (Phil Evans, MRC Cambridge)	51
A bug in PROTIN/PROLSQ (Eleanor Dodson, York)	53
Some UK crystallography JANET addresses (Andrew Lyall, Bristol)	55

Editor for
this Issue:

Dr. Phil Evans

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH.

Deputy Editor:

Sue Bailey

Science and Engineering Research Council,
Daresbury Laboratory, Daresbury,
Warrington WA4 4AD, England.

EDITORIAL

This issue of the Newsletter is overshadowed by the death of Pella Machin, who has always edited it in the past. We are all still trying to understand how we shall manage without her. I thank Tony North for writing the note below, and all the contributors for their articles.

Please note that this is not a formal publication, and permission to refer to this document should be sought from the authors.

Phil Evans

All readers of the Information Quarterly for Protein Crystallography will have been both shocked and saddened to learn of Pella Machin's death in a mountaineering accident on March 7th. She had played a central role in CCP4 from its inception - indeed for many of us, Pella was CCP4, or so it has seemed. After graduating from Bristol in Physics with the top first of her year, she joined David Phillips's laboratory in Oxford in 1967, subsequently moving to the Atlas Laboratory and then to Daresbury in 1977, becoming Head of the Applications Group in 1985.

Pella's long association with protein crystallographers was of immense help in the running of CCP4 as she knew all of us in the U.K. labs and very many of our friends overseas. The smooth running and pleasant atmosphere of a succession of CCP4 meetings were the most public evidence of her organisational abilities; those of us involved with one or other of the CCP4 working groups, and especially with preparing the case for renewal of funding for the Project, had an even better opportunity to appreciate her efficiency, reliability and firm support. Keeping track of the Project's finances, writing minutes of meetings, arranging for the production and distribution of the Quarterly, helping to plan conferences, her involvement ensured that all would be done well and with the minimum of fuss. We shall miss her help and common sense greatly - above all, we shall miss her warm and friendly personality.

We shall remember also with sorrow and gratitude Mike Elder, who died with her; his support for CCP4, if less well known, was nonetheless of great help to us. We have lost two good and valued friends.

A.C.T.N.

FAST DETECTOR PERFORMANCE

U.W. Arndt,

MRC Laboratory of Molecular Biology, Cambridge.

1. Introduction

The electronic area detector used in the Enraf-Nonius FAST (Fast-scanning Area-Sensitive Television) diffractometer is an analogue detector: the numerical output produced by the detector is a measure of the incident X-ray intensity but this output is not expressed directly in terms of X-ray photons per unit area per unit time. The purpose of the present note is to explain the functioning of the detector and the way in which its electronic performance is optimised for the particular crystallographic measurements to be made with it.

The detector is shown in schematic form in figure 1.

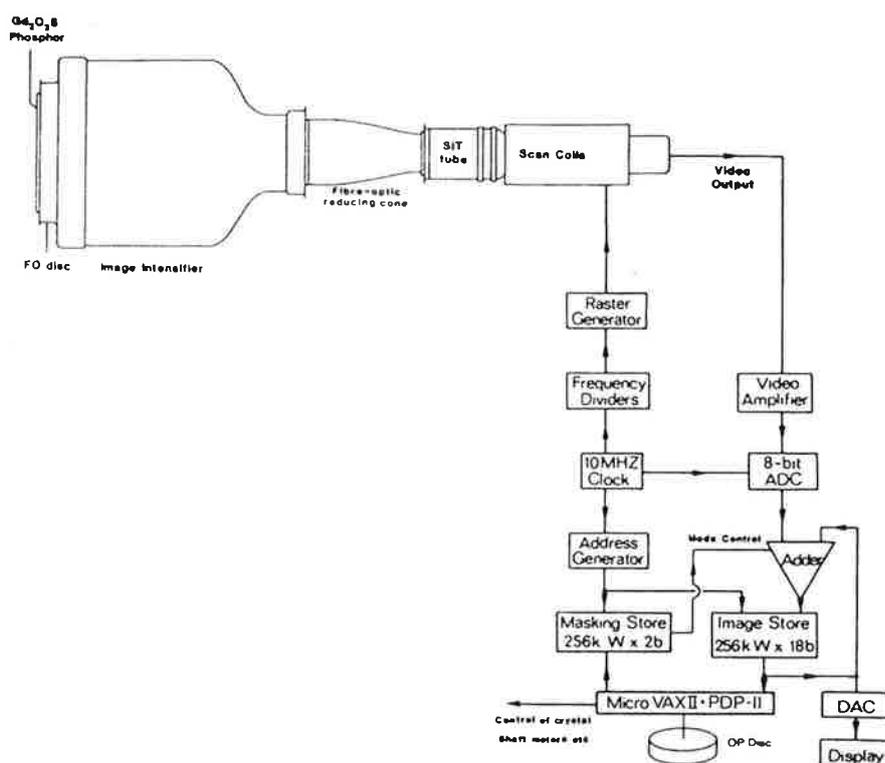


Fig.1

The design concepts have been discussed by Arndt & Gilmore (J.Appl.Cryst. 12, 1-9, 1979) and by Arndt (Nucl. Instrum. & Meth, 201, 13,20, 1982). The standard thickness of the Gd_2O_2S X-ray phosphor is 10 mg cm^{-2} . When the detector is to be used predominantly with radiation of wavelength less than 1\AA the deposit thickness can be increased to at least 30 mg cm^{-2} with a consequent gain in absorption at the

cost of a small loss in spatial resolution. A, the fraction of X-rays absorbed in these layers is shown in figure 2.

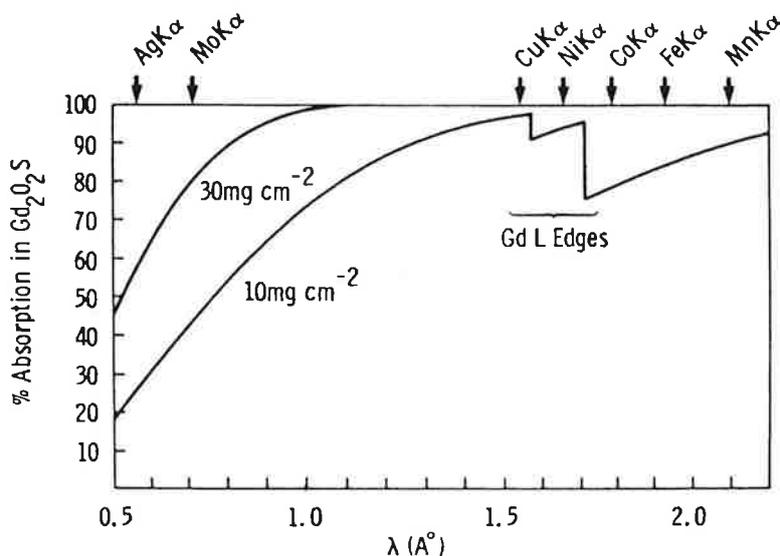


Fig.2

The necessary de-magnification between the 80 mm diameter input screen and the 18 mm diameter silicon intensifier target (SIT) TV camera tube is provided in part electron-optically in the image intensifier and in part by the tapered fibre-optics cone. The TV camera is operated in accordance with the European interlaced 625-line standard; the duration of each line is 64 μ s, 51.2 μ s of which constitute the active line period, the remainder being needed for the line fly-back. The active frame consists of 512 lines and each active line is sampled at 512 points. There is thus an inactive time of (625-512) line periods = 7.2 msec in every frame of 40 msec. Each frame consists of two fields of 20 msec, containing the 256 odd and even lines, respectively.

2. Definitions

Pixel (picture element, abbreviated px)

An element of the 2-dimensional area representing the digitized image (512 x 512 points). A pixel can refer to the direct image from the detector or to the

contents of the mass store. In the direct image a pixel value (intensity) is generated every 100 ns during the active period. In the mass store a pixel value represents the integrated value in the address corresponding to the specified pixel. The size of the active area of the detector is 64 mm (vertical) x 48 mm (horizontal). The nominal pixel size, therefore, is 0.125 mm x 0.094 mm.

ADU (ADC unit)

This is the digital unit in which the output of the detector Analog-Digital Converter is expressed. The detector uses an 8-bit ADC. Each pixel intensity in the live image, therefore, is represented by a number between 0 and 255.

MSU (Mass store unit)

This is the digital unit in which the mass store contents are expressed as seen by the host computer. Of the total mass store word width of 18 bits, the 16 most significant bits are available to the host computer. As the ADC output is mapped directly in the mass store the conversion factor between ADU and MSU is 4. An ADC value of 3 will be seen as 0 in the mass store and an ADC value of 20 as 5 MSU.

The truncation results in a small fixed mean error in the pixel intensity in the stored image when a positive integration is followed by a negative integration. The mean integrated intensities in MSU for the two operations, \bar{P} and $-\bar{N}$, will be

$$\bar{P} = \frac{\sum p}{4} + \bar{R} \text{ and}$$

$$-\bar{N} = \frac{-\sum n}{4} + \bar{R} \text{ where}$$

$$\bar{R} = (0+1+2+3)/4 = 3/2 \text{ is the mean round-off error and}$$

p and n are the individual ADC outputs in ADU. Then

$$\bar{P} - \bar{N} = (\sum p - \sum n)/4 + 3$$

so that the desired difference is given by

$$\sum p - \sum n = 4(\bar{P} - \bar{N}) - 3$$

Frame Period (abbreviated fp)

Time period of 40 ms, being the time to scan one complete image; each pixel value is read out once a frame. Thus the frame frequency is 25 Hz.

Dark current

The current from the camera tube in the absence of light on the input screen of the tube. This current originates in the leakage of the target diodes in the camera tube. It consists mainly of a DC component. There is a small AC component, called 'shot noise', due to the statistical fluctuations at the leakage current. Only the AC component contributes to the total detector noise.

Equivalent Background Illumination (EBI)

The small output signal of the image intensifier in the absence of an input due to the random emission of electrons from the photocathode which becomes visible

in the final output as individual scintillations. The time-integral of these scintillations contribute to the total offset which is removed in difference measurements. The fluctuations in this signal contribute to the total detector noise and are discussed further below.

ADC offset

DC offset added to the signal before it is digitized in the Analog to Digital convertor (ADC). The analog detector signal is digitized within the ADC span range. This span in volts is the sum of the ADC offset and the ADC range, set under program controls by the two parameters ADCZ and ADCR, respectively. The offset can be set between -0.1 volt and + 0.1 volt and the range between - 1.0 volt and -2.0 volt. The normal operating voltages are +0.1 volt and -2.0 volt, respectively, for all detectors. Note that maximum signal is a negative voltage.

The reason for giving a non-zero value to ADCZ is the following: noise spikes in the signal may be of either polarity so that at small signal levels the peak of these spikes may be positive. The ADC gives a zero digital value for any analogue input which is positive with respect to ADCZ; a zero value of ADCZ would thus lead to a 'rectification' of the noise instead of a long-term cancellation of positive and negative noise spikes. The processing circuits provide the facility of digitally subtracting a fixed quantity equal to the analogue offset before the signal is passed to the mass-store in order to increase the capacity of the store. In practice one is nearly always interested in difference signals and the black-level subtract (BLS) facility will only be used rarely. The DC offset does not contribute to the noise in the detector output.

3. Detector Sensitivity

The gain of the video amplifier which precedes the ADC is factory adjusted to give a 2.0 volt output for an output current of the camera tube of 300 nA. This corresponds to the tube manufacturers' specification of a 'peak-white' signal, ie the top end of the linear range of the camera tube transfer curve. The X-ray sensitivity of the FAST detector can be varied in three ways under computer control.

1. By varying the light gain of the image intensifier. This is done by changing the high voltage applied to the intensifier (HVII) in steps between 4.5 kV and 15 kV by specifying a value of HVII between 0 and 7, respectively. The ratio of the gains at at maximum and minimum settings depends on the individual detector and is between ~ 2 and ~ 10 .

2. By varying the gain of the intensifier stage of the SIT TV camera tube. This is done by changing the high voltage applied to the camera tube (HVCA) in

steps between 3 kV and 10 kV by specifying a value of HVCA between 0 and 7, respectively. The gain varies by a factor of about 100 between the minimum and maximum settings.

3. By varying the reference voltage of the ADC and thus defining the range of the latter as $(1 + \text{ADCR}/255)$ volts where ADCR can take any value between 0 and 255.

The three quantities HVII, HVCA and ADCR between them multiplicatively determine the sensitivity of the FAST detector expressed in MSU per X-ray photon. In principle, the sensitivity of a given detector can be varied from a maximum of between 1.25 and 2.5 MSU per keV of photon energy to about 1/500 of this value, but in practice, for precision measurements the choice of values is restricted:

HVII should be chosen so as to minimise the ratio of total electronic noise to sensitivity (see below).

HVCA settings below 3 should be avoided in the interest of maximum stability.

ADCR should normally have a value of 255 for a full-range signal current of 300 nA.

The sensitivity of the detector may be measured with the help of a radioisotope X-ray source (eg Fe-55) masked with a lead plate with a small (0.1 to 1 mm diameter) pin-hole rigidly attached to the source. The number of quanta emerging through the pin-hole is first measured with an X-ray scintillation counter of known efficiency (eg that fitted on a CAD-4 diffractometer, which is nearly 100% efficient). The source is then mounted in front of the FAST detector and the intensity over the entire spot image in MSU is measured by integrating for a known time. A background subtraction should be performed without the source. The sensitivity in MSU per photon for any given radiation can be obtained from that for the Mn $K\alpha$ X-rays from the Fe-55 by multiplying by the ratio of the photon energies, eg 1.36 for Cu $K\alpha$ X-rays. Note that the sensitivity is defined in terms of absorbed photons: it is quite independent of A, the absorbed fraction shown in figure 2.

4. Detector Noise

The crystallographer using the FAST system is primarily interested in the total noise of the detector. This may be determined by carrying out a large number (>20) of integrations in the absence of X-rays and calculating the standard deviation from the mean. Each integration should be over a box size equal to that used for a typical diffraction spot. (10 x 10 pixels). The noise is approximately Poissonian so that the S.D. is proportional to the square root of the integrating time. As a result of spatial variations of the noise the S.D. depends on the selected position in the image; because of spatial correlation between neighbouring

pixels due to a finite point spread function it also depends on box size and box shape. Because of this spatial correlation the SD of individual pixel intensities from the mean of a box after a single integration is smaller, by a factor of at least 2. It is less indicative of detector performance than that obtained from multiple integrations.

The latter type of measurement is, of course, much quicker to carry out and Enraf-Nonius have found it useful for a comparison of different detectors and for checking long-term deterioration.

The individual components of the noise are of interest to the manufacturers and must be known to appreciate their relative importance and the way in which detector performance can be improved, eg. by cooling. These components are

1. Amplifier Noise.
2. TV camera tube dark current shot noise.
3. Statistical fluctuations in the dark emission of the image intensifier.

These three components are the most important. They are Poissonian and add in quadrature. They will be discussed further below.

4. Large-amplitude scintillations, probably due to cosmic radiation and radioactivity in the phosphor and the first fibre-optics face-plate. In practice these are rare and random events, but they may result in a very occasional 'rogue' measurement.

5. Small electrical discharges in the detector, especially under conditions of high humidity, have only been found in one detector, where they could be avoided by operating the camera tube at a lower high voltage.

6. Random electrical interference due to external equipment manifests itself on the live TV image as bright streaks along TV lines and can be avoided by appropriate grounding and shielding or removal of the offending equipment.

7. Synchronised electrical interference resulting in a fixed-pattern noise was formerly quite serious since even a very small synchronous signal is integrated. This is now reduced to negligible proportions by improved grounding and shielding and is almost completely eliminated in difference measurements.

8. Variations in DC offsets, especially of the camera dark current due to temperature changes. These can be removed by operating the detector at a lowered and stabilised temperature (see below).

5. Individual Noise Components

1. Amplifier Noise

This is largely due to thermal noise in the preamplifier. The signal-to-noise ratio of the amplifier should be at least 50 dB, ie the rms noise is about $\ln A \text{ fp}^{-\frac{1}{2}}$

$\text{px}^{-\frac{1}{2}} = 0.22 \text{ MSU fp}^{-\frac{1}{2}} \text{ px}^{-\frac{1}{2}}$ (for ADCR = 255). This noise is Poissonian, so it will increase with the square root of the integration time and of the number of pixels in the integration box. It can be measured by determining the standard deviation from the mean of a large number of integrations for a given number of frame periods, with the scanning beam of the camera disabled. In principle the amplifier noise is proportional to the square root of the absolute temperature so that it cannot be reduced very much by a moderate amount of cooling.

2. Shot Noise in the SIT Tube Dark Current

The rms shot noise for a dark current of I_{dk} amps is $\sqrt{2eI_{\text{dk}}B}$, where

e = electronic charge = 1.6×10^{-19} C.

I_{dk} = DC dark current ($\sim 12 \times 10^{-9}$ A at 35°C).

B = Bandwidth = 5×10^6 HZ.

This gives a shot noise of $0.14 \text{ nA} = 0.03 \text{ MSU fp}^{-\frac{1}{2}} \text{ px}^{-\frac{1}{2}}$. In practice the dark current may consist of multi-electron bunches resulting in a much larger shot noise. (A value of $0.24 \text{ MSU fp}^{-\frac{1}{2}} \text{ px}^{-\frac{1}{2}}$ has been observed for one detector). The shot noise and the amplifier noise add in quadrature so that the shot noise may be measured by determining the total noise with the scanning beam enabled but the image intensifier high voltage off and by then correcting for the amplifier noise. The dark current is strongly dependent upon the temperature, (see section 8, below), and hence the shot noise is reduced by cooling.

3. Scintillations or 'Fleas'

These are due to dark emission in the image intensifier. In the live monitor image these manifest themselves as bright spots of light, usually more intense than X-ray photon scintillations, dancing about on the screen in the manner of a flea circus. The scintillation, or flea intensity, is different for individual image intensifiers, varies with position on the detector, being highest near the centre, and is non-linearly dependent on HVII. The intensity may be measured by carrying out a background-subtracted integration for a period of several minutes in the absence of X-rays. The 'Background' should really be measured with the Image Intensifier high voltage fully switched off, but in practice the scintillation intensity at HVII=0 is so low that it can be used for this purpose. A knowledge of the total flea intensity in a given time is useful as a test of the image intensifier. The statistical performance of the detector is determined only by the fluctuations in this intensity. The latter can be derived from the total noise variance, ie the square of the standard deviation, measured as described above, by subtracting the variance due to dark current and amplifier shot noise, since the individual standard deviations add in quadrature. The mean amplitude of a scintillation appears to vary more rapidly with HVII than does the mean amplitude

of an X-ray photon flash. At maximum HVII the average scintillation produces a signal equivalent to that from a photon with an energy of about 25 keV, at lower HVII of only about 5 keV. It will be apparent that the scintillations have a smaller relative effect when working with harder X-rays.

Typical noise component values are summarised in the table.

Noise Source	Rms Fluct.		Units $fp^{-\frac{1}{2}}$ $px^{-\frac{1}{2}}$
Amplifier	0.20		MSU
Scintillations at HVII 4	0.10		MSU
Scintillations at HVII 7	0.62		MSU
Dark Current at Temp.	35°C	10°C	MSU
	0.32	0.11	
Total (Max. Sensit.)	0.73	0.66	MSU
	0.034	0.031	Number CuK qu.
	0.016	0.014	Number MoK qu.
Total (Opt. Setting)	0.39	0.25	MSU
	0.022	0.014	Number CuK qu.
	0.010	0.006	Number MoK qu.

6. Optimum Operating Conditions

The parameter which determines the detector performance is the ratio of the total noise standard deviation (in $MSU fp^{-\frac{1}{2}} px^{-\frac{1}{2}}$) to the sensitivity in MSU per X-ray photon. This ratio is the S.D. expressed in equivalent X-ray photons $fp^{-\frac{1}{2}} px^{-\frac{1}{2}}$ and is denoted by σ . Its numerical value is inversely proportional to the photon energy. As discussed above, σ varies over the image surface but we can take as a comparison standard the value of σ measured at $(y,z) = (200,200)$ using a 10×10 box size. For optimum performance HVCA should be as high as possible while avoiding any discharges. σ should then be measured as a function of HVII and the operating point chosen so as to minimise σ . Curves of σ as a function of HVII for one particular detector are shown in fig 3 for $CuK\alpha$ and $MoK\alpha$ X-rays; it can be expected that the shapes of the curves, and the values of σ , will be somewhat different for different detectors.

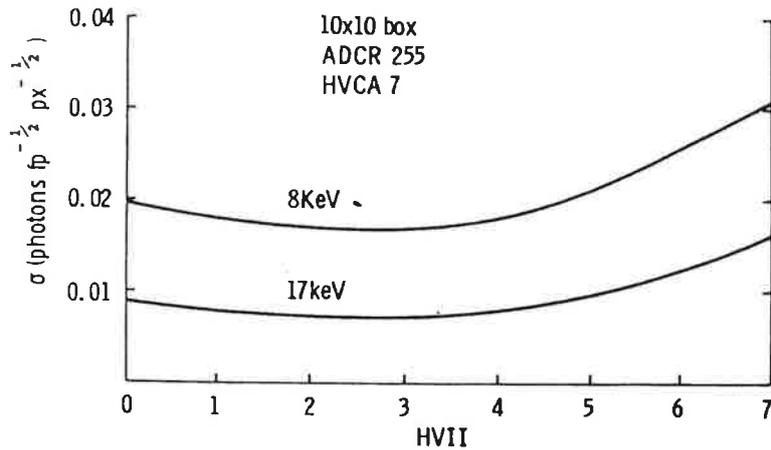


Fig.3

7. Detective Quantum Efficiency

ϵ , the detective quantum efficiency (DQE) of a photon detector can be expressed as

$$\epsilon = (Rn)^{-1},$$

where R is the relative variance in the output signal produced by n photons. An analogue detector with a DQE of ϵ thus behaves like a perfect Poissonian detector, or ideal counter, which responds to a fraction ϵ of the incident photons.

For the FAST detector, assuming that the photon variance and the noise variance are additive, we can write

$$R = (1 + \sigma^2/An)/An$$

and $\epsilon = A/(1 + \sigma^2/An)$

where n is expressed in photons $\text{fp}^{-1} \text{px}^{-1}$ and A is again the fraction of absorbed photons. ϵ thus varies with the count rate; it also varies somewhat with box size and shape and with position on the detector. We shall take ϵ for a 10 x 10 pixel box at (200,200), ie near but not at the centre of the detector, at a count rate

which corresponds to the average background count rate for a typical protein crystal, as the most descriptive single parameter for specifying the performance of the detector.

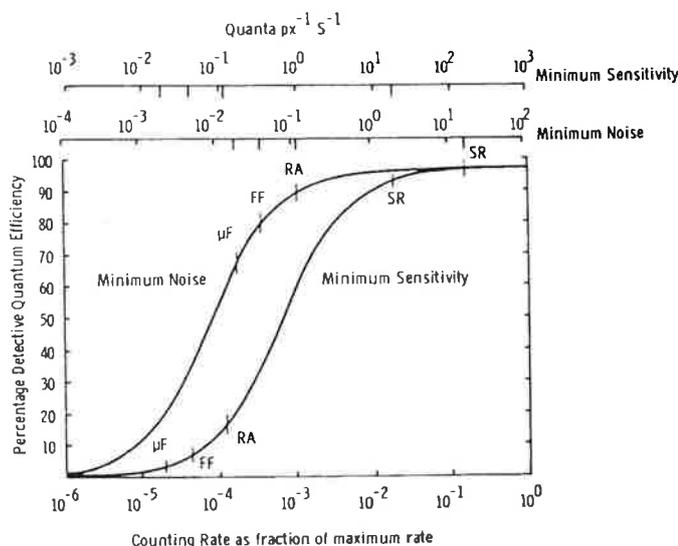


Fig.4

Figure 4 shows ϵ for $\text{CuK}\alpha$ X-rays as a function of count rate as measured for one particular FAST detector when operated at minimum sensitivity and when operated at the optimum values of HVII and HVCA. The mean background count rates are marked for a typical protein crystal with an 80\AA unit cell for four different X-ray sources. (μF = 800 watt microfocus tube; FF = 1.5 kW fine-focus tube; RA = 5 kW rotating anode tube; SR = Storage ring = $160 \times$ RA). Note that the rate for SR is simply scaled from X-ray tube measurements; in practice the background would be relatively lower than with conventional sources. These figures show that from the point of view of noise and counting statistics the FAST detector behaves almost like an ideal detector for the more intense sources; they also show that for more weakly diffracting specimens, (smaller crystal size or larger unit cells), intense rotating-anode sources are essential.

8. Effect of temperature on detector performance

The dark current of the SIT tube is reduced by cooling; typically, it is about 12 nA ($\sim 2.6 \text{ MSU fp}^{-1} \text{ px}^{-1}$) at 35°C and less than 1 nA ($\sim 0.2 \text{ MSU fp}^{-1} \text{ px}^{-1}$) at 0°C . (With the standard fan provided the SIT tube target has an equilibrium temperature of about 35°C when the room temperature is 20°C ; this equilibrium is established about 4 hours after switching on the detector). The statistical fluctuation in the dark current ('shot noise') is proportional to the square root of the current: the standard deviation due to this component can, therefore, be expected to vary by a factor 4 between 35° and 0° . Since this component is only a small part of the total noise the effect is small. However, it is essential for the DC offset to remain constant during the integration of one frame or during the scanning of any given reflexion, otherwise large errors may be introduced in the background correction. It is, therefore, more important for the temperature to be stable than for it to be particularly low, although at a lower temperature, and thus a lower dark current, the demands on stability can, of course, be relaxed. There is some evidence that the amplitude of the TV scan waveforms, and thus the size of the detector image change slightly with temperature. Fortunately, the large thermal capacity of the detector ensures that temperature fluctuations are slow, even in the absence of specific thermostatic control.

The amplifier thermal noise is proportional to the square root of the absolute temperature, so that this type of noise can only be reduced by a few percent by the amount of cooling which is possible in practice.

9. Maximum Counting Rate and Dynamic Range

The maximum rate at which the television detector can operate without losses is that at which the ADC produces an output of $255 \text{ ADU} = 63 \text{ MSU px}^{-1}$ in one frame period. This rate, therefore, is inversely proportional to the chosen sensitivity. We have seen above that the minimum sensitivity useable in practice is achieved for HVII and HVCA settings of 0 and 2, respectively where for a typical FAST detector the sensitivity is about $1.75 \text{ MSU (8keV photon)}^{-1}$. The maximum counting rate, therefore, is $63/1.75 = 36 \text{ photons px}^{-1} \text{ fp}^{-1} = 900 \text{ photons px}^{-1} \text{ s}^{-1}$. It is correspondingly lower for higher-energy X-ray photons. Unlike, for example, a multi-wire chamber, the television detector could function at the maximum rate simultaneously in all of its 262000 pixels. The limitations are local and not global count rates. In addition, overflows of the ADC affect only those pixels where they occur: it is perfectly possible to collect accurate single-crystal data even though the detector saturates on a few of the strongest reflexions. In the interests of achieving the highest detective quantum efficiency for the weakest

reflexions the detector should always be operated at the highest sensitivity which avoids saturation on the strongest reflexions which it is desired to measure. In figure 4 the DQE is shown as a function of the count rate expressed as a fraction of the maximum count rate at a) the minimum sensitivity and b) the minimum noise settings. The dynamic range of the detector may be defined as the range of counting rates over which the DQE has an acceptable value: we shall arbitrarily define this minimum value as 20%. On this basis the dynamic range and the maximum counting rates in the two gain settings are 5000:1 and 900 photons $s^{-1} px^{-1}$, respectively, and 35000:1 and 105 photons $s^{-1} px^{-1}$, respectively.

It is of interest to note that the maximum count rate is well matched to the maximum data transfer rate permitted by the digital circuitry of the FAST diffractometer controlled by a Microvax II computer. For a typical protein crystal a maximum peak count rate of 900 photons $px^{-1} s^{-1}$ for a strong reflexion corresponds to an integrated count rate of about 5000 $qu^{-1} s^{-1}$ for an average reflexion (Arndt & Gilmore, 1979). The time needed for an average reflexion to pass through the Ewald sphere is therefore 2 sec to achieve 1% counting statistics. Assuming a reflecting range of 0.2° the rotational speed of the crystal should, therefore, be $0.1^\circ s^{-1}$ or $6^\circ min^{-1}$. Using the Phase III data collection software system written by D. J. Thomas the maximum speed at which data can be collected from a Haemoglobin crystal (P_{21} $a=63\text{\AA}$ $b = 83\text{\AA}$ $c = 54\text{\AA}$ $\beta = 99^\circ$) is about 30 reflexions s^{-1} at a crystal rotation rate of 4° to $6^\circ min^{-1}$. The primary beam intensity required for the above reflexion intensities is about 200 times that obtainable from a 5 kW rotating-anode tube, ie about that available at the protein crystallography beam lines of the SRS.

10. Spatial Distortion

The output image of the television detector is distorted by imperfections in the electron-optical and the light-optical components and in the TV camera raster scan. These distortions are very stable with operation of the detector under constant conditions so that an infrequent distortion calibration is all that is necessary. However, they vary slightly with HVII and HVCA; because of local variations in the external magnetic field they vary with position of the detector on the detector arm and with the swing angle of this arm. In addition, they may vary with temperature, an additional reason why the temperature of the detector should be stabilised. The distortion calibrations must, therefore, be carried out at the position at which the detector is used for data collection and at operational voltages and temperature. Calibration is carried out with the help of an accurate shadow mask fixed in front of the detector. A material which

fluoresces strongly in the incident X-ray beam, ie one which has an excitation potential just below the photon energy of this beam, should be mounted in the crystal position. Suitable materials are compounds of iron for $\text{CuK}\alpha$ radiation and of strontium for $\text{MoK}\alpha$ radiation. The positional variation of the distortion can be greatly reduced by demagnetising those parts of the detector arm which are made of hard magnetic material (the rails), but the variation can never be totally eliminated and great care should always be taken with the calibration as it affects not only the positions of the reflexions but also the uniformity of response of the detector, because of variations in the pixel area.

11. Non-Uniformity of Response.

The measurements made with the detector must be corrected for spatial variations of its response which may amount to a factor of 4 between the centre and the corners of the image. The major cause of the non-uniformity is a variation in the pixel size as a result of spatial distortions; accordingly calibration of this effect must also be made at the operational position of the detector. A fluorescent material is a suitable isotropic source for this purpose also, provided that the actual positions of Debye-Scherrer rings and of the backstop shadow are avoided. Corrections must, of course, be made for the variation of the path length with position which causes an inverse-square law variation of intensity and a change in the air absorption. Oblique-incidence corrections are not necessary as they affect calibration and data equally for the same detector position.

Polynomial correction functions have, recently, been found preferable to look-up tables for both types of calibration.

Analysis of Reflection Data collected using the
Phase III Software for FAST area detector.

A.C. Bloomer, G. Fermi, D.J. Thomas

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH.

The phase III software developed for FAST uses the detector in 'diffractometer' mode rather than in 'film' mode such that the response of the camera is transferred from the mass store to computer memory (and hence eventually to disk) for only those pixels lying within pre-specified windows. The position of the window centre depends upon the crystal lattice and its orientation and the size of the windows depends upon the characteristics of the incident X-rays and the crystal mosaicity. Appropriate positioning of the windows thus requires both accurate knowledge of these parameters and also careful calibration of the geometry of the instrument with its spatial distortion characteristics (to obtain the appropriate transformations between pixels and millimetres).

Any errors in the window determination may prove unrecoverable once the data are recorded. However, our experience has shown that prior knowledge of the other major instrumental calibration - for non-uniformity of pixel response - is not as important. Present imprecision in this calibration is such that, with the LMB detector, it is better to regard all active pixels as having the same response than to attempt a pixel by pixel correction for any local non-uniformity of response. This conclusion is dependent upon the phase III prediction algorithm admitting for measurement only those reflections whose window falls totally within the active area of the detector (i.e. excluding the corners where the beta-lights are recorded and the backstop shadow. Whenever the backstop support bracket does not fall within the blind region [i.e. with the standard backstop mounting, whenever $Kappa \neq 0$] a simple rectangular exclusion is used to eliminate from the prediction any reflections whose window enters the support shadow).

Improvements in the determination of this pixel response calibration are in hand but, in the meantime, data are corrected only for a global non-uniformity. With the detector in the symmetrical setting ($\theta = 0$) this shows up as a small negative relative temperature factor between the uncorrected FAST data and a standard data set. A slightly better result is achieved by determining a smoothly varying correction from internal redundancy within the FAST data and/or a standard dataset obtained by other means. Since this correction is a function of the instrumental response, and not of the particular crystal used, a correction determined from one standard dataset can also be usefully applied to data for which no such standard sets are available.

The attached table shows the results of scaling these data from one crystal in five different ways which each gave some improvement in both the internal consistency within the data and also the absolute agreement with an external standard. The overall agreement between Bijvoet pairs after data reduction is 3.0% on intensity compared with an expected real difference of about 1%.

The data represent 80% of the 3.6 Å reflections from a hybrid form ($\alpha\text{Fe}, \beta\text{Co}$) of deoxyhaemoglobin ($P2_1$, a , b , c = 63, 83, 54 β = 99) collected at a continuous rate of 22 reflections/minute overall (crystal rotated at constant speed of 0.2 deg./min.) on a sealed tube running at 1400 watts. The data were collected in four separate runs due to such problems as disk storage at that time, giving an overall average redundancy of 3 for each member of a Bijvoet pair. The four runs comprise overlapping rotation ranges (Φ = 0-85, 0-52, 40-120 and 116-180) and are treated as four separate BATCHES (i.e. as if four "films") in the LCF

file. All statistics are given with no measurements rejected. Any reflection which is measured is thus considered a 'good spot'. For these data, reflection intensities were determined from the 3-D boxes by simple peak minus background subtraction calculated from the 1-D profile along the coordinate of the angle of rotation. Thus all pixels within the reflection window on the detector are used equally in calculating both peak and background estimates.

Considering the five columns of the table in order, the four data runs were merged as follows:

(1) without any scaling.

The entire file of 31871 observations of 5723 separate reflections was admitted directly to AGROVATA without any prior correction. Considering the speed of data collection relative to the X-ray power, the overall symmetry R-factor of about 7% was surprisingly good in the complete absence of any scaling, or correction for absorption or crystal illuminated volume.

(2) scaling as determined by ROTAVATA.

Only scale factors were refined with no temperature factors, as a single overall BTEMP was felt to be more appropriate than 4 separate values (although these could have given a crude correction for crystal degradation). The effect of the 3 additional scaling parameters used here was to give a small improvement in R_{sym} reducing it by under 1% but no change in the agreement with the standard dataset used as an external comparison.

(3) scaling by primary (incident) beam correction.

This takes into account both the change in illuminated volume of the crystal with the rotation angle PHI and also the primary beam component of crystal absorption. The correction was applied by a new program (BEAMCOR) which reads the LCF file of background-subtracted intensities and writes a new LCF file of corrected intensities for subsequent input to AGROVATA in parallel with a set of scale factors set to unity for every batch (i.e. as in (1) above).

The correction was determined by a least squares refinement of the coefficients of an appropriate Fourier series to give a smoothly varying function of PHI. Repeated measurements of a reflection typically occur at different values of PHI such that the differences between these observed intensities contain the necessary information for determining the primary beam correction.

This calculation used five Fourier terms (i.e. up to order ± 2) for each batch giving a total of 20 refined parameters. By treating the four batches separately it was possible to determine the correction as a function of TIME rather than angle directly (within a single run, these are directly proportional) to allow for the possibility of time-dependent decay. However, the program has the facility for determining the primary beam correction as a one-dimensional (as here) or complete two-dimensional Fourier series for future use when different scanning axes are used on FAST for parts of the dataset from a single crystal (e.g. torus and cusp) or if the axis is a complex function of two or more of the instrumental axes, possibly changing with time. Eigen value filtering ensures that the refinement proceeds appropriately and the facility for modifying the variances of the observations allows compatibility between the weighting employed in this least squares refinement procedure and the subsequent statistical analysis performed in AGROVATA.

This correction gave a small improvement 0.4% in both R_{sym} and R_{merge} . When it was tried as a two-dimensional correction (8 parameters with filtering instead of 20) the results were essentially the same.

(4) scaling also by a secondary beam correction.

This takes into account not only the factors considered in (3) as being functions of the primary beam direction but also those which are functions of the secondary (diffracted) beam direction. This mainly includes the diffracted beam component of the crystal absorption, assuming that this effect is separable into the two components.

The correction was determined with the same program, by specifying that a secondary beam correction be refined as a two dimensional Fourier series, simultaneously with the primary beam correction as above. A further 24 parameters are thus required to give coefficients up to order ± 2 in both dimensions ($5 \times 5 - 1$, excluding the overall constant term).

This correction gave still further improvement in the overall R_{sym} with a decrease of 1% due to the secondary beam component. There was a far larger improvement, however, in the external agreement with a reference data set where R_{merge} on amplitudes decreased by 2.4% on application of this correction for absorption by the crystal.

(5) scaling also by detector response correction.

This considers all the corrections included in (4) as being functions of the primary and secondary beam directions and in addition also those which are functions of the pixel coordinates on the detector for each reflection. The coordinates used are those of the centroid of the reflection peak within the two dimensions (Y,Z) of the detector after summing all the successive angular slots for each pixel independently. A further 24 parameters were refined to give two dimensional Fourier coefficients over the area of the detector.

In this case, we included also a reference set of deoxy haemoglobin intensities (from a rotation camera film dataset collected at LURE) which were used as a scaling set for the detector correction but ignored for the primary and secondary beam corrections. In principle a scaling set is not strictly necessary, as sufficient information is present either from multiple observations in two or more quadrants of the detector, and/or from additional information for this correction could have been obtained by the use of batches of observations measured in different geometries (e.g. rotating the crystal about the incident beam and/or varying the crystal to detector distance).

The result of this correction gave, as expected, no improvement in the internal R_{sym} and only very slight improvement in the external R_{merge} . The relative temperature factor between the FAST and film data now appears as zero because it has been accounted for earlier within the detector correction of the least squares refinement program.

Further improvements in the quality of FAST data will clearly come with the use of profile fitting, whether in one, two or especially three dimensions. This is expected to improve the weak reflections particularly, which are those for which the present data do not show good agreement with the film data.

Analysis of Phase III Data Collected on FAST at Cambridge(LMB)

e.g. Deoxy Haemoglobin (Fe/Co hybrid) to 3.6A resolution
collected on a sealed tube (1400w) at a rate of 22 reflections per min.

Column .5. ALLunscaled for 4 batches of partially overlapping rotation ranges
Column .6. ALLscaled by ROTAVATA as 4 separate batches
Column .7. ALLscaled by primary beam correction for crystal illuminated volume
Column .8. ALLscaled as .7. plus secondary beam correction for absorption
Column .9. ALLscaled as .8. plus detector non-uniformity of response correction
Columns 7,8,9 corrections were all applied as Fourier coefficients determined by
least squares refinement of smoothly varying functions over angles and/or pixels.

=====

AGROVATA : after scaling the data as described in text. Nobs = 31871 Nhkl = 5723
Rsym on INTENSITIES is shown for ranges chosen such that I_{mean} lies
in the middle of the third range. For these ranges, the two members
of a Bijvoet pair are treated separately (by setting KANOM = 3).
The overall Rsym is shown (I_o) for all data (by setting KANOM = 0)
ignoring any anomalous differences - expected to be about 1% .

Column No.	.5.	.6.	.7.	.8.	.9.	
Nparams	1	4	20	44	68	
Mean I	1005	1038	1001	936	878	
I<400	.169	.173	.183	.172	.169	
800	.082	.078	.080	.071	.069	
1200	.065	.061	.060	.048	.046	
1600	.058	.049	.049	.037	.037	
2000	.055	.047	.045	.033	.032	
2400	.056	.047	.044	.031	.031	
2800	.049	.039	.043	.029	.028	
3200	.054	.045	.037	.027	.028	
3600	.049	.039	.040	.025	.025	
.....	.044	.035	.034	.023	.022	
Overall						
(I+/I-)	.066	.058	.055	.045	.045	*****
Ignoring anom.						
(I _o)	.074	.068	.064	.053	.053	

=====

AGROVATA : after re-running the program to compare I+ and I- from the output
above, (setting KANOM = 0) the overall Rsym (I_o) on INTENSITIES is
shown for all reflections for which both I+ and I- are available.
Nobs = 9638 Nhkl = 4819

Rsym						
(I _o)	.038	.038	.035	.030	.030	*****

=====

COMPARISON with the standard Deoxy Haemoglobin 1.9A film data;
scaled on amplitudes for the 5833 reflections common to both files
using KSCALE & BTEMP as determined by the program RSTATS.
Rmerge is shown for amplitude binss where meanF = 353.

KSCALE:	.080	.080	.075	.079	.091
BTEMP:	-3.8	-3.8	-5.4	-4.5	+0.01
F<140	.371	.371	.372	.365	.364
280	.093	.094	.093	.079	.078
420	.067	.068	.068	.048	.047
560	.062	.064	.064	.038	.036
700	.058	.060	.061	.030	.028
840	.056	.058	.058	.028	.026
980	.056	.059	.058	.022	.019
1120	.044	.047	.049	.022	.021
.....	.052	.050	.060	.014	.014
Overall					
Rmerge	.083	.083	.079	.055	.054

=====

Progress report on the FAST diffractometer and a new method
=====

of response calibration of an area detector using
=====

single-crystal diffraction intensities.
=====

by Glaucius Oliva and Ian J. Tickle,
Department of Crystallography, Birkbeck College, London.

Introduction.
=====

Delivery of an Enraf-Nonius FAST diffractometer was taken by this department on November 17th, 1986, and although the machine was installed and collecting protein diffraction data of acceptable quality ($R_{sym} \sim 6\%$, see Appendix) within 4 days, it soon became apparent that there were unforeseen problems with the calibration of the spatial non-uniformity of response of the area detector using the currently available software (Enraf-Nonius: "PHASE 1").

We were particularly concerned with this because an inadequate calibration of the detector would obviously militate against measurement of data of the highest quality, as required in particular for anomalous scattering studies, for analysis of anisotropic thermal parameters, and for comparisons and differences between FAST intensity data and data collected by other means such as single-counter diffractometer and photographic film. There is no doubt that the precision necessary for these analyses can in principle be attained with this type of detector.

What also became apparent was that the currently recommended method of non-uniformity calibration (by use of an uncollimated flood of X-radiation at the maximum source-detector separation) probably made things worse by introducing systematic errors not present in the raw uncorrected image. Some time and effort was spent in attempting to improve this method, but the results obtained did not come up to expectation. However, for the sake of completeness of this report, these abortive attempts will be described in detail, and then a new method which uses diffracted intensities from a single crystal will be described. This has so far been tested only on simulated data, owing to pressure of work on the machine and a number of hardware problems, but the results obtained to date look very promising.

FAST detector - sources of response non-uniformity.
=====

The non-uniformity of response of the FAST television detector arises from a number of sources, the ones identified by other workers in this field being vignetting in the fibre optics coupling and variations in the photocathodes of the image and camera tubes (Arndt, 1985); and pin-cushion distortions in the image intensifier and edge-effects due to point-spread (Kalata, 1985; Thomas, 1987).

The majority of these effects are directly related to the spatial distortion which would be relatively easily calibrated by means of a precision machined pin-hole grid fitted directly on the detector surface, if it were not for the fact that the spatial distortion is magnetic-field dependent, and the rails on which the detector, and its associated electronics, runs become permanently magnetised by the very act of moving the detector during the calibration procedure. However, this is a side issue as the problem can be overcome by demagnetising the rails periodically and employing a procedure which calibrates at the measuring position.

The variations in response mentioned above amount to about 50% of the maximum and are all smoothly varying over the detector surface; in addition there are local irregular variations of about 1.5% due to phosphor graininess and a small number of blemishes ("black" and "white" bad pixels) (Arndt, 1985).

Detector calibration - 1st method.
=====

The calibration technique that we initially proposed was partly based on the manufacturer's recommended one and partly as suggested by D. Thomas (1987). The first step is to use the uncollimated flood beam at 5 degree tube take-off angle, the maximum source to detector distance (~ 121 cm.), minimum kV/mA settings and minimum detector sensitivity : HVCA=0, HVII=1. We were somewhat concerned that the instruction manual warned us not to set the camera voltage HVCA below setting 3 due to stability problems, but we had no choice if we were to avoid overflowing the analogue-digital converter (ADC), and in any case no problems with stability were apparent.

We had previously checked the vertical uniformity of the beam by varying the inclination of the X-ray tube about the horizontal axis; this turned out to be surprisingly uniform and no correction was necessary. However in the horizontal direction the intensity naturally varies with position since the detector is looking at a line focus. We therefore obtained a calibration curve of intensity versus take-off angle, which was converted into correction factor versus horizontal detector coordinate, Z (in cm.).

An image was made with the shutter open and corrected for the dark current. "Black" bad pixels were identified as a specified fraction (0.2) of the maximum of this image. A dark image was also made, to identify "white" bad pixels and beta-light areas. Spatial distortion polynomial coefficients were produced according to the instruction manual. These coefficients allow interconversion of the coordinates (in cm.) of a point on the detector faceplate and the coordinates (in pixels) of the same point mapped into the mass-store.

Then the detector was moved to the normal measuring position (5.5 cm from the centre of goniostat rotations and at zero inclination angle), the normal collimator replaced, the X-ray generator set to normal values of kV and mA for the tube in use, and the detector voltages set to normal values (HVCA=7, HVII=5). A piece of standard magnetic computer tape folded once was mounted on the goniostat and set at the centre of rotations and perpendicular to the beam, and the production of the spatial distortion polynomial coefficients repeated using the powder diffraction and incoherent scattering from the tape.

The first correction to be made using this information is to the spatial distortion coefficients for the image at the measuring position to compensate for the fact that the calibration grid is spaced somewhat (~ 3mm.) from the detector faceplate. The square root of the ratio of the areas bounded by 4 pixels in the images at the calibration and measuring positions gives a linear scale factor. The pixels used were (128,128), (128,384), (384,128) & (384,384), the whole image being 512 by 512 pixels; the area of a quadrilateral with vertices $\underline{V1}$, $\underline{V2}$, $\underline{V3}$, $\underline{V4}$ in the same order is most conveniently computed as $|\frac{(\underline{V4}-\underline{V1}) \times (\underline{V3}-\underline{V2})}{2}|$.

Then the following calculation is performed to obtain a value of the response for each pixel in the image at the measuring position: the detector coordinates in cm. are computed together with the area (=A) of the pixel mapped onto the detector. The coordinates and area are converted back to pixel units (area=B) at the calibration position. The new pixel coordinates are used either to look up the value (C) in the calibration image, or if the pixel is bad to return a bad pixel flag. The take-off correction (T) is interpolated from the take-off calibration curve using the detector Z coordinate. The value of the response is then $T.B.C/A$. The ratio of the areas A/B is just the area of a pixel in the calibration image when mapped onto the detector.

The effect of applying this response calibration to the diffraction data was not encouraging; the same reflection (obtained by rotating the crystal around the beam, not symmetry-equivalent) measured at different places on the detector produced significantly differing integrated intensities, varying by as much as 20%. The only explanation for the discrepancy appears to be that in the calibration position at the maximum source-detector separation the spectral distribution is heavily

modulated by air absorption, and this effect is combined with a wavelength-dependent detector response (Arndt, 1987). (The transmission factor for 115cm of air is about 0.26 for 1.54A radiation and 0.85 for 0.71A.)

One alternative method of calibration that has been suggested is the use of a computer-controlled X,Y stage to position accurately a collimated Fe-55 X-ray source anywhere over the detector surface (Hamlin, 1985). We are currently investigating the acquisition of a "dial source", which consists of a gamma-emitter behind a rotatable dial of different metal foils, producing their characteristic X-ray fluorescence, but it is not clear how feasible it is to calibrate the detector at its measuring position, which appears to be essential if we are to avoid magnetic distortion problems.

Detector calibration - 2nd method.

=====

The basis of the new method is to use the differences in intensity observed for the same reflection at different places on the detector to model the smoothly varying components of the non-uniformity with a low order (up to 5) bivariate power series in the mass-store coordinates, retaining the part of the previously described procedure which calibrates the irregular pixel-to-pixel variation. The data consists of multiple measurements of the same set of integrated intensities corrected for pixel-to-pixel non-uniformity, and obtained by rotating the crystal around the X-ray beam axis and also by moving the detector to scale the image up or down. It is assumed that this latter movement of the detector will not be sufficiently great as to introduce errors due to changes in spatial distortion.

According to the usual least-squares procedure we require to minimise the weighted sum of squares of differences :

$$S = \sum_i \sum_j W_{ij} \cdot (M_{ij} - G_{ij} \cdot I_i)^2 \quad \dots (1)$$

where i indexes the list of different reflections,
 j indexes the measurements of the same reflection,
 W_{ij} is the weight attached to the measured intensity M_{ij} ,
 I_i is the true (but unknown) value of the intensity,
 G_{ij} is the mean detector gain (response) for the measurement.

Note that it is not necessary to have all measurements of a particular reflection; unmeasured values are assigned zero weight.

The weight W_{ij} should be the reciprocal of the variance of M_{ij} ; the processing program makes an estimate of this assuming Poissonian statistics but does not take detector gain into account, hence if the estimate of the weight obtained from this is now W_{ij} , the true weight is W_{ij}/G_{ij} , and equation (1) has to be replaced by :

$$S = \sum_i \sum_j (W_{ij}/G_{ij}) \cdot (M_{ij} - G_{ij} \cdot I_i)^2 \quad \dots (2)$$

The gain function G_{ij} is modelled as a bivariate power series :

$$G_{ij} = \sum_{k=0}^N \sum_{l=0}^k C_{kl} \cdot Y_{ij}^k \cdot Z_{ij}^l \quad \dots (3)$$

where N is a small number (4 or 5), the C_{kl} are the polynomial coefficients to be determined, and (Y_{ij}, Z_{ij}) are the mass-store coordinates. S in (2) is minimised by differentiating with respect to the unknown I_i and the shifts ΔC_{kl} in the C_{kl} .

$$\text{Hence } I_i = \frac{\sum_j W_{ij} \cdot M_{ij}}{\sum_j W_{ij} \cdot G_{ij}} \quad \dots (4)$$

and the observational equations for the unknown ΔC_{kl} (call these P_m) are :

$$\begin{aligned} \sum_m \sqrt{(W_{ij}/G_{ij})} \cdot I_i \cdot (Y_{ij} \cdot Z_{ij} - G_{ij} \cdot \frac{(\sum_{j'} W_{ij'} \cdot Y_{ij'}^k \cdot Z_{ij'}^l)}{(\sum_{j'} W_{ij'} \cdot G_{ij'})}) \cdot P_m \\ = \sqrt{(W_{ij}/G_{ij})} \cdot (M_{ij} - G_{ij} \cdot I_i) \quad \dots (5) \end{aligned}$$

The problem is actually very similar to the inter-film layer scaling problem, the equations for which were set out by Hamilton, Rollett & Sparks (1965), with the modifications that the scale factors G_{ij} and the weights W_{ij} become functions of the coefficients C_{kl} . Because multiplication of all C_{kl} and hence all G_{ij} by a constant merely changes S by the same constant factor, the normal matrix obtained from equations (5) is singular. The normal equations are therefore solved by eigenvalue filtering, as suggested by Fox & Holmes (1966). There should be precisely 1 zero eigenvalue; in practice due to rounding errors it is necessary to pre-condition the normal matrix by dividing element A_{mn} by $\sqrt{A_{mm} \cdot A_{nn}}$, and to set to zero the eigenvalue which is insignificant compared with the largest. The equations are solved iteratively from initial estimates of the C_{kl} (in practice initial $C_{00} = 1$, else $C_{kl} = 0$, i.e. constant initial G_{ij}).

Results and conclusions. =====

Simulations were run using polynomial coefficients which approximated the response variation previously obtained. Mass-store coordinates and intensities for 100 "reflections" were produced with a random-number generator and Poissonian errors

introduced into the intensities. Measurements at either 4 or 8 values of the rotation angle around the beam and for each value 2 positions of the detector with distance ratios in the range 1.1 to 2 were simulated. Refinement of the coefficients from the initial values previously stated proceeded by first iterating with only the zero and first order coefficients to convergence (small maximum absolute shift to e.s.d. ratio), then adding successively higher order coefficients, iterating to convergence with each new order. This procedure is necessary because the terms in a power series do not form an orthogonal set of functions.

The overall appearance of the gain function used to calculate the intensities was reproduced in all cases; however simulations using 8 values of the rotation angle were much better than those using 4, and the optimum value of the distance ratio was found to be about 1.4; smaller values presumably lead to higher correlations, larger values reduce the number of data by scaling a proportion of the reflections off the detector.

Acknowledgement.

=====

We thank Dr. David Moss for useful discussions about parameter-dependent weighting in non-linear least-squares optimisation.

Appendix - summary of data collected (21-Nov-86 to 24-Mar-87).

=====

Protein	Dmin	No of Reflections	Exposure time	Rsym
Endothia pepsin (native)	3.0A	25000	31h	6.0
" "				
(+ BW624C inhibitor)	2.0A	60000	70h	8.0
Serum amyloid P-component (native)	3.5A	46000	36h	5.5
" "				
(Th(NO3)4 derivative)	4.0A	32000	31h	6.7
	2.8A	100000	40h	-

References.

=====

- ARNDT, U.W. (1985). Meth. Enzymol. 114, 472-485.
- ARNDT, U.W. (1987). Proc. FAST users' meeting, Enraf-Nonius, Delft.
- FOX, G.C. & HOLMES, K.C. (1966). Acta Cryst. 20, 886-891.
- HAMILTON, W.C., ROLLETT, J.S. & SPARKS, R.A. (1965). Acta Cryst. 18, 129-130.
- HAMLIN, R.C. (1985). Meth. Enzymol. 114, 416-452.
- KALATA, K. (1985). *ibid.* 486-510.
- THOMAS, D.J. (1987). Proc. FAST users' meeting, Enraf-Nonius, Delft.

A NOTE ON THE PERFORMANCE OF THE XENTRONICS/NICOLET AREA DETECTOR
SYSTEM INSTALLED AT YORK UNIVERSITY

Zygmunt S. Derewenda

Chemistry Department, University of York, Heslington, York, YO1 5DD

The Xentronics/Nicolet area detector system installed in the University of York consists of the Xentronics Imaging Proportional Counter, sealed-tube X-ray generator, 3-axis goniostat, PCS computer dedicated to the IPC and the goniostat, an Ethernet link, and a Microvax II computer. It was installed in November and commissioned between November and March 1st. Since March 1st several data sets were collected and processed (in addition to those used for diagnostics purposes). They have all served as a basis for a brief analysis of the performance of the detector, presented here.

According to the manufacturer there are two types of distortions occurring in the system: a high frequency spatial distortion, smoothed out on-line by pixel-mapping procedure; and global spatial distortion, handled separately by the data-reducing software with the use of a calibrating brass-plate image. No photon-response calibration is done on the assumption that the response is uniform and no photons are lost. No specification is provided by the producer and therefore it is impossible to assess the quality of the particular detector without diagnostic data-collection.

During the commissioning period we have found a number of hardware and software problems, which significantly affected the quality of the X-ray data. Three of these problems were particularly serious.

1. Time instability.

We have developed a simple diagnostics software package which enabled us to compare flood-field images. In simple terms the method amounted to re-scaling two frames and dividing one frame by the other, pixel by pixel. The resulting information was then converted into data-like frame format to allow viewing on the system's graphics monitor. The time instability was manifested by sharp vertical stripes indicating drastic shifts in the mapping of incident photons. The problem was traced mostly to vibration transmitted from the omega motor, and steps were taken to minimize the effect.

2. Point-spread function variations.

It was found, the the profiles of spots varied considerably across the detector, and that the total intensity of a spot influenced its shape as it appeared on the frame image. In some areas of the detector the spots are circular and fit into the integrating box of 11*11 pixels (as used by the Harvard SCAN program); in some areas, however, the spots have elongated profiles and are 16 - 20 pixels wide. This effect gives rise to systematic errors in the data. Although this is a deficiency of the chamber itself, we have now learned how to minimize the errors. The problem can also be solved simply by modification of the scanning program to allow a larger integrating box.

3. Uneven detector response.

The elementary underlying assumption is that identical doses of radiation will cause identical response in the detector, regardless of the location of the impulse on the face of the detector. To test this we have performed a simple experiment of measuring the response to an attenuated primary beam across the detector. For this purpose, 10s exposures of attenuated primary beam were taken for various 2 theta settings of the detector, from -25 deg to 30 deg, at 5 deg intervals; 8.0 cm detector-to-crystal distance was used. In order to obtain accurate estimates of total intensities, 10 exposures were taken at each setting, and deviations were computed from the spread of the values. 25*25 pixel box was used to ensure that all photons are included, and the background was estimated from 216 pixels outside this box (i.e. two extra rows and columns). The background had a minimal effect, since it had total values between 5 and 20 counts for all 216 pixels. Fig. 1 shows the results of this experiment.

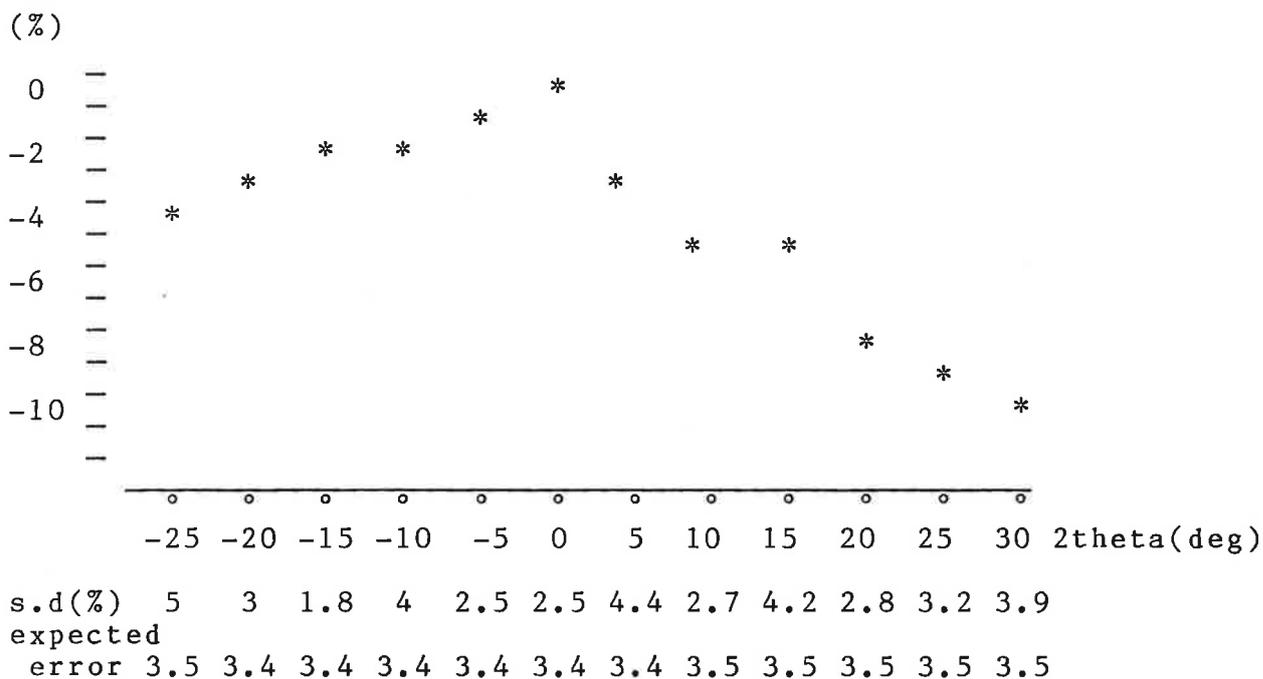


Fig. 1. A plot illustrating the fall off in intensity of the attenuated primary beam. The intensity 2 theta = 0.0 is assumed to be 100 %; the table underneath the chart shows the standard deviations determined for each measurement from the spread of values and compares them with values of expected errors calculated from counting statistics.

It is easily seen, that apart from the uneven spatial distortion across the detector, the enlargement of spots is accompanied by a loss of counts up to 10% for the left side of the detector. The shape of the curve suggested that some of this effect may be due to obliquity effects in beryllium and different optical paths. To test this hypothesis the experiment was repeated at 23.0 cm detector-to-crystal distance. This distance is the radius of curvature of the beryllium window, and therefore any obliquity effects as well as differences in optical paths would be reduced to zero at this setting. Any loss of

counts under these conditions represents a genuine deficiency of the detector. Fig 2 gives the results of this experiment.

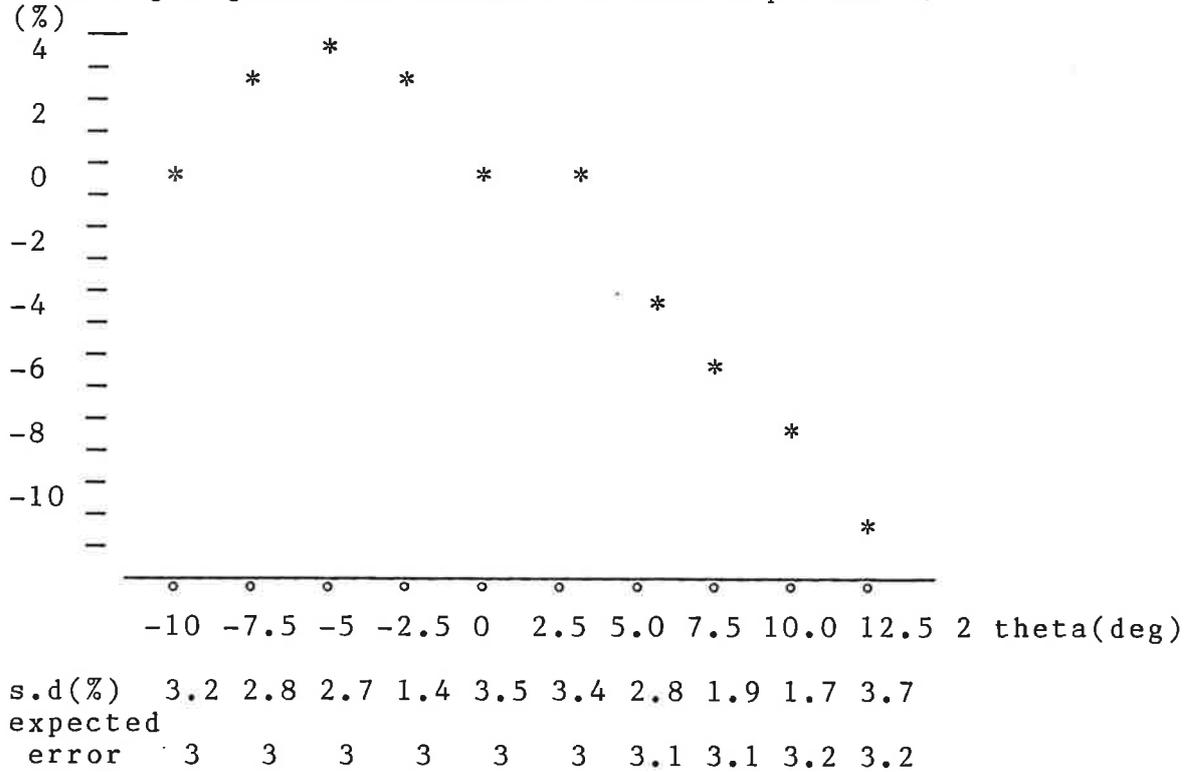


Fig. 2. The response of the detector at 23.0 cm.

It is clearly seen, that there is a considerable loss of counts towards one side of the detector.

It must be therefore concluded, that the variations in size of the spots, invalidating the present mode of integration, as well as variations in response to identical doses of radiation are a characteristic property of the detector.

Although the extreme sensitivity of the pre-amplifiers to vibrations is a phenomenon known to occur in practically all detectors, the two problems outlined above seem to reflect the poor quality of the particular chamber installed at York. We are at present negotiating with Nicolet a replacement of the detector.

No attempt was made to alter in any way the existing software package written by Andrew Howard (Genex, U.S.A.) and supplied by Nicolet, but simple steps were taken to minimize the effects of the identified sources of errors. For example, the use of relatively small crystals and an introduction of rectangular slits in the collimator to reduce the effective cross-fire of the primary beam, substantially reduced the size of the spots.

Having identified major sources of errors and having learned how to minimize or eliminate their effects, we proceeded to collect real data. Table 1 shows basic statistical information for a number of data sets collected so far.

TABLE 1

data set	number of obs.	number of refl.	rejected obs.	R (%)	res. (A)
diag.				ca.2.5-4.0	2.8
dapi	7870	5551	137	4.3	2.0
ins (1)	7168	4889	197	4.3	2.0
ins (2)	7990	5250	1035	4.1	2.0
ortho	17463	8036	86	4.4	2.0
x8	14746	3115	388	4.8	2.8
p.a.Pt	40669	16071	274	6.1	2.8
p.a.nat.	52093	20099	1026	5.9	2.8
enz.Sr	13229	6280		5.9	2.8
enz.Hg	31887	7592		7.3	2.8

diag. - diagnostic data sets on orthorhombic insulin crystal;
space group (SG) P212121
dapi - A-21 des-amido porcine insulin ; SG R3
ins (1) & ins (2) - crystals from two preparations of 2Zn human
insulin; SG R3
ortho - orthorhombic crystals of porcine insulin; full data set
x 8 - mutated insulin (A8 His)
p.a.Pt - Pt derivative of penicillin acylase; SG P1
p.a.nat.- native penicillin acylase
enz.Sr & Hg - two derivatives of a commercial enzyme SG C2221

The structure of des-amidated insulin (DAPI) was refined from the model of native porcine molecule (R value of 17.8%) to an R value of 13.9%. Small alterations resulting from des-amidation have been identified and reported elsewhere (E.E.C. Biotechnology Action Programme, Sectoral Meeting: Enzyme Engineering, Capri '87; Collected Abstracts). Most recently the structure of orthorhombic crystals of insulin was solved using the detector data by the molecular replacement procedure. Work is continuing on the structure solution of two larger enzymes.

In general, we found the system very powerful and useful even in the presence of the faults described above. In spite of the obvious systematic errors, the quality of the data was, to say the least, not inferior to data obtained by photographic methods.

THE AREA DETECTOR SOFTWARE DISCUSSION GROUP

Report on the first meeting John W. Campbell, Daresbury Laboratory

1) INTRODUCTION

The first meeting of the 'Area Detector Software Discussion Group' was organised by John Helliwell under the auspices of CCP4 and the BCA Biological Structure Group. It was held during the BCA Spring meeting in April at the Heriot Watt University, Edinburgh. All PX groups in the UK had been invited to send a representative. The meeting was attended by six people including the four speakers with several groups being unable to send a representative. The topics covered were:

The functions of the discussion group

A report on the EEC initiative

Discussion Theme 1 - Autoindexing

Suggestions for future meetings and discussion themes

2) THE FUNCTIONS OF THE DISCUSSION GROUP

John Helliwell outlined some of the reasons why there was a need for the discussion group. He felt that the information available about the developments of area detector software should be made available to the UK protein crystallography groups through a group which might act almost as a sub-committee of Working Group II of CCP4.

Two specific areas of particular importance were identified:

- a) An opportunity for the UK representative(s) involved in the EEC initiative (see below for more details) to report back to the UK groups. Though there were several UK representatives at the first phase of this initiative, there has been only one, Rob Stansfield, at the later phases.
- b) An opportunity for representatives on Specialist User Groups for various devices (e.g. FAST, XENTRONICS) to report on the latest developments and problems on those devices.

The material reported back would be a source of discussion for the group and it would be hoped that good ideas could be fed back to the EEC initiative and to the Specialist User Groups through the UK representatives.

It was felt that CCP4 had a role to play in these discussion groups and a request for some funding from CCP4 had already been made. Some money may be available to fund a small technical meeting in the year 1987/1988.

3) A REPORT ON THE EEC INITIATIVE

Rob Stansfield reported on the EEC initiative to set up a Cooperative Workshop on Position-Sensitive Detector Software. This was the brainchild

of Gerard Bricogne who has previously participated in a similar venture in a workshop on Isomorphous Replacement. The present initiative was thought to be different in character as there was much less general expertise available and an input of new ideas was required. The EEC workshop is being funded by the EEC and about six phases are planned of which the first four phases have already been completed. The workshop meetings are held at L.U.R.E, Orsay, in France.

Phase I (May 1986) involved about 30 scientists and involved a general discussion of the requirements for the software with a review of software already in existence. It was agreed that the aim of the workshops should be to produce device independent software and it was agreed that the Pflugrath and Messerschmidt package MADNES should be used as the starting point.

Phase II (June 1986) involved a dozen scientists and led to detailed short and long term programming projects aimed at developing the MADNES software.

In Phase III (Nov 1986), participants reported on the progress of a number of projects. The results of these projects were discussed and further plans made.

Phase IV (Jan 1987). There were no UK participants in this workshop and details of the work carried out have not yet been received.

Rob gave a flavour of the ways in which the members of the workshop were approaching the problems. Consideration was given as to what parameters needed to be defined and what algorithms were required. Questions as to how the various topics were inter-related were discussed and topics were classified as dependent or independent. Dependency diagrams were drawn up for all topics. Further suggestions for the structuring of MADNES were made with the definition of Active and Passive objects. A typical active object might be a subroutine for collecting a set of data and a passive object might be a Fortran Common block containing a set of goniometer parameters. Import and Export mechanisms would be available to access and update items of information.

Two documents had been produced by Gerard Bricogne, the first giving details of Phases I and II of the workshop and the second giving details of Phase III. A version of the EEC software might be available during the summer of 1987.

After Rob's description of the progress of the EEC initiative a number of points were discussed. These included the following:

- a) The relative merits of collecting whole images (suited to generalised software) and collecting selected parts of the data only (could achieve much faster data collection rates though would be much more device dependent and some possibly useful information might be lost) were discussed. The possible role of Transputers in both these approaches was considered.
- b) It was felt that the MADNES approach of describing the detector in terms of a virtual flat plate could cause problems in some cases when dealing with real detectors which were not flat.
- c) The development of ideas for extracting integrated intensities could

well have relevance to obtaining better integrated intensities from film methods (both Oscillation and Laue) than could be obtained using box integration or current profile fitting methods.

4) DISCUSSION THEME 1 - AUTOINDEXING

Recent developments in Autoindexing were described and discussed for the following three areas of work:

Laue data processing

Autoindexing on the FAST for micro crystals

Autoindexing and Oscillation films

a) Laue Data Processing

John Campbell described the Autoindexing method for Laue photographs programmed by Mike Elder. Details of the method used may be found in an earlier version of the Information Quarterly (reference 1.)

b) Autoindexing on the FAST for micro crystals

Steve Andrews described a method, which he had programmed, for using autoindexing with micro crystals on the FAST. The basic principles of the method were those of Jacobson (reference 2). An orientation matrix is defined by choosing three reference reflections:

$$R = \begin{array}{ccc|ccc} & & & & & -1 \\ |x_1 & x_2 & x_3| & |h_1 & h_2 & h_3| \\ |y_1 & y_2 & y_3| & |k_1 & k_2 & k_3| \\ |z_1 & z_2 & z_3| & |l_1 & l_2 & l_3| \end{array}$$

The indices of any other reflection can then be calculated:

$$\underline{H(j)} = R^{-1} \underline{V(j)}$$

Then h_1 , h_2 , h_3 are varied until three sets of linearly independent indices are obtained which generate the indices $H(j)$ to within a preset error level.

The steps in the autoindexing process are as follows:

Choose 5 to 10 low resolution reflections

Select three non-coplanar reference reflections and set an upper index limit for these reflections

Find an autoindexing solution as indicated above and calculate the orientation matrix and cell dimensions

Include all the reflections and refine the orientation matrix

Calculate the Kappa, Omega and Phi angles for the FAST required to align the cell axes relative to the FAST axes and calculate initial

missetting angles

Refine the crystal/camera parameters in MADNES in the normal way or use the indices obtained through the autoindexing to constrain the refinement

The method had proved successful with the micro crystal of Piperazine Silicate and had enabled its structure to be determined using data collected on the FAST. An attempt to use the method on a Copper complex had not been successful, perhaps because the crystal was too well set. The method was suited to small molecules rather than proteins where an autoindexing method using difference vectors would be more appropriate.

c) Autoindexing and Oscillation Films

John Helliwell reported that Peter Brick had attempted to use Andy Howard's autoindexing program with data picked off still photographs. The method had not proved to be successful probably because the program needs a large number of reflections measured with good Phi centroids. An alternative approach, using the oscillation data photographs themselves, has not yet been tried.

5) SUGGESTIONS FOR FUTURE MEETINGS AND DISCUSSION THEMES

The layout of the first meeting could be a model for future meetings with feedback from the EEC workshop and specialist user groups and with a main discussion theme. It was not yet clear where the meetings should be held or how many participants would be expected. Small scale discussion groups could perhaps be satellite meetings to the CCP4 working group meetings and would be appropriate for very specialised topics. Larger discussion meetings of more general interest would probably have to be held in conjunction with meetings such as the BCA meetings where many interested persons were already likely to be present. There would in any case be only very limited finance available to fund them.

It was felt that the next discussion group should be a small scale one with a theme of 'Calibration and Diagnostic Software'. Future discussion themes might concentrate on the integration step in processing the data or on the impact of new computer hardware (e.g. Transputers, array processors, laser disks).

REFERENCES

- 1) 'LGEN and LCHK: Utility programs for investigating Laue photographs', M. Elder, Information Quarterly for Protein Crystallography, (1986), 19, 31
- 2) Jacobson, J. Appl. Cryst., (1976), 9, 115

PROGRESS REPORT ON THE X-RAY STRUCTURAL ANALYSIS OF DIFERRIC
RABBIT SERUM TRANSFERRIN.

S. Bailey**, R. W. Evanst, R. C. Garratt*, B. Gorinsky*,
S. S. Hasnain**, H. Jhoti*, P. F. Lindley* and R. Sarra*.

* Department of Crystallography, Birkbeck College
(University of London), Malet St., London WC1E 7HX.

† UMDS Division of Biochemistry, Guy's Hospital,
London SE1 9RT.

** SERC Daresbury Laboratory, Daresbury,
Warrington WA4 4AD.

The transferrins are a class of homologous two-sided iron-binding proteins, molecular weight ≈ 80 kD, comprising serum transferrin (from blood), lactoferrin (from milk and other secretory fluids), ovotransferrin (from egg white) and melanotransferrin (a cell surface protein on human melanoma cells), [for reviews see Morgan, (1981) and Brock, (1985)]. Although significant amino acid sequence similarities exist between the various transferrins, there are thought to be important differences in their biological functions; serum transferrin is the only member of the class whose primary function is to transport iron. Extensive biochemical and spectroscopic studies have shed comparatively little light on either the nature of the iron binding sites or the uptake, release and transport mechanisms. Structural studies should provide an insight into this important metabolic activity.

The molecular structure of diferric rabbit serum transferrin has been solved at 3.3 Å resolution by MIR in conjunction with solvent flattening techniques. The MIR map alone, based on protein phases derived from three heavy atom derivatives, (figure of merit $\approx 62\%$), clearly revealed the molecular boundary, the positions of the two iron atoms in the molecule, and substantial stretches of connected electron density corresponding to the polypeptide backbone chain. However, there were many apparent breaks in the chain and this, together with the presence of some 19 disulphide bridges, made complete interpretation almost impossible.

A recent determination of the structure of human lactoferrin at 3.2 Å resolution, (Anderson *et al.*, 1987), suggested that the use of solvent flattening techniques at 3.3 Å resolution might resolve the interpretation dilemma. Previous attempts at solvent flattening at lower resolutions, (6.0 Å and 4.5 Å), but using phases determined from only two heavy atom derivatives, had not been successful in sharpening the detail within the molecular boundary.

Initially, the 3.3 Å MIR map was contoured at a sufficiently high level to enable the major volumes of molecular density to be readily distinguished from the solvent region, ($\approx 68\%$ by volume), and thereby facilitating the manual designation

of a molecular boundary. All solvent density was then set to zero, negative regions within the envelope were multiplied by 0.1 and the resulting map back transformed using the standard CCP4 programmes. In the phase recombination step equal weight was given to the MIR and back transformed phases. This procedure was repeated four times before calculation of the final map. The resulting solvent flattened map was a significant improvement on the MIR map and enabled, for the first time, convincing regions of secondary structure, and particularly the α -helices directed away from the iron binding sites, to be identified. However, several connecting loops near the edge of the envelope were either poorly defined or completely absent. A second approach to solvent flattening was to use Leslie's, (1987), reciprocal space averaging modification to the algorithm of Wang, (1985). Two maps were computed using averaging spheres of 8Å and 10Å respectively and a solvent content of 65%, only 3% less than the estimated value of 68%. There were very few significant differences between these two maps and in both of them the definition of the elements of secondary structure was generally poorer than in the first solvent flattened map. However, the connectivity between such elements was much improved particularly in the exposed regions and this has allowed most of the ambiguities of the original map to be resolved. A further Wang map, for which the protein content was deliberately set some 15% greater than predicted, failed to produce any further interpretable features. It should be stated, that with hindsight, much of the secondary structure can be identified even in the MIR map prior to solvent flattening.

The initial chain tracing of the polypeptide backbone, some 650 amino acids, was undertaken from a combination of minimaps of all the syntheses described above. In order to transfer this information as rapidly as possible to the PS300 we opted to use the automatic electron density skeletonisation of Greer, (1974), as implemented in the programme 'Bones' of Jones and Thirup, (1986). Using a basis level of 1.3 x the standard deviation of the map and an increment value equal to the standard deviation as recommended by Jones, a chain trace was determined largely in agreement with that obtained from the minimaps. The most common errors in the Bones trace were bifurcations in the main chain at disulphide bridges and a failure to find some of the more diffuse surface loops. After manually correcting the connectivity of the Bones atoms we have adjusted their positions to represent approximate C_{α} coordinates.

The iron atoms can be clearly located in the solvent flattened maps and correspond to the two highest peaks in all cases. These positions have also been unequivocally identified from anomalous difference electron density syntheses using $[\alpha = -\pi/2]$ phases, (see, for example, Harada *et al.*, 1986), on datasets collected with CuK_{α} radiation, and synchrotron radiation tuned to 1.725 Å in the EXAFS region above the Fe K edge. The corresponding anomalous difference Patterson syntheses were not convincing.

The overall tertiary structure and iron site geometry in rabbit serum transferrin is similar to that found in human lactoferrin. The disulphide bridge pattern, however, is rather different; additional disulphide bridges can be identified in both the N- and C-terminal lobes in agreement with predictions based on the amino acid sequence of human serum transferrin. In the N-terminal lobe, however, disulphides 4,5 and 11 are in a different arrangement to that proposed by MacGillivray *et al.*, (1982,1983), giving an improved sequence match with other transferrins. The presence of extra disulphide bridges results in structural changes and these may account, in part, for some of the significant differences in the physiological roles of the serum and lacto proteins.

The structure is bilobal, with one iron binding site per lobe, and has a pseudo two-fold axis relating the two lobes of the molecule. A similar folding pattern is observed in each lobe consistent with a gene duplication event. The super secondary structure of each lobe exhibits two domains built up from $\beta\alpha\beta$ units, with the majority of the helices pointing away from the iron site and with the iron site at the domain interface. Each iron is coordinated to at least four protein ligands which, on the basis of homology with lactoferrin, can be identified as two tyrosines, one histidine and an aspartic acid residue. An arginine residue is also in the vicinity of the iron site and this residue and the iron may be bridged by the (bi)-carbonate anion required for iron binding. In the vicinity of the C-terminal lobe there is also electron density, particularly prominent in the MIR map, which may correspond to a single branched-chain carbohydrate moiety.

EXAFS studies, (Garratt *et al.*, 1986), suggest that the iron is at least six-coordinated and comparative spectra on freeze-dried and solution protein samples further suggest that another ligand is a water molecule, (Hasnain *et al.*, 1987). This is consistent with the electron density distributions around the iron sites in the various maps which clearly show that the four protein ligands are clustered on one side of the iron atom, leaving space for a water molecule to complete the coordination sphere via the cleft. Such a molecule may be involved in ligand exchange reactions which lead to the release of iron intracellularly.

Currently, we are using the fragment fitting approach of Jones and Thirup, (1986), to build in the remaining backbone atoms on our Bones C_α skeleton, including side chains where they are evident and consistent with homologous sequences. Hopefully, cycles of refinement and computer graphics model building will clarify the few regions of the protein which, at this stage, appear obscure. One such region, in strong contrast to the case of human lactoferrin, is the connecting peptide region between the two lobes. Ultimately, it is hoped to undertake a refinement of the iron site geometry on the basis of the EXAFS data where an improvement of an order of magnitude in precision can be expected for metal-ligand coordination distances.

References:

- Anderson, B.F., Baker, H.M., Dodson, E.J., Norris, G.E., Rumball, S.V., Waters, J.M. and Baker, E.N. (1987). *Proc. Natl. Acad. Sci. (USA)*, in press.
- Brock, J.H. (1985). In *Metalloproteins, Part 2.*, Ed. P.M. Harrison, 183-261, MacMillan, London.
- Garratt, R.C., Evans, R.W., Hasnain, S.S. and Lindley, P.F. (1986). *Biochem. J.*, 233, 479-484.
- Greer, J. (1974). *J. Mol. Biol.*, 82, 279-302.
- Harada, S., Yasui, M., Murakawa, K. and Kasai, N. (1986). *J. Appl. Cryst.*, 19, 448-452.
- Hasnain, S.S., Evans, R.W., Garratt, R.C. and Lindley, P.F. (1987). Submitted to *Biochem. J.*.
- Jones, T.A. and Thirup, S. (1986). *Embo. J.*, 5, 819-822.
- Leslie, A. (1987). *Acta Cryst.*, A43, 134-136.
- MacGillivray, R.T.A., Mendez, E., Sinha, S.K., Sutton, M.R., Lineback-Zins, J. and Brew, K. (1982). *Proc. Natl. Acad. Sci. (USA)*, 79, 2504.
- MacGillivray, R.T.A., Mendez, E., Shewale, J.G., Sinha, S.K., Lineback-Zins, J. and Brew, K. (1983). *J. Biol. Chem.*, 253, 3543.
- Morgan, E.H. (1981). *Mol. Aspects. of Med.*, 4, 1-123.
- Wang, B.C. (1985). In *Methods in Enzymology*, Eds. Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., 115, 90-112.

PROGRAMME TO CONTOUR PROTEIN CHANNELS AND CAVITIES

D. Barford and L.N. Johnson,
The Rex Richards Building,
Laboratory of Molecular Biophysics,
University of Oxford,
Oxford. OX1 3QU.

1. Introduction

Richards, (1974), has emphasized the role played by regions of low atomic packing density, representing cavities and channels, in the structure and function of proteins. Their analysis is useful in attempting to understand ligand binding specificity and enzyme catalysis. Knowledge of the structure and size of cavities and channels can assist in the model building approach of fitting ligands to protein sites. Moreover, they represent likely loci for conformational change.

To facilitate the delineation and analysis of regions of low packing density, a programme was written which calculates a map, the density of which is inversely proportional to the packing density. After contouring the map is displayed interactively on a PS300, analogously to an electron density map. In this method, by highlighting regions of low packing density, the position and structure of cavities and channels may be investigated easily. A higher contour level is displayed according to the depth of the channel.

The programme was applied to the enzyme glycogen phosphorylase b, comprising 7000 atoms, excluding waters. Of particular interest here was the catalytic site, and the channels leading from the surface 15A away.

2. Programme description

A grid with spacing of 1A is established and the proximity of atoms to each grid point determined in the following way. If the grid point is within a Van der Waals radius of any atom, it represents a region of high packing density, and a density value of zero is assigned to the point. Atoms are treated equivalently each with a Van der Waals radius of 1.5A. When a grid point is greater than 1.5A from the closest atom, the grid point is assumed to be within a protein cavity and assigned a positive density value. Grid values are set at arbitrary values of 200, grid-atom distance greater than 1.5A; 400 for 2.5A; 600 for 3.5A; and 800 for 4.5A.

The co-ordinates of the atoms are sorted by dividing the protein into blocks with dimensions of 10A x 10A by the length of the protein along Z. The spatial interval between adjacent blocks along X and Y is 1A. Atoms which fall within the

extent of a particular block are stored in an array. For a grid point of value X, Y and Z, all atoms within the block which has its centre at X and Y are assessed for their proximity to the grid point.

The programme output consists of a direct access file of integers, where the grid coordinate is stored as the integer position in the file and the grid density as the integer size. Contouring of the map is performed by running MAPBRICK, (P.R. Evans, modified for Oxford map format by D.I. Stuart). The final contour map is viewed interactively aided by the programme FRODO, Jones, (1978), implemented on a PS300.

3. Results

Running the programme on a DEC VAX-11/750 with an 8 megabyte memory for the protein glycogen phosphorylase b, dimensions 80A. 80A, 72A took 132 mins. of cpu time.

Figure 1 shows the polar and solvated cavities of the catalytic site. The largest in extent and depth corresponds to the glucose-1-phosphate binding site. From here extend 3 channels, one of which is large enough to accommodate glucan-hydroximo-1,5-lactone-N-phenylurethane. All three channels have their paths to the surface blocked.

Elsewhere, polar cavities and channels are water filled, some despite having no access to the surface. Non polar cavities are empty.

We thank Drs. P.J. Mclaughlin and D.I. Stuart for helpful advice on implementing this programme.

4. References

- Jones, T.A., (1978), J. App. Cryt., 11, 272-288.
Richards, F.M., (1974), J. Mol. Biol., 82, 1-14.

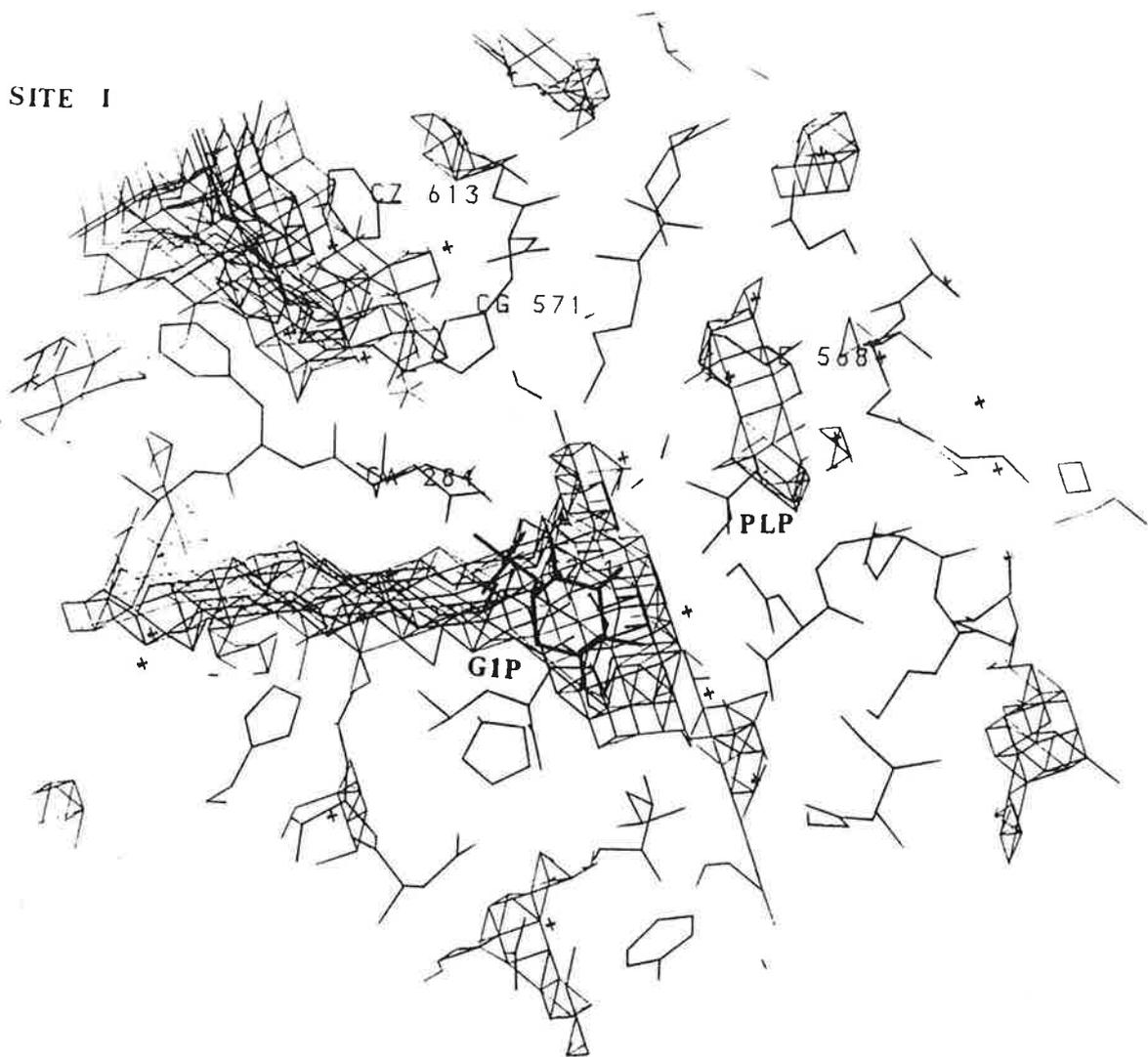


Figure 1.

Cavities and channels at the catalytic site of phosphorylase generated at distances greater than 2.5Å from any atom. The substrate glucose-1-phosphate is shown. The channel to its left accommodates glucanhydroximo-1,5-lactone-N-phenyl-urethane. The inhibitor site at the protein surface is at the top left, with access to the active site blocked by histidine 571.

Steven Glover, The Astbury Department of Biophysics, Leeds

Introduction

The program MOPAC¹ is a semi-empirical Molecular Orbital PACKage which was designed to allow the study of chemical reactions. Although the program can be used for calculations of molecular orbitals, eigenvectors, vibrational spectra and thermodynamic quantities, we have mainly used it for geometry optimisation and partial charge calculations on small molecules whose structures are not known with sufficient precision to allow accurate conformational energy calculations, as a precursor to performing protein:ligand docking both interactively on a graphics system (with partial charges indicated by colour coding), and in 'batch mode' with an energy minimisation program.

It is not necessary to be a specialist in quantum chemistry or advanced mathematics to make use of the program because it was written in such a way that the non-theoretical chemist can run it and obtain useful results from a simple set of input data. It is possible to treat the program as a 'Black Box', modelling all calculations on the examples supplied in the manual, but to get the best use from the program it is desirable to understand at least the basics of data input, running the program and interpretation of results.

Data input

As implied above, the control file for the program is quite simple. The first line of the file contains a list of keywords which define the type of calculation to be carried out. The next two lines are available for the titling and comments on the run. If these lines are left out they should be replaced with blank lines, but it is a good idea to include as much information here as possible. The next group of lines define the geometry of the input molecule in internal coordinates - that is, each atom is defined by a bond, a bond angle and a dihedral connecting it to previous atoms in the list. Each value is followed by an optimisation flag, which is set to 0 if the parameter is not to be optimised, or to -1 if it is to be used as a reaction coordinate. The end of the geometry data is signalled by a blank line, and in most cases this is also the end of the file. If symmetry is used to define one part of the molecule in terms of another, or if a geometrical parameter is being allowed to vary systematically (this may be as simple as rotating a dihedral angle or slowly withdrawing a proton from a carboxyl group - or it may be as complex as attacking an amide or ester linkage with a hydroxyl anion - a 'reaction path' calculation), the extra information is given after the blank line. In any case, the file is finished with a blank line.

There are 65 possible keywords to define the calculation and the flow of control within the program, but a number of these are used for printing out details of working in various subroutines and are used mainly for debugging, checking or educational purposes. Assuming that the main reason for a calculation is to obtain a better information on (for example) an enzyme inhibitor of poorly determined geometry, only a few of the keywords are necessary.

The geometry specification is best described by use of an example, in this case methanol.

```
      H5
      |
H1-O2-C3-H4
      |
      D6
```

atom	bond	angle	dihedral	nb	na	nt	comment	space
H	0.0 0	0.0 0	0.0 0	0	0	0		
O	0.97 1	0.0 0	0.0 0	1	0	0		
C	1.38 1	110.0 0	0.0 0	2	1	0		
H	1.12 1	109.4 1	180.0 1	3	2	1		
1	1.12 1	109.4 1	-60.0 1	3	2	1	hydrogen	
H2.0	1.12 1	109.4 1	60.0 1	3	2	1	deuterium	
	0.0 0	0.0 0	0.0 0	0	0	0	(blank line)	

The first atom has no coordinates specified because it is at the origin. The second atom is displaced 0.97 Angstrom and this distance is allowed to optimise during the calculation. The third atom is displaced a further 1.38 Angstrom, making an angle of 110° degrees with the O-H bond and this angle is held fixed (not marked for optimisation). Similarly, the fourth, fifth and sixth atoms are all attached to the third atom, making various dihedral angles with the first. These last two atoms also indicate the differing ways of specifying atoms. Atom 5 is represented by its atomic number while atom 6 is a deuterium, and hence is given the atomic mass to be used immediately after the symbol. Atomic masses are only necessary for force constant, vibrational spectra and thermodynamic calculations.

Running the program

Obviously, the running of the program is going to depend on the machine it is implemented on (and MOPAC is available for CDC, Digital, IBM and both Cray-XMP and CDC-205 supercomputers), but on any machine array sizes in the program may be changed when it is initially compiled. The default array sizes allow the use of 40 atoms, of which up to twenty may be non-hydrogen, and this would seem to be satisfactory for most purposes - especially as cpu time used goes up dramatically with molecule size. For example, while formamide takes about 30s, N-acetyltyrosine methyl amide can take 30hr (these times are for approximate starting coordinates and measured on a VAX-11/750). This means that large molecules should be run in conveniently sized fragments whenever possible. A dipeptide, for example is best modelled as two single amino-acids with acetyl and N-methyl groups attached.

On memory limited systems, or on systems where some options are never going to be used, it is possible to compile and use a shorter version which has some subroutines replaced by dummies. If an attempt is made to use the removed routines, the program halts with an error message before significant cpu time is used. Another useful feature is the ability to shut down a calculation in a recoverable manner. This can be important if MOPAC's use of resources begins to interfere with other users, or if the computer has to go 'down' for any reason during a run.

Interpretation of results

A successful run produces two output files. The first is the listing file, which contains the results of the calculation, and the other is the archive file, which contains the run summary, the input file and the input file as amended to contain the new geometry.

The output file starts off with a header, showing which MNDO Hamiltonian was used. This is followed by a list of the keywords recognised in the input file (useful if one has been mistyped or over-abbreviated). A copy of the input file follows, for checking that the data has been input correctly. If symmetry is used, the next part of the file show which parameters are related. The cartesian coordinates of the input molecule follow, together with the number of doubly occupied orbitals. A table of interatomic distances is next, and as a crude geometry check the program halts execution if two atoms are closer than 0.8 Angstrom. This is

sufficient to trap a surprisingly large percentage of erroneously input geometries. The next information output is the calculated heat of formation for the input geometry, and this is followed by the values of the gradient norm and heat of formation for the geometry after each least squares cycle until either the number of cycles performed exceeds 100 or the gradient norm reaches a minimum or drops below a given value.

The gradient quoted is a measure of how far each parameter is from its optimum value, and it is given as a scalar length of the optimisation gradient vector (in kcal/mole/Angstrom). It should be below five for a well optimised amino acid, but values of up to ten are acceptable for larger structures, or cyclic molecules. The next part of the output is the final values of electronic energy, core-core repulsion, ionisation potential and number of SCF calculations along with the amount of cpu seconds used.

This is followed by the optimised geometry and a table of interatomic distances. It will usually be seen that bad contacts have been relieved by opening angles or lengthening bonds rather than by rotating a torsion angle. It must be emphasised that this geometry optimisation is not the same as an energy minimisation. For example, if the geometry of N-methyl acetamide is specified, with the exception of a 90° twist applied to the amide linkage, then an energy minimisation program would probably rotate the molecule back to planarity, possibly even finding two minima corresponding to *cis* or *trans* configurations of the amide bond. The MOPAC geometry optimisation routine however, will hardly change dihedral angles, and will produce an output molecule which would correspond more to a 'nitrogen ester', in that there will be less partial double bond character of the N-C bond, and the C=O bond will be slightly shorter.

This is followed by the table of eigenvalues for the molecular orbitals and the table of partial charges for the molecule, along with information on the dipole if the molecule is not ionised. The next information given is the cartesian coordinates of the new geometry. Finally, if they have been requested, the bond matrix, localised bond table and energy levels for the occupied localised molecular orbitals are given. This can be useful for determination of which bonds would weaken on ionisation, excitation or on changing the amount of electron density available to the molecule.

Summary

Although the MOPAC program is complex and uses large amounts of cpu time and memory to run (a molecule of about 30 atoms can use about 4Mb memory on a VAX-11/750), it is better than geometry regularisation or force field calculation when the molecule contains functional groups or sequences of atoms that are not sufficiently common to appear in most sets of parameters used. After a run through MOPAC to optimise the geometry, any further energy calculations can be done with less resource-intensive programs. It is of course, possible to verify the accuracy of the new conformation by running it through MOPAC. The partial charges calculated by MOPAC are also 'better' than those given by programs that either use look-up tables for each type of atom found in, say, a protein, or perform approximate calculations based on the immediate connectivity of each atom.

Reference

1) MOPAC: A General Molecular Orbital Package.
By James J.P. Stewart, Dewar Group, University of Texas, Austin, Texas 78712.
Available from the Quantum Chemistry Program Exchange, Department of
Chemistry, Indiana University, Bloomington, Indiana 47405, USA

An improved ribbon algorithm for proteins

Phil Evans

MRC Laboratory of Molecular Biology, Cambridge

An algorithm for generating protein main-chain ribbon pictures using the B-spline function of the PS300 has been described by M. Carson & C.E. Bugg (J.Molec.Graphics (1986) 4,121-122). This algorithm defines a set of guide points for the spline curve which give a ribbon of closely-spaced strands approximately parallel to the peptide plane and passing close to the α -carbon atoms. However, a spline curve does not pass through the guide points, so a series of guide points at the α -carbon positions would give a very narrow helix in an α -helical region. It is therefore necessary to displace the guide points away from the run of the chain to compensate. In the Carson & Bugg algorithm, the guide points are halfway between $C\alpha$ positions, and in helical regions are displaced along the normal to the peptide plane. This means that the helical regions must be determined or specified, which is a nuisance. In the algorithm presented here, the guide points are displaced from the $C\alpha$ positions along the bisector of the $C\alpha_{i-1} \rightarrow C\alpha_i$ and $C\alpha_i \rightarrow C\alpha_{i+1}$ vectors. This works well for all types of protein structures, so helix assignment is unnecessary. The ribbon plane is parallel to the mean plane of the peptides on either side of the $C\alpha$, and passes close to the $C\alpha$ positions with an appropriate offset.

This algorithm has been implemented in the MRC version of Frodo, as a MOL option.

The algorithm: the notation here follows that of Carson & Bugg

Parameters:

- w ribbon width in Å
- n number of ribbon strands
- f offset from $C\alpha$ position (1.2Å seems a good value)
- m number of chords to interpolate between guide points
(for the B-spline function)

For each residue i , we need the coordinates of the α -carbon $C\alpha_i$ and the carbonyl oxygen O_i .

$$\begin{aligned} \mathbf{a}^i &= \mathbf{C}\alpha_{i+1} - \mathbf{C}\alpha_i \\ \mathbf{b}^i &= \mathbf{O}_i - \mathbf{C}\alpha_i \\ \mathbf{c} &= \mathbf{a}^i \times \mathbf{b} && \text{normal to peptide plane} \\ \mathbf{d}' &= \mathbf{c} \times \mathbf{a} / |\mathbf{c} \times \mathbf{a}| && \text{normalized vector across ribbon} \\ \mathbf{d}_i^i &= \text{sign}(\mathbf{d}_{i-1} \cdot \mathbf{d}_i') \mathbf{d}_i' && \text{sign change stops ribbon from flipping} \\ &&& \text{over} \\ \mathbf{e} &= (\mathbf{d}_{i-1} + \mathbf{d}_i) / |\mathbf{d}_{i-1} + \mathbf{d}_i| && \text{mean vector across ribbon} \\ \mathbf{p}' &= (\mathbf{a}_{i-1} - \mathbf{a}_i) / |\mathbf{a}_{i-1} - \mathbf{a}_i| && \text{direction of offset from } C\alpha \text{ position} \\ \mathbf{p} &= \mathbf{C}\alpha_i + f \mathbf{p}' && \text{position of offset guide point for} \\ &&& \text{centre of ribbon} \end{aligned}$$

Then the i 'th guide point for the j 'th ribbon strand \mathbf{g}_{ij}

$$\mathbf{g}_{ij} = \mathbf{p} + [j-(n-1)/2][w/(n-1)] \mathbf{e} \quad \text{for } j=1,n \quad (\text{if } n=1, \mathbf{g}_i = \mathbf{p})$$

For the first $C\alpha$ atom:

$$\mathbf{e} = \mathbf{d}_i \quad \mathbf{p}' = 0$$

For the last $C\alpha$ atom:

$$\mathbf{e} = \mathbf{d}_{i-1} \quad \mathbf{p}' = 0$$

A bug in PROTIN/PROLSQ

There has been an error in the Hendrickson-Konnert refinement program PROTIN in the calculation of chiral volume and torsion angle pointers if some atoms in a residue are missing. This is only serious in residues other than amino-acids, such as protein ligands, substrates eg sugars, nucleotides etc. The program is correct as long as all atoms in these residues are included. If atoms are missing, it often caused the refinement in PROLSQ to blow up, with the conjugate gradient refinement diverging. The error has been corrected in my version, and the corrections sent to Daresbury.

Eleanor Dodson
Chemistry Department, University of York

