

DARESBUURY LABORATORY
INFORMATION QUARTERLY
for
PROTEIN CRYSTALLOGRAPHY

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

Number 16

SEPTEMBER 1985

Contents

Editorial	1
Progress on the Antigen-Antibody Complex Structure at Leeds/Institut Pasteur, Simon Phillips	3
Hendrickson-Konnert Refinement on the DAP Jim Raftery	5
Progress with Glutamate Dehydrogenase, P. Baker, G. Farrants, D. Rice, F. Rodgers	7
A study on the inter and intra subunit contacts in M-type Pyruvate kinase, D. Clayden	11
Use of Post refinement for Analysis of Systematic Errors in GAPDH High resolution Dataset, Alan Wonacott	13
Crystallographic Studies on the Vertebrate Lens Proteins, Huub Driessen	19
News from York University, Guy Dodson	23
Frodo and the PS300, P. R. Evans	25
Renewal of CCP4 in Protein Crystallography, Pella Machin	33
Meeting on Protein Structure Prediction at Daresbury, 24/25 January 1986 Pella Machin	35
Unscrambling of harmonic reflection intensities from spots on Laue patterns: Results on Pea Lectin S. Zurek et al.	37
Computing at the Laboratory of Molecular Biophysics Garry Taylor	41

Editor: Pella Machin

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

Deputy Editor
for Imperial
College: Dr. Alan Wonacott

Imperial College of Science & Technology,
The Blackett Laboratory, Prince Consort Road,
London SW7 2BZ

Deputy Editor
for Birkbeck
College: Dr. David Moss

Department of Crystallography, Birkbeck College,
University of London, Malet Street, London WC1E 7HX

Editorial

It is most pleasing to be able to circulate this edition of the newsletter so soon after the last, and I would particularly thank David Rice (Sheffield) who stimulated and collected the contributions so efficiently. The articles are, as far as I know, published in the order of submission and it has been suggested that the last or non-contributory group should take responsibility for organising the next newsletter! I would also like to encourage non-UK protein crystallography groups to contribute if they have information on software which would be of general interest to the scientific community.

As you may know CCP4 (Collaborative Computing Project in Protein Crystallography) has recently applied for funds from the SERC to continue beyond September 1985 and I have included a note about the success of this application later in the newsletter.

Pella Machin

Progress on the Antigen-Antibody Complex Structure at Leeds/Institut Pasteur

After the initial shock of arriving in Leeds in January, work has continued on the complex in collaboration with Roberto Poljak and others at the Institut Pasteur in Paris. The crystals contain the complex between the Fab fragment of a monoclonal antibody (D1.3) raised in mice against hen egg lysozyme, and lysozyme itself. The space group is $P2_1$, with dimensions $56 \times 144 \times 49 \text{ \AA}$ and $\beta = 120^\circ$, and crystals diffract to about 2 \AA on the synchrotron, although we are currently working with 2.8 \AA data from a single counter diffractometer.

Initial work at 6 \AA resolution was done using 3 heavy atom derivatives: PtCl_4^{2-} (1 major site), PHMBS (1 major site 4 \AA from Pt + several minor sites) and pentafluoro-uranyl (3 major + several minor sites independent of other derivatives). The 6 \AA MIR map was of such good quality that it was possible with FRODO + PS300 unequivocally to fit rigid lysozyme and Fab New (human myeloma protein) molecules into the density. This crude model already shows that the antigenic determinant is "topographic" (i.e. relies on the native lysozyme structure for its existence) and no large conformation changes take place in complex formation. The observation that the single mutation Gln 121 \rightarrow His in lysozyme is sufficient to abolish complex formation fits well with the model, where the side chain pokes into the "hapten pocket" of the Fab. This was reported in Nature 313, 156-158.

Current work at 2.8 \AA is not going so smoothly. The cell dimensions are such that there are very few centric terms and this, together with the proximity of the Pt and Hg sites, has made refinement difficult. Gérard Bricogne is using our data to develop his full-matrix heavy atom refinement and maximum entropy techniques, so far without notable success but we are optimistic. Current MIR maps show some features, pieces of helices, sheets etc., but density is often broken and hard to interpret. We have recently refitted the lysozyme, VL and VH domains as rigid bodies into the density, but antibody domains still are a problem, as the D1.3 sequence is quite different from Fab New in places, and there are insertions and deletions. Since one of the objects of the exercise is to look for conformation change on binding, we are loath to calculate phases directly from co-ordinates at this stage. Any comments on experience refining out bias from a poor starting model at 2.8 \AA would be gratefully received, especially suggestions on how to prove that any bias really has gone!

I will spend August in Paris wrestling with the problems (and eating well). Current ideas include solvent flattening based on envelopes defined by the model, and use of model phases to assist heavy atom refinement. I will report success or failure of these, and any other ideas, in a later newsletter.

A bientôt.

Simon Phillips

HENDRICKSON-KONNERT REFINEMENT ON THE DAP

Jim Raftery,
Napier College and the Biochemistry Department*
Edinburgh University

The ICL distributed array processor, the DAP, has been in operation for a number of years and David Fincham wrote an article about its capabilities and potential applications in CCP4 News, Number 10. We have adapted the Hendrickson-Konnert refinement program to run on the DAP in Edinburgh in an attempt to keep the cycle time down to something manageable i.e. to stop the threatening phone-calls!

The program uses the host, an ICL 2976, to handle all of the I/O, to calculate some of the derivatives (eg. distance restraints) and, at present, to apply the conjugate gradient method.

The DAP firstly rearranges all of the data so that all of the equivalent terms are adjacent rather than all of those relating to the one atom. This is more convenient for parallel computation and the rearrangement is accomplished by CRINKLE, a library utility which allows regularly spaced elements to be addressed as if they are contiguous. The present version of the program orders the data on entry to the DAP and reorders it on exit. All scattering factor and trigonometric terms are calculated as they are needed. Incidentally, we have found significant differences in Fcalc determined this way compared with that obtained from the lookup tables in the original program, although the overall results seem to be unaffected. Whether this is an effect of the ICL machine is not yet clear.

The DAP then steps through the reflections calculating the structure factors and, if required, the derivatives which are used for the final stage of setting up the normal matrix.

Our initial trial with beta-lactoglobulin, space group B22/12, 1196 atoms, 3564 reflections and 30 iterations took 5500s on the 2976 for the space group specific version. The DAPperised version takes 415s for the identical calculation of which 190s on the host is the same for both. A spacegroup general version takes a rather disappointing 1166s and some thought is going into trying to improve this.

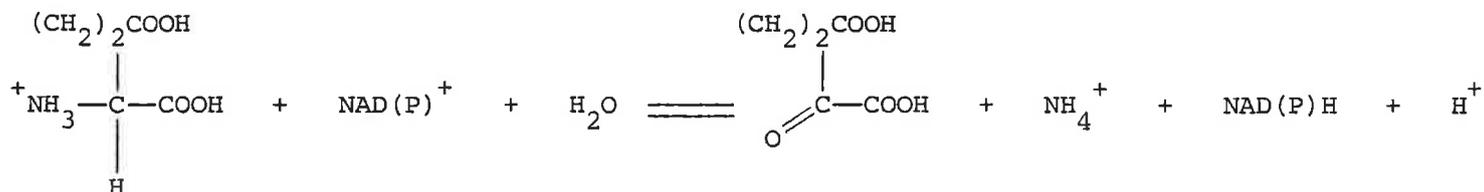
* Present address

Progress with Glutamate Dehydrogenase

Patrick Baker, George Farrants, David Rice and Fiona Rodgers

Department of Biochemistry, University of Sheffield

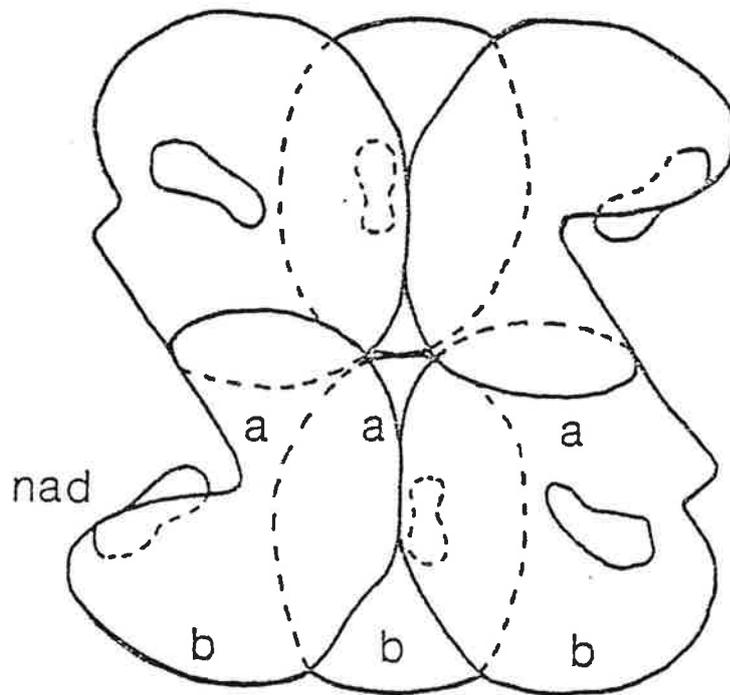
Glutamate dehydrogenases catalyse the oxidative deamination of L-glutamate to give 2-oxoglutarate and ammonia with concomitant reduction of NAD^+ or NADP^+ .



They fall into two metabolic categories: those involved in ammonia assimilation are predominantly NADP^+ -dependent, whereas those involved in glutamate catabolism are usually NAD^+ -specific. GDHs in vertebrates, however, are able to use both coenzymes with similar efficiency. GDHs from a variety of sources have been well studied partially because of their interesting regulatory properties and out of a desire to reveal their relationship to other dehydrogenases such as LDH. However, difficulties in crystallizing this enzyme have led to the absence of a 3D structure and hence confusion over the interpretation of many of these results. Recently however crystals of GDH from two bacterial sources have been obtained. The group at Leeds are working on an NADP-linked GDH from *E.coli* whilst in Sheffield we are studying crystals of an NAD-linked enzyme from *Clostridium Symbiosum*. Both of these enzymes are hexamers with a subunit molecular weight of $\sim 50,000$ Daltons and the clostridial enzyme is arranged in 32 symmetry with a half molecule (3 subunits) as the asymmetric unit of the cell.

Data were collected for these crystals and two related mercury derivatives on the 5 circle diffractometer in Oxford to 6\AA resolution and the Patterson function of both derivatives was easily solved in terms of a single site for each of

Figure 1. Schematic representation of a hexamer of glutamate dehydrogenase from *C.Symbiosum*.



the subunits. The isomorphous map calculated from these derivatives was then averaged around the non crystallographic 3 fold axis to produce a greatly enhanced electron density map. Interpretation of the averaged map was quite straightforward due to the high α helical content (probably greater than 55%) of the enzyme and a preliminary analysis revealed that the molecule was organized into two distinct domains, A and B, separated by a deep cleft. The overall arrangement of the molecule is shown in Fig. 1.

The interactions within a trimer and the contact between the two sets of trimers are mediated primarily by domain A, which lies closest to the 32 symmetry point and the most extensive interactions appear to occur around the 3 fold axis. The domain interface between domains A and B is dominated by 4 large α helices, the longest of which is about 40Å in length and these are very well defined in the averaged map. Inspection of the electron density of domain B showed that it contained 5 large helices arranged in groups of 2 and 3 with these two groups of helices being separated by some 20Å across a region of diffuse electron density. This arrangement corresponds to the separation of α helices on opposite faces of a β sheet and the structure seems to closely resemble the structural arrangement of the well characterized NAD binding fold. Data collected on crystals soaked in NAD confirmed the assignment of this domain to the enzyme's NAD binding function and placed the NAD moiety on one face of the cleft on domain A facing domain B.

Work now is concentrating on the extension of the resolution of this structure and data have now been collected on the SRS for the native enzyme and two derivatives to 2.5Å and for the substrates NAD, NADH, NAD + Glutamate and NADH + Glutamate to 3.0Å. Clearly a busy time data processing lies ahead but with the first derivative data set nearly processed it may not be too long before the exact relationship between GDH and the other dehydrogenases is revealed.

A study on the inter and intra subunit contacts in M-type Pyruvate Kinase

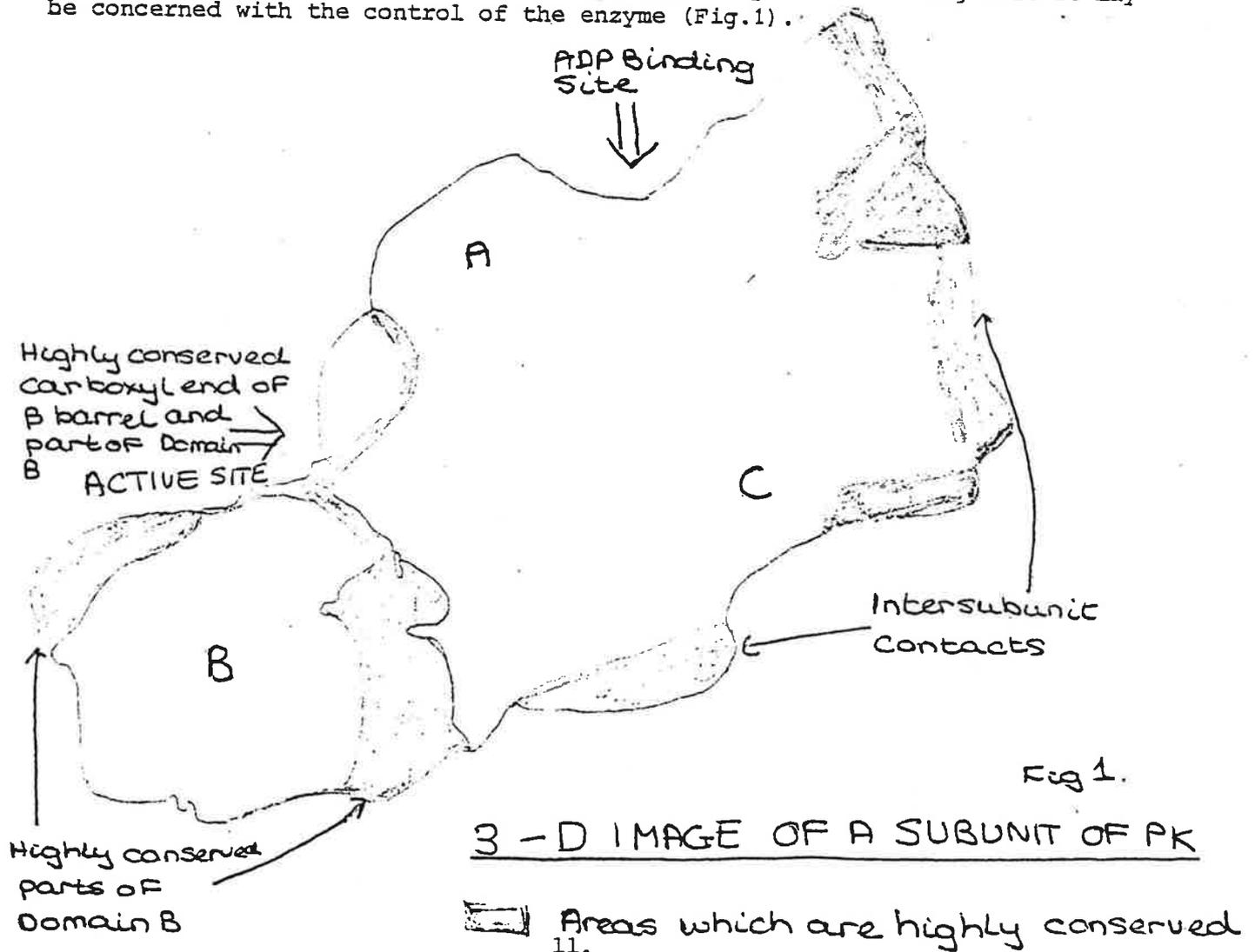
D. Clayden, Dept. of Biochemistry, University of Bristol, Bristol BS8 1TD

Pyruvate Kinase (PK) is a tetrameric enzyme which catalyses the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) to ADP, in the presence of both bivalent and monovalent cations. It is unique as a kinase in that it requires two bivalent cations as well as one monovalent cation per active site. The regulation of this enzyme is important in the control of glycolysis, especially in tissues capable of gluconeogenesis.

Amino acid and nucleic acid sequence studies have found that feline muscle pyruvate kinase has 530 amino acids per subunit; chicken muscle PK has 529 amino acids per subunit and yeast PK has 499 amino acids per subunit. Feline muscle PK amino acid sequence is 88% identical to the chicken muscle PK sequence. Yeast PK amino acid sequence is 42% identical to muscle PK amino acid sequences.

The tertiary structure of PK consists of a small N terminal domain plus three major domains A, B and C (Stuart, Levine, Muirhead and Stammers J. Mol. Biol. (1979), 134, 109-149).

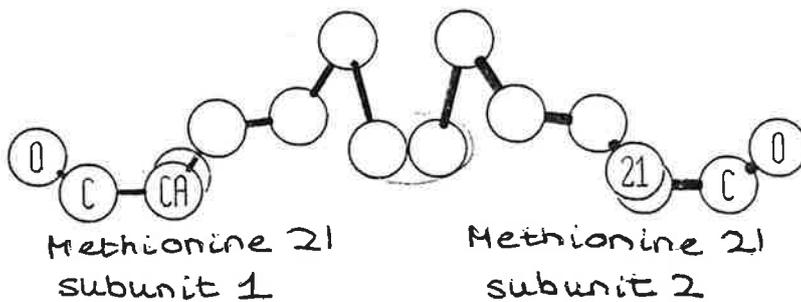
A study was made on the areas of conservation of the primary sequence in the three sources of enzyme (named above). The side chains close to the active site are 100% conserved (as one might expect). The degree of conservation differs in the three types of intersubunit contacts. There is only one contact between subunits 1 and 2 and this is conserved in all three sources. 55% of residues forming contacts between subunits 1 and 3 are conserved in all three sources. Only 27% of residues forming contacts with subunit 4 are conserved in all three sources. There is also another area in Domain B (in addition to the entrance to the active site) which is highly conserved; the reason for it being so conserved is not obvious. A possible explanation being that it may be concerned with the control of the enzyme (Fig.1).



A systematic check of all inter and intra subunit contacts is currently being made, as initially only bad contacts within the single subunit had been looked at. It is expected that areas with bad contacts would be inter-domain and inter-subunit contacts. Also those residues which make up areas connecting different elements of secondary structure may also form bad contacts with other residues. All interatomic distances of less than 4\AA are being recorded. Interatomic distances of less than 2.5\AA (except in the case of atoms involved in hydrogen bonding) give an indication that these atoms are too close together and that these areas may need rebuilding e.g. helix 6 in Domain A of subunit 1 which was packing badly against part of the N terminal peptide of subunit 3.

The plot of Fig. 2 shows an example of a bad intersubunit contact:-

PK CONTACT 21 21X



Residues involved in binding heavy atoms have been listed. There are six heavy atom sites - 3 Au, 1 Gd and 2 Pt. It is thought that Au would bind to a cysteine residue. The closest residue to the major Au site is cys 316. The second Au atom is situated close to cys 475. The third Au site (which is the weakest Au site) is closest to Met 493 and is also accessible from the surface. The Gd binds to the carboxyl group of Glu 271. The enzyme bound Mg is at the same site. The first Platinum site is close to lys 366 which is at the entrance to the active site. This lysine is active and has been chemically labelled. The second Pt bound close to the weak Au site and the nearest side chains to it were Arg 515 on subunit 2 and Isoleucine 509 on subunit 1. The side chains of Asn 522 and Thr 523 of subunit 1 are also in the vicinity. This is an area which may need rebuilding.

Thus a study of inter and intrasubunit contacts will prove to be an invaluable help in assessing whether our interpretation of the tertiary structure is correct. It will also be an aid to elucidating those regions where we are still uncertain of the tertiary structure e.g. the N terminal domain.

Use of Postrefinement for Analysis of Systematic Errors in
GAPDH High Resolution Dataset

Alan Wonacott (Imperial College)

A dataset for HoloII crystals of GAPDH to a resolution of 1.8A has been collected over an extended period on the SRS at Daresbury using vee-shaped cassettes. It consists of data from 13 crystals one of which was recorded on a GX6 rotating anode (X0) and the remainder using monochromator (X1-X9) or mirror/monochromator optics (X10-X15) on PX7.2 station.

As previously reported in Daresbury annual reports several problems have been encountered; among these are severe radiation damage when the higher fluxes available with 2.0Gev and mirror/monochromator were employed.

The final merged dataset includes 160 film packs. For fully recorded reflexions alone $R_{merge} = 7.3\%$ for a total of 116,000 independent reflexions with R_{merge} varying from 4.6% at low resolution to 20% for the 1.9-1.8A shell. Severe partial bias was observed in the merged dataset which resulted in a markedly increased R_{merge} ($=10.2\%$) when partial reflexions were included.

The nature of the partial bias ($= (I_{full} - I_{partial})/I_{full}$) is shown by the analysis as a function of partial fraction in the range 0.5 - 1.0 carried out by AGROVATA. Partial bias varies from -17% for 0.5-0.6 range to +6.4% for 0.9-1.0 range. Thus reflexions close to half-recorded are grossly underestimated, the opposite of the commonly occurring situation.

Further analysis was carried out in an attempt to diagnose the cause of this problem. The program POSTREF which has been developed by Phil Evans from the original program of Schutt & Winkler was thought likely to shed some light on the source of this systematic error.

Refined Parameters in POSTREF

Unit cell parameters were not refined since they were consistent for all crystals and well-defined from the original stills. For the synchrotron films the wavelength was allowed to vary in case there were any small differences in the wavelength setting between shifts on the PX station.

The main parameters to be refined were the misorientation and beam divergence (mosaic spread in the SRS case since the beam parameters are defined).

The program allows one to refine either the mean orientation for a film pack or individual missetting angles at the start and end of a film pack.

For the rocking curve either a cosine or hyperbolic tangent function (TANH) can be used. In practice the TANH profile proved to result in slightly better residuals.

Refinement was carried out against the mean of the safe fully-recorded reflexions.

Results

For data from crystal X0, recorded on a GX6, the refined parameters shown in Table 1. were very consistent for all batches when the divergence was refined independently for each pack. The refined value for divergence was close to that used for measurement of the data, 0.30 degrees. There is no indication of crystal slippage and the values of Φ_{12} at the boundary between contiguous packs agree to within better than 0.01 degrees (cf. below). This crystal which gave the lowest R_{merge} values in the overall merging process also resulted in the best R-factor (0.12) after post-refinement.

For crystals X1-X9, recorded with monochromator alone, the results are very satisfactory. Crystal slippage was small (less than 0.04 degrees) and consistent with the values obtained on individual pairs of contiguous films using the programs POSTCHK and IDXREF. Values of the mosaic spread, refined when using the synchrotron model, are fairly consistent and show only a slight trend upwards at the end of the crystal lifetime (see Table 2). Contiguous film packs show a difference in Φ_{12} between the end of one pack and the start of the next of the order of 0.03 - 0.04 degrees. It has been suggested that this apparent gap between film packs could be caused by slow opening of the camera shutter with the short exposures and fast rotation rates on the SRS. The magnitude of this gap is quite uniform for all of these crystals, some of which were collected with low beam flux and much slower rotation speeds. It is very unlikely that the effect is real. Rather, this is a consequence of the refinement procedure which attempts to scale up the partial reflexions to match full reflexion intensities by modifying the calculated fraction. The only variables that can influence this agreement are Φ_{11} and Φ_{12} and it follows that these tend to change.

Crystals X10-X14 which exhibited catastrophic radiation damage and which have the poorest internal agreement in the merged dataset show a similar gap between contiguous film packs (see Table 3). Results from X12 and X13 indicate a small change in orientation; this was detected during measurement of the films. The most significant result is a very marked increase in mosaicity with batch number, from < 0.05 degrees to > 0.20 degrees within a few exposures. Thus the physical nature of the crystal degradation is revealed. By allowing for this increase of mosaic

spread in the measurement of the films or by using the reclassified reflexion list generated by POSTREF we would expect to improve the merging of these crystals in the 3-D dataset.

Conclusions

Although the use of post-refinement has not revealed the cause of the excessive partial bias in this dataset, it has certainly proved to be of value in pointing out problems and has also enabled a clear distinction to be made between good and poor batches which will ultimately lead to improved data.

Table 1. Refined parameters for crystal X0

Batch	Phix1	Phiy1	Phiz1	Phix2	Phiy2	Phiz2	Divergence	R-factor
02	0.21	0.52	-0.67	0.21	0.52	-0.68	0.217	0.12
03	0.21	0.52	-0.68	0.21	0.52	-0.68	0.219	
04	0.21	0.52	-0.68	0.21	0.53	-0.69	0.213	
05	0.21	0.52	-0.68	0.21	0.52	-0.69	0.240	
06	0.21	0.53	-0.66	0.21	0.53	-0.67	0.236	
07	0.21	0.52	-0.67	0.21	0.52	-0.68	0.245	
08	0.21	0.51	-0.67	0.21	0.51	-0.68	0.231	
09	0.21	0.52	-0.67	0.21	0.52	-0.68	0.244	
10	0.21	0.52	-0.67	0.21	0.52	-0.68	0.259	

Table 2. Refined parameters for crystal X2

Batch	Phix1	Phiy1	Phiz1	Phix2	Phiy2	Phiz2	Divergence	R-factor	
2002	0.52	-0.59	0.14	0.52	-0.59	0.10	0.002	0.15	
2003	0.52	-0.59	0.13	0.52	-0.59	0.10	0.017		
2004	0.52	-0.59	0.13	0.52	-0.58	0.10	0.013		
2005	0.52	-0.58	0.13	0.52	-0.58	0.10	0.004		
2006	0.52	-0.58	0.13	0.52	-0.58	0.09	0.013		
2007	0.52	-0.58	0.12	0.52	-0.58	0.09	0.040		
2008	0.52	-0.59	0.12	0.52	-0.59	0.09	0.019		
2009	0.52	-0.59	0.11	0.52	-0.59	0.09	0.008		
2011	0.51	-0.59	0.12	0.51	-0.59	0.09	0.040		
2012	0.49	-0.60	0.12	0.50	-0.60	0.08	0.029		
2013	0.49	-0.61	0.12	0.49	-0.61	0.09	0.004		
2114	0.59	-0.72	0.16	0.59	-0.72	0.13	0.046		0.16
2115	0.59	-0.72	0.17	0.59	-0.72	0.13	0.041		
2116	0.58	-0.73	0.16	0.59	-0.72	0.13	0.036		
2117	0.58	-0.73	0.16	0.58	-0.73	0.13	0.047		
2118	0.58	-0.73	0.16	0.58	-0.73	0.13	0.055		
2119	0.58	-0.73	0.16	0.58	-0.73	0.13	0.061		
2120	0.58	-0.73	0.16	0.59	-0.72	0.13	0.058		
2121	0.59	-0.73	0.16	0.59	-0.73	0.13	0.050		
2122	0.59	-0.73	0.16	0.59	-0.73	0.14	0.056		
2123	0.59	-0.73	0.16	0.59	-0.73	0.14	0.040		
2124	0.59	-0.73	0.16	0.59	-0.73	0.14	0.077		
2125	0.59	-0.72	0.16	0.59	-0.73	0.14	0.074		
2126	0.59	-0.73	0.16	0.59	-0.72	0.14	0.089		
2127	0.59	-0.73	0.16	0.59	-0.73	0.14	0.057		
2128	0.59	-0.72	0.16	0.59	-0.73	0.14	0.082		

Table 3. Refined parameters for crystals X12,X13

Batch	Phix1	Phiy1	Phiz1	Phix2	Phiy2	Phiz2	Divergence	R-factor
12001	0.98	-0.06	0.27	0.98	-0.06	0.22	0.075	0.18
12002	1.00	-0.10	0.22	1.00	-0.09	0.17	0.063	
12003	1.00	-0.12	0.20	1.00	-0.10	0.17	0.084	
12004	1.00	-0.13	0.20	1.00	-0.13	0.16	0.123	
12005	1.00	-0.15	0.18	1.00	-0.14	0.15	0.164	
13002	-0.10	1.67	0.05	-0.10	1.67	0.01	0.004	0.19
13003	-0.10	1.68	0.07	-0.10	1.70	0.02	0.040	
13004	-0.10	1.71	0.06	-0.09	1.72	0.01	0.076	
13005	-0.09	1.74	0.07	-0.09	1.76	0.02	0.088	
13006	-0.10	1.79	0.08	-0.09	1.80	0.03	0.117	
13007	-0.09	1.82	0.08	-0.09	1.83	0.04	0.172	
13008	-0.09	1.85	0.08	-0.09	1.85	0.04	0.258	
13009	-0.10	1.85	0.08	-0.09	1.86	0.05	0.283	

CRYSTALLOGRAPHIC STUDIES ON THE VERTEBRATE LENS PROTEINS

by Huub Driessen

Laboratory of Molecular Biology,
Department of Crystallography, Birkbeck College,
Malet Street, London WC1E 7HX

The transparency of the lens depends not only on the regular arrangement of its fibre cells, but also on the interactions of the constituent structural proteins, the crystallins (Blundell et al., 1983). Throughout the course of lens development a whole range of crystallin genes is differentially expressed (Piatigorsky, 1981) producing a variety of crystallins occupying different locations along the refractive index gradient, which increases from the cortex to the nucleus. Most light scattering centres in cataracts must have involved an alteration of the supramolecular organisation of the crystallins. It is therefore of the utmost importance to understand the inter- and intra-molecular interactions of the crystallins. The Department of Crystallography of Birkbeck College is investigating this problem in two ways. Firstly a reductionist approach is used by looking at individual components and their possible aggregation behavior, secondly the aggregates themselves are being studied in a more holistic path of approach.

The crystallins of mammals are generally classified as α , β and γ according to their relative molecular masses (Bloemendal, 1981). The α -crystallins comprise two gene products organised as apparently spherical oligomers of 500,000 to 1,000,000. The β -crystallins comprise at least seven gene products which form homo- and hetero-oligomers ranging from dimers to aggregates with as many as eight chains (Berbers et al., 1982), while the γ -crystallins—a group of six gene products—are monomeric (Schoenmakers et al., 1984). The β - and γ -families are highly homologous. In avian and some reptilian lenses there is a further protein, tetrameric δ -crystallin, which replaces the γ of the mammals.

The three-dimensional structure of the γ -crystallins.

1. YII

The three-dimensional structure of bovine lens YII crystallin, which crystallizes with one molecule in the asymmetric unit of $P4_12_12_1$, has been solved by the method of multiple isomorphous replacement (Blundell et al., 1978, 1981) and has been refined to 1.6 Å resolution (Wistow et al., 1983, Summers et al., 1984) using the restrained least-squares program RESTRAIN on the ULCC Cray-1s (Haneef et al., 1985). Currently Shabir Najmudin is refining two different data sets for YII, which have been collected from crystals under different reducing conditions. This protein has seven cysteines, more than the average for 174 residues, of which two are capable of forming an intra-molecular S-S bond (Summers et al., 1984). This S-S bond has now been confirmed for the absence of reducing agent. It is important to

understand the function of the cysteines in view of the ease with which oxidation may lead to disulphide formation and denaturation in the lens.

2. YIII

The three-dimensional structure of bovine lens YIII crystallin, which crystallizes with two molecules in the asymmetric unit of $P2_12_12_1$, has been solved by the method of multiple isomorphous replacement (Chirgadze et al, 1981) and has been refined to 2.7 Å resolution. During a visit by Yuri Chirgadze of the Protein Crystallography Group, Institute of Protein Research, Poustchino, USSR, from may-july 1985, 65 degrees of 1.9 Å resolution data were collected and processed in a collaborative effort with Ian Glover and Peter Lindley and Dr. John Helliwell's group at the Daresbury Synchrotron Radiation Source. The remaining data will be collected and further collaborative work will take place. Especially the intermolecular contacts between the two pseudo-symmetrically related molecules will be investigated in order to understand the in vivo interactions of the monomeric γ -crystallins. To this end further restrained refinement will take place.

3. YIV

The three-dimensional structure of bovine lens YIV crystallin, which crystallizes with one molecule in the asymmetric unit of $C222_1$, and is homologous to the extent of 80 % with YII, has been solved by the method of molecular replacement using YII as the search molecule (Driessen and White, 1985) using data collected to 2.3 Å resolution (Christine Slingsby and Peter Lindley). To this end the CCP4 versions of the rotation function ALMN and the R-factor search program SEARCH were used. The solution was initially refined as one rigid body with the refinement program CORELS (Sussmann et al., 1977), followed by a refinement of the two domains of this molecule as independent rigid bodies. The structure is currently being refined by Helen White to 2.3 Å with RESTRAIN, and we hope soon to be able to understand some of the physiological differences between YII and YIV.

The three-dimensional structure of the β -crystallins.

For the major bovine β -crystallin subunit β Bp, which is a dimer in solution but crystallizes with two dimers in the asymmetric unit of $C222$, native single crystal X-ray diffraction data have been collected to 3.3 Å resolution (Linda Miller, Christine Slingsby and Peter Lindley). Cross-rotation studies with YII (which is homologous to the extent of ca. 30 %) and self-rotation studies have given a tentative solution for the rotation parameters (Huub Driessen). Over the last years data have been collected for both a Pt and a Hg derivative by Daruka Mahadevan to 5.5 Å resolution. Currently work is being done on an initial interpretation of a multiple isomorphous replacement map, and further derivatives are being investigated by Ben Bax.

The three-dimensional structure of the δ -crystallin.

δ -Crystallin has been purified by gel filtration under non-dissociating conditions to give tetramers of four similar subunits, each of relative molecular mass 50,000 (Narebor et al., 1980, 1985). At room temperature the

protein crystallizes with one tetramer in the asymmetric unit of space group $P2_12_12_1$, which form is currently used by Sangari Mylvaganam for native single crystal data collection at 3.3 Å resolution, both here and at the SRS at Daresbury.

Future work.

Work on the refinement of the YII, YIII and YIV will continue and initial studies on the effect of ultraviolet radiation on the oxidative state and stability of YII in the crystal, using Laue methods at the SRS will be initiated. The data collection and processing of the native δ -crystallin will be pursued. The interpretation of the β Bp data will be continued. Work has been initiated by Christine Slingsby and Fiona Giblen in the finding of a well-characterized α -crystallin oligomer, which can be used for crystallographic studies.

REFERENCES

- Bloemendal, H., Molecular and Cellular Biology of the Eye Lens, H. Bloemendal Ed., John Wiley, New York, 1981, p1.
- Berbers, G.A.M., Boerman, O.C., Bloemendal, H. and de Jong, W.W., Eur.J. Biochem., 128, 495 (1982)
- Blundell, T.L., Lindley, P.F., Moss, D.S., Slingsby, C., Tickle, I.J. and Turnell, W.G., Acta Cryst., B34, 3653 (1978).
- Blundell, T.L., Lindley, P.F., Miller, L.R., Moss, D.S., Slingsby, C., Tickle, I.J., Turnell, W.G. and Wistow, G.J., Nature, 289, 771 (1981).
- Blundell, T.L., Lindley, P.F., Miller, L.R., Moss, D.S., Slingsby, C., Tickle, I., Turnell, B. and Wistow, G., Lens Research, 1, 109 (1983).
- Chirgadze, Y.N., Sergeev, Y.V., Fomenkova, N.P. and Oreshin, V.D., FEBS Lett., 131, 81 (1981).
- Driessen, H.P.C. and White, H., in Proceedings of the Daresbury Meeting on Molecular Replacement (15-16 February, 1985), in press.
- Haneef, I., Moss, D.S., Stanford, M.G. and Borkakotti, N., Acta Cryst., in press.
- Narebor, E., Slingsby, C., Lindley, P.F. and Blundell, T.L., J.Mol. Biol. 143, 223 (1980).
- Narebor, E.M. and Slingsby, C., Exp. Eye Res., 40, 273 (1985).
- Piatigorsky, J., Differentiation, 19, 134 (1981).
- Schoenmakers, J.G.G., den Dunnen, J.T., Moormann, R.J.M., Jongbloed, R., Van Leen, R.W. and Lubsen, N.H. in Human Cataract Formation, CIBA Foundation Symposium, 106, 208 (1984).
- Summers, L., Wistow, G., Narebor, M., Moss, D., Lindley, P., Slingsby, C., Blundell, T., Bartunik, H. and Bartels, K. in Peptide and Protein Reviews, Vol. 3, M. Hearn Ed., Marcel Dekker, New York and Basel, 1984, p.147.
- Sussmann, J.L., Holbrook, S.R., Church, G.M. and Kim, S.-H., Acta Cryst. A33, 800 (1977).
- Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. and Blundell, T., J. Mol. Biol., 170, 175 (1983).

UNIVERSITY OF YORK

HESLINGTON, YORK, YO1 5DD

TELEPHONE 0904 59861

TELEX 57933 YORKUL

DEPARTMENT OF CHEMISTRY
Professor G G Dodson

12 August 1985

Dear David

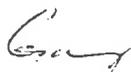
The main development at York has been the appointment of John Helliwell to the Physics Department as a new blood lecturer in protein crystallography. Obviously John will continue to contribute to the development of the SRS and he will be contributing a certain amount for a fair period. Keith Wilson meanwhile has gone to Hamburg for 3 years to look after the protein crystallography there. Fortunately we are not going to lose Keith. He will be collaborating closely with us while at Hamburg and, marvellous news, he will return to York in 3 years to a lectureship in Physics.

As probably most people will know, Dorothy Hodgkin is a frequent visitor to York; the University has made her a Visiting Professor which is nice for us. Another connection is with Peter Murray-Rust who has been made a Visiting Fellow; this recognises our extensive shared interests in graphics and structure. We are at present enjoying a short visit from Ted and Heather Baker from Palmerston North in New Zealand. Ted and Heather have been in Brian Matthews laboratory at Oregon, on sabbatical leave.

Spicek and Mirka Dauter are about to return to Gdansk after another 6 month visit working on the metal binding environment in insulin and some small molecule structures. Our Polish connection will be maintained by Zygmunt and Ursula Derewenda's return next October; they will be working on various insulin structures. Finally Keith Moffat now at Cornell, will be coming next Academic year to work at York and Daresbury with John Helliwell on synchrotron radiation and structural problems.

The arrival of the York PS 300 (and imminently the VAX) has prevented the expected hiatus following the return of the Leeds, Sheffield, York PS 300 system to Sheffield. From our point of view the moving picture system was very successful; it got us into efficient habits and gave us time to evolve our computing organisation and support. But we are now facing a move of the graphics and biochemical laboratories and this, together with the refinement studies, the water structure analyses and the films Rod Hubbard and Max Perutz are making is beefing up the pressures; so much so that Rod is going to Harvard for a week to get some work done!

Yours



Guy Dodson

Dr D Rice
Biochemistry Dept
University of Sheffield
Sheffield
S10 2TN

Frodo and the PS300

P.R.Evans
MRC Laboratory of Molecular Biology
Hills Road
Cambridge
CB2 2QH

Over the last two or three years, the Evans & Sutherland PS300 has been displacing the earlier PS2 and Multi-Picture Systems as the standard machine for molecular graphics, and many crystallographers are now familiar with using them. The PS300 differs from previous devices in that it performs many of the functions which were previously performed by the host computer. This has the advantage that most of the real-time operations (rotations, translations etc), and most of the book-keeping of display lists are handled by the PS300, but the disadvantage that operations on the transformed coordinates are hard to do, in particular distance calculations between different parts of a structure and bump checking for docking. Conversion of programs to run on the PS300 is not straightforward, as decisions have to be made as to what is done in the PS300 (most operations that directly linked to interactive dials, buttons etc) and what is to be done in the host (the setting up of display lists, ie 'drawing' the molecule, maps etc). Most installations at present have only a slow link (9600 or 19200 baud) between the host and the PS300, which discourages frequent interchange of large amounts of data, and this also complicates writing programs for the PS300. Thus although the PS300 is in general more powerful than earlier graphics devices in that it can display more vectors etc, a few facilities are difficult to implement on it.

Probably the most widely used molecular modelling program is Frodo, written by Alwyn Jones. Frodo was originally written for the interpretation of electron density maps, and this probably remains its major use, but facilities have been added that make it useful for general display and structure comparison, and to a lesser extent for speculative model building. Frodo has been implemented on a number of graphics devices, and various people have modified it over the years, retaining its basic structure, but confusing its history. Frodo was implemented on the PS300 by the group at Rice University, Houston, Texas, originally by Jim Pflugrath and Mark Saper: this version has been widely distributed. Over the last two years, I have developed an alternative PS300 version, starting from a Fortran-77 version constructed by Rod Hubbard from a Multi-Picture System Frodo and pieces of the Rice University code. This has by now been extensively altered to add new facilities and to make it easier to use. This note describes briefly some of the differences from other versions of Frodo.

1. Handling of density maps.

The parts of the program handling the display of maps has been completely rewritten. The program will take maps as packed density maps (contoured on the fly), as precontoured vector maps, or as ridge-line maps. These are all treated in similar ways in that the selection of which bricks to be displayed is common to all types of map. The minimum element of the display for each map is the vectors at one contour level for a 'brick', ie a volume of space defined by $NX \times NY \times NZ$ grid points. For all types of map, once a brick of

vectors at a particular contour has been sent to the PS300, it need not be sent again in the same session. Up to 4 maps (of different types) can be handled, provided they are on the same grid, and up to a total of 8 different contour levels, all of which may be coloured and switched on and off independently. Usually, the program is run such that whenever a new set of atoms are selected, all bricks close to those atoms are automatically displayed. The volume displayed may also be edited by adding or removing a region around a point which may be moved around the screen.

2. Symmetry.

The symmetry routines have also been rewritten. Both crystallographic and non-crystallographic symmetry operations can be used. Symmetry can be used in two distinct ways: in the selection of both foreground and background atoms symmetry can be used to choose all atoms which lie in a defined sphere. Alternatively, for background display (see below), a molecule or subunit can be displayed with a specified symmetry operation, so that oligomeric molecules and crystal packing can be generated.

3. Background displays.

The background display part of Frodo (MOL) provides a very flexible way of drawing molecules. These pictures can provide a static backdrop for model-building, or can be used to examine structures. This part of the program provides flexible colouring options (colouring by atom type, B-factor or by region), molecular surfaces (Van der Waals, or solvent-accessible surfaces calculated

externally), external vector lists, extra lines and points, and atom or residue labels. The display may be divided into parts which are separately switchable and colourable, and sub-pictures can be created which are then displayed ('instanced') with different symmetry operations. The PS300 facilities for creating hierarchical data structures makes this sort of display easy: once a picture has been created with a unique name, the host computer can forget all about it except what it is called.

The background display is very useful to provide the context of the region of the molecule being changed. The foreground atoms, which are altered in model-building, are all drawn the same colour: this would be difficult to change on the PS300 since for efficiency vectors must be sorted into their colours. A superimposed background picture coloured by atom type then provides both a visual identification of the atoms, and a record of the starting point of the model-building.

4. Regularization

Some minor modifications have been made to the regularization option REFI. The anchoring (ANCH) of the ends of the refinement zone now allows only one end to be anchored, and the definition of atoms to be fixed can be changed, so that nucleic acids can be dealt with.

5. Keyboard commands.

The format of keyboard commands is more flexible than in earlier versions: parameters to commands may be given on the same line as the command, or otherwise the program prompts for the required input. All

commands may be abbreviated provided that they are unique. All typed input can be recorded in a file, and 'played back' later (after editing if necessary). In particular, this provides a convenient way of saving complex background pictures.

6. Extra facilities.

Various minor extra facilities are provided. These include an interactive Ramachandran plot, a ruler that can be moved around the screen, and a hydrogen-bond search (with editing of the H-bond list).

7. Missing facilities.

The major missing facility present in earlier Frodo's is the continuous distance calculation between parts of the structure moved by FBRT or TOR. This is difficult to provide on the PS300, although Evans & Sutherland have now provided functions which could be used for a limited distance monitor on pre-nominated distances (not a continuous neighbour search), and it is possible to implement such facilities doing the distance calculation in the host, but to avoid swamping the speed of the slow interface, it is necessary not to update the distances too often.

Appendix A list of the most recent changes (for current users)

- 19/8/85 E2.3 Fixed small array dimension error in SCOLVW (MOLCUL)
The setting of atom colours and VDW radii for MOL needs making more flexible sometime. Failed to get soft control/C interrupt to work: it blows the GSR routines.
- 14/8/85 E2.3 New version of Frodonet1, all global, FBRT and move networks replaced with simpler versions using the E&S supplied user-written functions AAROT, TRANSPOSE, ORTHONORM.
Network LAST.FUN also changes the sampling rate on all dials to 12 /second (this removes the sluggishness of the new rotate networks), and increases the sampling rate of the tablet ('L')
- 14/8/85 E2.3 Nick Strauss fixed bugs in RIDGE
- 11/7/85 E2.2 Changed MAKERD so that residues have a legal number, numbered from 1 at the beginning
- 10/7/85 E2.2 Added VDW radii for Cl (instead of Zn)(subroutine SCOLVW/MOLCUL)
- 25/6/85 E2.2 removed switches of picking off and on from FBRT display structure: this caused spurious atom picking when FBRT was active. Changed DATASTRUC.FUN
- 24/6/85 E2.2 Added .Yes/.Fbrt option to activate all foreground atoms as FBRT object (changed FBRT)
Fixed buglet in HBND, .NoDA can't restart h-bond search if .No has been cleared by .NoID. .NOid now does not clear .No (INTER)
- 27/5/85 E2.1 Key O or o restores last orig centre after 'C'
- 24/5/85 E2.1 Small bug in ridge-line display in MAPSUBS
- 23/5/85 E2.1 Fixed bug in READWH to allow you to read non-standard "Hendrikson" format files.
- 22/5/85 E2.1 Fixed small bug in SYMGEN (SYMTRY) which caused the basic molecule to be drawn as symmetry-related if the sphere was larger than the unit cell.
- 21/5/85 E2.1 Changes to MOL: LABEL now sets flag, so that labelling is done by all atom-drawing commands label is on.
Added options to read and draw molecule surface file (from Connolly program): MSURFACE sets flag so that atom-drawing commands select atoms, then read corresponding surface, and DOTS reads whole surface file. See SURFACE.TXT for generating surfaces.
Key 'C' now centres and scales picture to foreground atoms (see FASGEN, FRODONET3)
Added SYMPrint command to Chat, now prints cell dimensions

- 8/5/85 E2.0 Fixed bug in PSUPDT, since firmware A1 changes first vector in Send VL to P, so may have to be changed back to L.
- 7/5/85 E2.0 Changed stereo options to shear the two views rather than to rotate them: this gives less problems with slab clipping. The stereo "angle" is now fixed at ~3 degrees. Changes in DATASTRUC.FUN, FRODONET2.MAC
- 2/5/85 E2.0 Trapped zero distances in Hbond search. Allow H-bond search in MOL (command HBOND to draw H-bonds, BOND to draw covalent bonds). MOL can now instance permanent objects BOND, FBND, and HBND
- 30/4/85 E2.0 Made HBND object separately switchable, on function key F4 instead of axes (now included in menu object). Changes to DATASTRUC.FUN, FRODONET2.MAC.
HBND/DANG calls function ARCCOS instead of fortran function ACOS so that $\cos(a) > 1$ is trapped.
- 23/4/85 E2.0 Change ZLNK command to allow two pairs of link atoms so that can draw proteins and nucleic acids together
- 19/4/85 E2.0 Added MOL option RLIST to draw list of residues. Also made MOL and SAM help commands look at rest of line
MOL/FLASH checks for recursive instancing
- 4/4/85 E2.0 Added menu option .Hbnd, like .Bond except draw dashed bonds. Routine CLRHBN, entry HBND, also does H-bond search if .yes & .no are not set.
- 1/4/85 E2.0 Changed to using GSR routines instead of PSIO routines. All GSR calls are in P3GSRIO, alternative PSIO versions in P3PSIO
Small changes made to MAKMAP (MAPSUBS), NAYBR, FRODOMAIN to make two versions compatible. Chat command TERM sets PS300 interface type (async, DMR-11 or parallel) as well as whether you have a separate terminal or not. Displays are now switched on or off on going between Chat and Inter, instead of being toggled.

RENEWAL OF CCP4 IN PROTEIN CRYSTALLOGRAPHY

Pella Machin (Daresbury)

The CCP4 in Protein Crystallography was initiated in 1979 as one of approximately 10 collaborative computing projects supported by the UK Science and Engineering Research Council (SERC) at the Daresbury Laboratory. CCP4 is funded by the SERC and has aimed to coordinate UK computing efforts in Protein Crystallography. It was initially funded for three years and was renewed for a further 3 year term in 1982.

The proposal to renew the CCP again has now been considered by the various SERC committees, and I am very pleased to report that the renewal has been agreed, with the recommendation that the associated RA post be for 5 years, subject to review after 3 years.

Dr. John Campbell who has been working with CCP4 as an RA for the past six years, has recently taken up a full-time post at Daresbury to work in the Applications Group of the Computing Division on SRS software development and support. I am sure that everyone connected with CCP4 will join me in thanking John for his efforts over the past years and will wish him well in his new job. The RA post is currently being advertised and I hope that this vacancy will be filled in the near future.

MEETING ON PROTEIN STRUCTURE PREDICTION,
AT DARESBURY 24/25 JANUARY 1986

Pella Machin (SERC, Daresbury)

CCP4 is organising a meeting on "The Prediction of Protein Structure" which is to be held at the Daresbury Laboratory on 24/25 January 1986. Dr. Mike Sternberg (Birkbeck College, London) agreed to organise the scientific programme and we hope that this subject will be of interest both to protein crystallographers and some gene sequencers who are non specialists in structure prediction but will be familiar with principles of protein information.

The meeting will be advertised more formally later this year and copies of the poster and application form will be circulated to all those on the CCP4 mailing list. Anyone requiring further information should contact me (Tel. 0925-65000 Ext 350).

Unscrambling of harmonic reflection intensities from spots
on Laue patterns: Results on Pea Lectin

By S.Zurek, M.Z.Papiz, P.A.Machin and J.R.Helliwell
S.E.R.C. Daresbury Lab.

1. Introduction to problem and background
2. Method of solution
3. Implementation
4. Results from Pea Lectin

1. Introduction to problem and background

An unfortunate complication of the Laue method is the introduction of harmonics superimposed to form composite spots on the film. The intensity of one of these spots is given by the sum of the intensity contributions from the spot's constituent reflections. Thus a particular spot's intensity may be a function of several wavelengths, which cannot be deconvoluted with the information on just one film. Before unscrambling, these multiplet spots cannot be used in processing since a one to one mapping from intensity to wavelength is required.

As a fraction of the total number of reflections, multiplet spots account for some 20% of Laue pattern data, none of which can be used in structure analysis involving the Fourier transform. Additionally it is desirable to obtain low resolution data from the Laue pattern. Ironically low resolution spots are also usually associated with harmonics and need to be unscrambled.

The two points above illustrate the potential usefulness of unscrambling multiplet spots.

2. Method of solution

It is clear that more than one film is necessary for unscrambling to be possible, and in fact the number of films in the pack determine the degree of unscrambling that can be accomplished. A pack of six films, as used in the Pea Lectin study, gives a theoretical limit of six reflections, though in practice the limit is much lower, since over- and under-exposed spots must be rejected.

Assuming, however, that M good films are available, then we can unscramble multiplets of order N, where $N \leq M$, as follows:

For a given spot (x,y),

let $I_{obs}(f)$ be the observed intensity on film f
let $I_{wav}(w)$ be the intensity component on film l due to a reflection
of wavelength w
let $t_f(f,w)$ be the TRANSMISSION factor from film l of the pack
to film f, at the wavelength w

then

$$[I_{obs}] = [t_f] * [I_{wav}]$$

or

$$I_{obs}(f) = \sum_w (t_f(f,w) * I_{wav}(w))$$

This is a set of linear simultaneous equations which can be solved by using a linear least squares method. The observed intensities are weighted

according to their standard deviations, and the unscrambling process gives an estimate of standard deviation for each unscrambled intensity.

3. Implementation

The above algorithm was implemented in FORTRAN 77 on a VAX 11/750 computer. Called UNSCRAM, the program accepts two files:

- o a .VC file containing a set of Victoreen function coefficients from which transmission factors are derived.
- o a .GE file which describes the integrated Laue pattern

Certain other parameters are also required which are obtained via interaction at the terminal:

- o unit cell dimensions
- o intensity cut-offs -- used for rejecting over- and under-exposed spots
- o sigma cut-off -- rejects intensities with large standard deviations
- o wavelength cut-offs -- maximum and minimum wavelengths for reflections

After unscrambling, R_{symm} factors are calculated to give an indication of reliability, although the acid test for unscrambling will be the comparison of unscrambled Laue intensities and intensities derived from monochromatic experiments (as has already been done for singlet reflections of Pea Lectin Laue patterns and reported in the previous newsletter and in press in the literature).

Finally the program allows the output of unscrambled reflections with L_p corrections, in a file format suitable for LAUESCAL, a CCP4 program which performs wavelength normalisation and allows comparison with monochromatic data.

4. Results from Pea Lectin

Unscrambling has only been applied to one pack - a further 4 packs are now being processed. The following is edited output from the unscrambling program.

Input .ge file is pea

Victoreen coeffs file is peas12345.vc

direct-lattice axes	50.3700	60.5800	135.500
intensity range	9000.00	100.000	
number of SD's before cut-off	5.00000		
wavelength range	0.300000	2.50000	

Number of reflexions read and unscrambled = 711

index for pairing up is h

wavelength mismatch limit 0.05

R-factor rejection threshold 0.0 i.e. off

Multiplicity = 2

For the PACK

R-symms on intensity

200	400	600	800	1000	1200	1400	1600	1800	2000	>2000
763	130	56	36	25	18	8	9	4	8	49
22.0	15.2	9.6	13.1	8.5	9.4	8.2	6.2	5.7	5.2	7.0

R-symms on wavelength

<0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	>2.3
30	120	120	168	180	144	108	114	48	54	20
17.4	14.7	7.0	11.4	5.4	5.9	10.0	13.9	22.6	17.1	30.7

R-symms on dhkl

>20	10	7	5	4	3.33	2.9	2.6	<2.6
0	6	77	358	143	180	234	108	0
-1.0	5.0	11.2	6.6	11.7	10.6	11.7	25.9	-1.0

R-symm distribution across pack

0.0-2.5	2.5-5.0	5.0-7.5	7.5-10	10-12.5	12.5-15	15-17.5	17.5-20	>20
92	113	98	92	72	72	54	52	461

Overall R-symm across pack = 9.5

Multiplicity = 3

For the PACK

R-symms on intensity

200	400	600	800	1000	1200	1400	1600	1800	2000	>2000
98	19	8	10	6	4	0	3	1	3	9
35.0	33.4	17.5	18.1	26.3	16.6	-1.0	9.1	3.0	9.5	16.0

R-symms on wavelength

<0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	>2.3
0	18	36	42	24	0	12	6	12	6	5
-1.0	55.8	15.4	19.0	11.5	-1.0	21.1	31.5	25.4	97.1	21.3

R-symms on dhkl

>20	10	7	5	4	3.33	2.9	2.6	<2.6
0	0	24	11	66	18	24	18	0
-1.0	-1.0	33.1	35.3	14.7	24.7	24.2	38.0	-1.0

R-symm distribution across pack

0.0-2.5	2.5-5.0	5.0-7.5	7.5-10	10-12.5	12.5-15	15-17.5	17.5-20	>20
8	6	7	6	8	8	10	7	101

Overall R-symm across pack = 18.3

Multiplicity = 4

*** No suitable reflexions at this multiplicity ***

Multiplicity = 5

*** No suitable reflexions at this multiplicity ***

Multiplicity = 6

*** No suitable reflexions at this multiplicity ***

Overall R-symm across pack and all multiplicities = 10.8

Computing at the Laboratory of Molecular Biophysics

Garry Taylor (Oxford University)

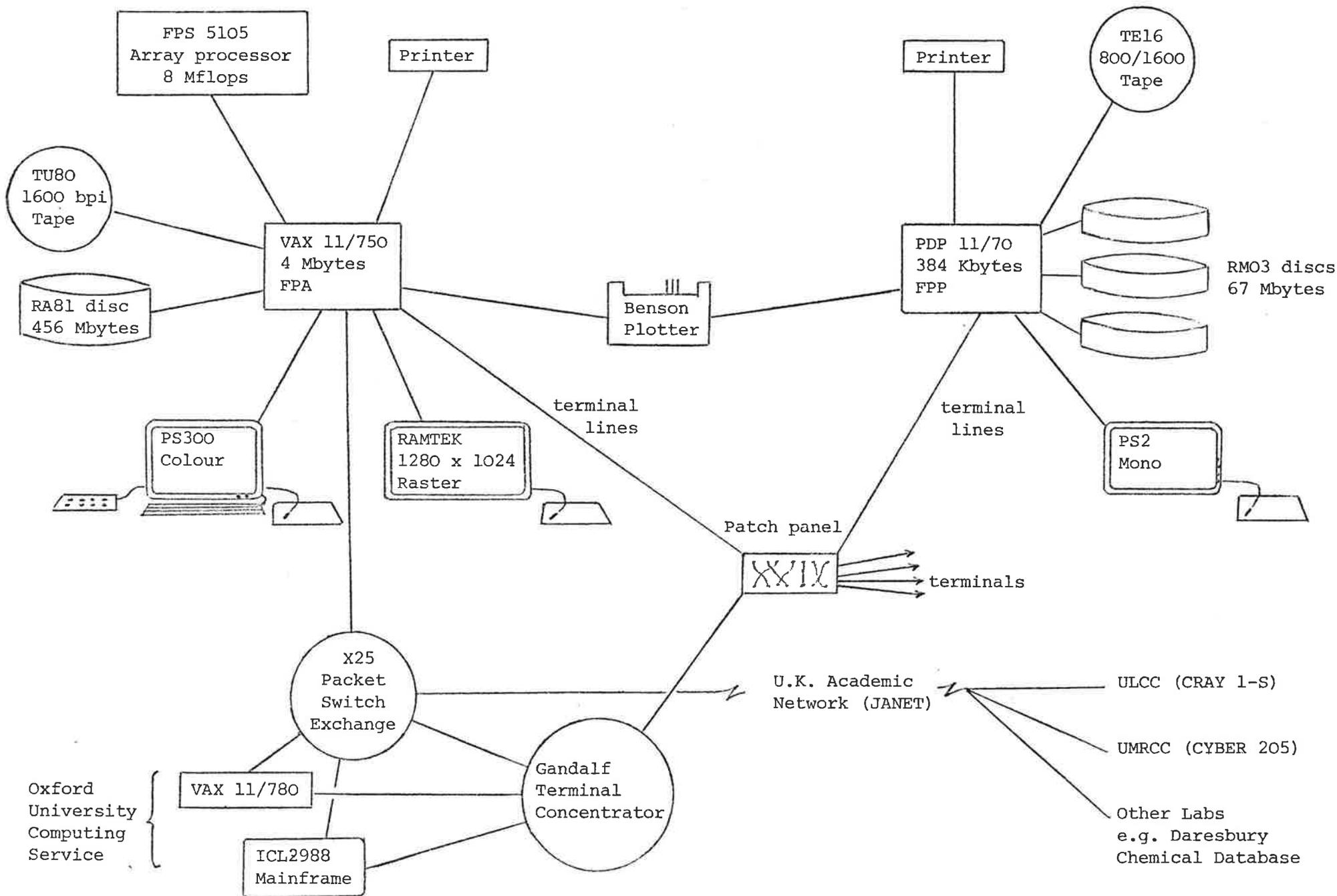
The past year has seen a major improvement in the Laboratory's computing facilities. The move to the newly-completed Rex Richards Building provided the opportunity to plan a purpose-built computer suite and we now have graphics, machine and terminal rooms with false flooring and integral air conditioning. A patch panel allows allocation of terminal lines to most of the other rooms on the two floors of the building occupied by Molecular Biophysics. A VAX 750 and an Evans & Sutherland PS300 were installed in January; an FPS array processor and a RAMTEK high resolution colour raster screen in March.

The figure shows a schematic view of the present facilities.

Graphics

(i) PS300 : Installation of this has stimulated a great increase in the use of graphics, both from the ability to display more vectors and through the use of colour. We run the version of FRODO most recently updated by Phil Evans and Rod Hubbard. This is proving an excellent version of the program, with FRODO being used in earlier stages of structure analysis by molecular replacement into a low resolution map for the Fv fragment of the mouse myeloma protein M315 (S. Collett & B. Sutton) and also for studying the structural homology between β -lactamase I from *B. cereus* and a transpeptidase (B. Samraoui & P. Artymiuk). The flexible MOL option is proving useful in model building studies : in particular in the docking of the loop region of lysozyme into model-built Fv structures to search for possible molecular interactions between the monoclonal antibodies and lysozyme (B. Sutton & P. de la Paz).

The demand for the PS300 is such that we are retaining the black & white PS2. It is still a very reliable device for electron density fitting and has the advantage over the PS300 of having a refresh buffer. This allows plots to be made of exactly what is displayed on the scope.



(ii) RAMTEK : A high resolution (1280 x 1024 x 8 bits) colour raster screen was acquired primarily for use in image analysis of electron micrograph pictures, but has also proved useful for the display of surface images of proteins and in analysis of diffraction photographs. We have installed SEMPER from the High Resolution Electron Microscopy Laboratory in Cambridge. This is a very powerful program suite for the analysis of pictures, providing, *inter alia*, masking, Fourier transformation, filtering and correlation averaging on images of up to 1024 x 1024 pixels. The RAMTEK has a pen and tablet allowing interaction with the picture as well as rapid zooming, panning and grey/colour scale manipulation of the image. The system is being used by members of the Zoology Department to study ordered 2-D arrays of the membrane cytoskeleton network in squid photoreceptor microvilli, while future projects include study of the anion transport protein of the human erythrocyte membrane.

We have also installed Arthur Lesk's CPK image program to complement our vector graphics facilities and are using the device in the analysis of Laue photographs taken at the SRS, Daresbury (D. Stuart).

VAX 11/750

The VAX supports the PS300, RAMTEK and array processor, as well as acting as a node on the academic network JANET. This means that we are now able to transfer files to and from the University machines (VAX 11/780 and ICL2988) as well as to and from the CRAY at ULCC and the CYBER at UMRCC. The VAX is used for structural analysis in general and we have recently installed Dr. van Gunsteren's GROMOS suite of programs for energy refinement and molecular dynamics. Program development includes use of the array processor (G. Taylor analysis of Laue photographs containing up to 34000 diffraction spots (D. Stuart) and searching for favoured ligand binding sites (P. Goodford).

Array Processor

The FPS 5105 consists of one adder and one multiplier, 256 Kwords (38 bit) of data memory and separate table and program memories. Through the use of a multiple bus structure, each 64 bit instruction word can achieve several operations in each clock cycle

of 250 ns : this gives a theoretical maximum performance of 8 Mflops (8 million floating point operations per second). Dr. Furey (Pittsburg) kindly provided us with his programs written for the precursor of the 5105, the AP 190-L, including Fourier syntheses and Hendrickson-Konnert refinement (PROLSQ). The 190-L was limited to 64 Kwords of data memory, limiting the number of atoms in PROLSQ to about 2000. The program has therefore been extended to make use of the larger memory on our processor - not a trivial exercise as the memory is arranged into 64 Kword pages and some thought was needed as to which page one placed various vectors and matrices for greatest efficiency. Only the structure factor and derivative calculation is carried out on the array processor. For α -lactalbumin with 1000 atoms and 15000 structure factors, one refinement cycle took 18 minutes (3 minutes of VAX cpu time & 15 minutes of AP time). This compares with almost 5 hours of VAX time when run totally on the VAX. When we have introduced FFT's, the array processor time will also be dramatically reduced and we expect to perform several cycles of refinement in house in the time taken to send and retrieve one cycle from the CRAY in London!

We are currently adapting the computer intensive parts of energy calculations and molecular dynamics simulations for the array processor.

The Future

The major part of our crystallographic computing continues to be carried out on the ICL2988 because of the large investment of effort in providing a coherent suite of programs over the past 20 years. The University will soon be looking for a replacement for this machine and we shall be looking with interest, both from the view of interfacing our facilities with it and with respect to our crystallographic suite which, because of the idiosyncrasies of the ICL machine, has become somewhat machine-dependent. Another area we are looking at actively is the upgrade of the computer control of our diffractometers where a system based on micro PDP's or micro VAX's connected via an ethernet to our main VAX would seem a good solution.